V

CELLS IN THEIR SOCIAL CONTEXT

Part V
Chapters

19 Cell Junctions, Cell Adhesion, and the Extracellular Matrix

20 Cancer

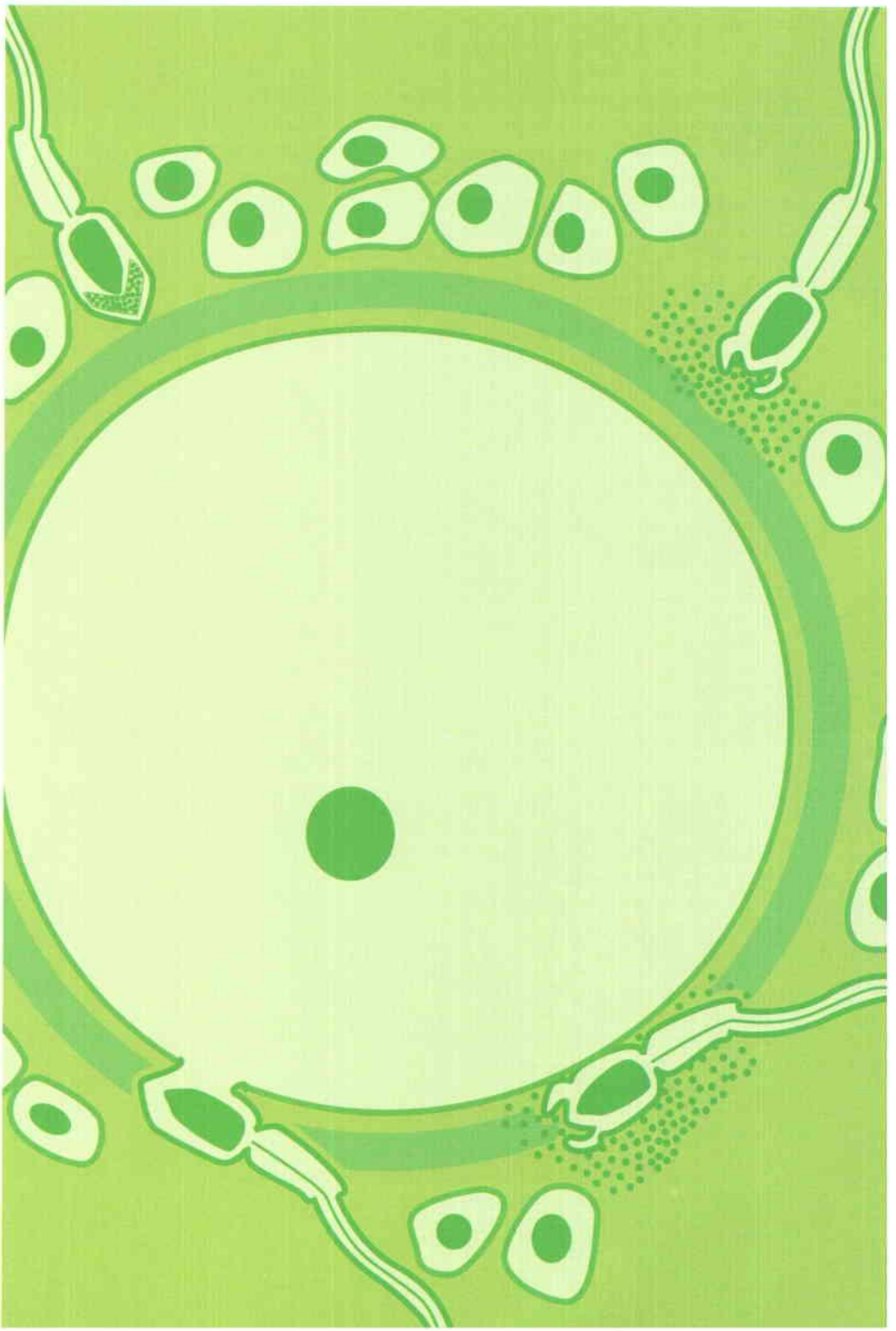
21 Sexual Reproduction: Meiosis, Germ Cells, and Fertilization

22 Development of Multicellular Organisms

23 Specialized Tissues, Stem Cells, and Tissue Renewal

24 Pathogens, Infection, and Innate Immunity

25 The Adaptive Immune System



Cell Junctions, Cell Adhesion, and the Extracellular Matrix

19

Of all the social interactions between cells in a multicellular organism, the most fundamental are those that hold the cells together. Cells may cling to one another through direct cell–cell junctions, or they may be bound together by extracellular materials that they secrete; but by one means or another, they must cohere if they are to form an organized multicellular structure.

The mechanisms of cohesion govern the architecture of the body—its shape, its strength, and the arrangement of its different cell types. The junctions between cells create pathways for communication, allowing the cells to exchange the signals that coordinate their behavior and regulate their patterns of gene expression. Attachments to other cells and to extracellular matrix control the orientation of each cell's internal structure. The making and breaking of the attachments and the modeling of the matrix govern the way cells move within the organism, guiding them as the body grows, develops, and repairs itself. Thus, the apparatus of cell junctions, cell adhesion mechanisms, and extracellular matrix is critical for every aspect of the organization, function, and dynamics of multicellular structures. Defects in this apparatus underlie an enormous variety of diseases.

As examples of structural engineering, large multicellular organisms represent a most surprising feat. Cells are small, squishy, and often motile objects, filled with an aqueous medium and enclosed in a flimsy plasma membrane; yet they can combine in their millions to form a structure as massive, as strong, and as stable as a horse or a tree. How is this possible?

The answer lies in two basic building strategies, corresponding to two ways in which stresses can be transmitted across a multicellular structure. One strategy depends on the strength of the *extracellular matrix*, a complex network of proteins and polysaccharide chains that the cells secrete. The other strategy depends on the strength of the cytoskeleton inside the cells and on *cell–cell adhesions* that tie the cytoskeletons of neighboring cells together. In plants, the extracellular matrix is all-important: plant tissues owe their strength to the cell walls that surround each cell. In animals, both architectural strategies are used, but to different extents in different tissues.

Animal tissues are extraordinarily varied, as we shall see in Chapter 23, but most fall into one or other of two broad categories, representing two architectural extremes (**Figure 19–1**). In **connective tissues**, such as bone or tendon, the extracellular matrix is plentiful, and cells are sparsely distributed within it. The matrix is rich in fibrous polymers, especially *collagen*, and it is the matrix—rather than the cells—that bears most of the mechanical stress to which the tissue is subjected. Direct attachments between one cell and another are relatively rare, but the cells have important attachments to the matrix, allowing them to pull on it and to be pulled by it.

By contrast, in **epithelial tissues**, such as the lining of the gut or the epidermal covering of the skin, cells are closely bound together into sheets called **epithelia**. The extracellular matrix is scanty, consisting mainly of a thin mat called the *basal lamina* (or *basement membrane*), underlying one face of the sheet. Within the epithelium, the cells are attached to each other directly by cell–cell adhesions, where cytoskeletal filaments are anchored, transmitting stresses across the interior of each cell, from adhesion site to adhesion site.

In This Chapter

CADHERINS AND CELL–CELL ADHESION 1133

TIGHT JUNCTIONS AND THE ORGANIZATION OF EPITHELIA 1150

PASSAGEWAYS FROM CELL TO CELL: GAP JUNCTIONS AND PLASMODESMATA 1158

THE BASAL LAMINA 1164

INTEGRINS AND CELL–MATRIX ADHESION 1169

THE EXTRACELLULAR MATRIX OF ANIMAL CONNECTIVE TISSUES 1178

THE PLANT CELL WALL 1195

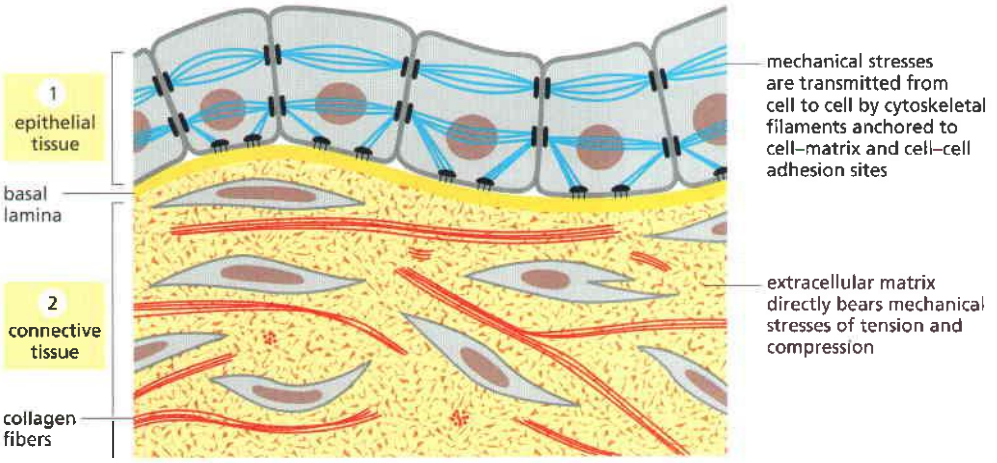


Figure 19–1 Two main ways in which animal cells are bound together. In connective tissue, the main stress-bearing component is the extracellular matrix. In epithelial tissue, it is the cytoskeletons of the cells themselves, linked from cell to cell by anchoring junctions. Cell–matrix attachments bond epithelial tissue to the connective tissue beneath it.

Physical attachment is critical, both in epithelia and in nonepithelial tissues, but junctions between cell and cell or between cells and matrix are diverse in structure and do more than just transmit physical forces. Four main functions can be distinguished, each with a different molecular basis (**Figure 19–2** and **Table 19–1**):

1. **Anchoring junctions**, including both *cell–cell adhesions* and *cell–matrix adhesions*, transmit stresses and are tethered to cytoskeletal filaments inside the cell.
2. **Occluding junctions** seal the gaps between cells in epithelia so as to make the cell sheet into an impermeable (or selectively permeable) barrier.
3. **Channel-forming junctions** create passageways linking the cytoplasms of adjacent cells.
4. **Signal-relaying junctions** allow signals to be relayed from cell to cell across their plasma membranes at sites of cell-to-cell contact.

Chemical synapses in the nervous system (discussed in Chapter 11) and immunological synapses, where T lymphocytes interact with antigen-presenting cells (discussed in Chapter 25), are the most obvious examples of signal-relaying junctions, but they are not the only ones. Sites of cell–cell communication via transmembrane ligand–receptor pairs such as Delta and Notch, or ephrins and Eph receptors, as discussed in Chapter 15, fall under this heading: the cell membranes must be held in contact with one another for the ligands to activate the receptors. Moreover, we shall see that anchoring junctions, occluding junctions, and channel-forming junctions, in different ways, all can have important roles in signal transmission.

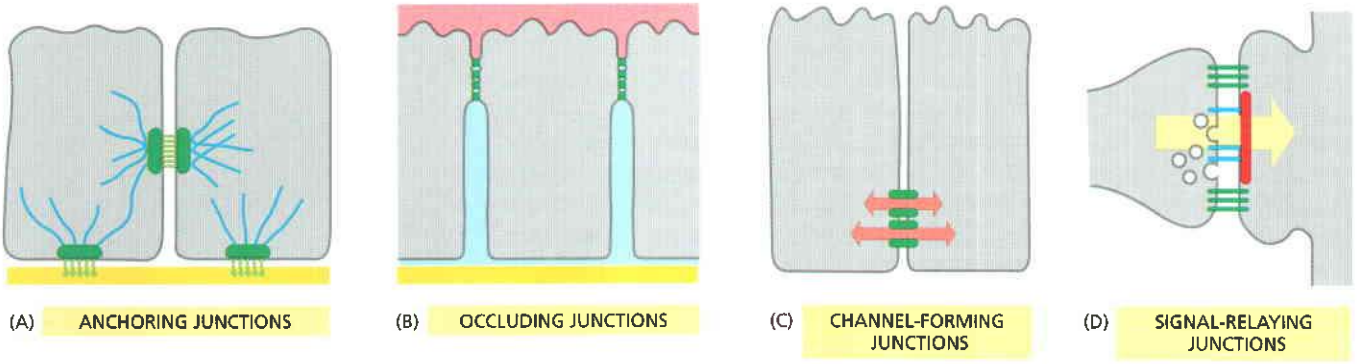


Figure 19–2 Four functional classes of cell junctions in animal tissues. (A) Anchoring junctions link cell to cell (typically via transmembrane *cadherin* proteins) or cell to matrix (typically via transmembrane *integrin* proteins). (B) Occluding junctions (involving *claudin* proteins) seal gaps between epithelial cells. (C) Channel-forming junctions (composed of *connexin* or *innexin* proteins) form passageways for small molecules and ions to pass from cell to cell. (D) Signal-relaying junctions are complex structures, typically involving anchorage proteins alongside proteins mediating signal transduction.

Table 19–1 A Functional Classification of Cell Junctions

ANCHORING JUNCTIONS	
Actin filament attachment sites	
1.	cell–cell junctions (adherens junctions)
2.	cell–matrix junctions (actin-linked cell–matrix adhesions)
Intermediate filament attachment sites	
1.	cell–cell junctions (desmosomes)
2.	cell–matrix junctions (hemidesmosomes)
OCCLUDING JUNCTIONS	
1.	tight junctions (in vertebrates)
2.	septate junctions (in invertebrates)
CHANNEL-FORMING JUNCTIONS	
1.	gap junctions (in animals)
2.	plasmodesmata (in plants)
SIGNAL-RELAYING JUNCTIONS	
1.	chemical synapses (in the nervous system)
2.	immunological synapses (in the immune system)
3.	transmembrane ligand–receptor cell–cell signaling contacts (Delta-Notch, ephrin-Eph, etc.). Anchoring, occluding, and channel-forming junctions can all have signaling functions in addition to their structural roles

The first part of this chapter will focus on animal cells and tissues, beginning with the cell–cell adhesions, occluding junctions, and channel-forming junctions that link cell to cell directly. As examples of signal-relaying junctions, we shall briefly examine neuronal synapses from the point of view of their adhesion mechanisms and assembly. We shall see how the different kinds of junctions together organize cells into polarized epithelial sheets. We shall then discuss the extracellular matrix in animals and the ways in which the cells interact with it through cell–matrix adhesions. Last, we shall turn to plants and the central role of the plant cell wall in their construction.

CADHERINS AND CELL–CELL ADHESION

The structures of **cell–cell adhesions** are most clearly seen in mature epithelia and in some other tissues, such as heart muscle, that are held together by strong direct anchorage of cell to cell. Study of these tissues by electron microscopy provided the first general classification of cell junctions. Biochemistry and molecular biology have since shown that the different structures seen in the electron microscope relate to distinct systems of molecules, important not only in adult epithelia but also in other tissues where the junctional specializations are not always so plainly visible.

Figure 19–3 illustrates schematically the types of junctions that the electron microscope reveals in a section of mature epithelium and shows how the cell–cell adhesions (anchoring junctions) that will concern us in this section are distributed in relation to other types of junctions to be discussed later. The diagram shows the typical arrangement in a *simple columnar* epithelium such as the lining of the small intestine of a vertebrate. Here, a single layer of tall cells all stand on a basal lamina, with their uppermost surface, or *apex*, free and exposed to the extracellular medium. On their sides, or *lateral* surfaces, the cells make junctions with one another. Closest to the apex lie occluding junctions (known as *tight junctions* in vertebrates), preventing molecules from leaking across the epithelium through gaps between the cells. Below these are two types of cell–cell

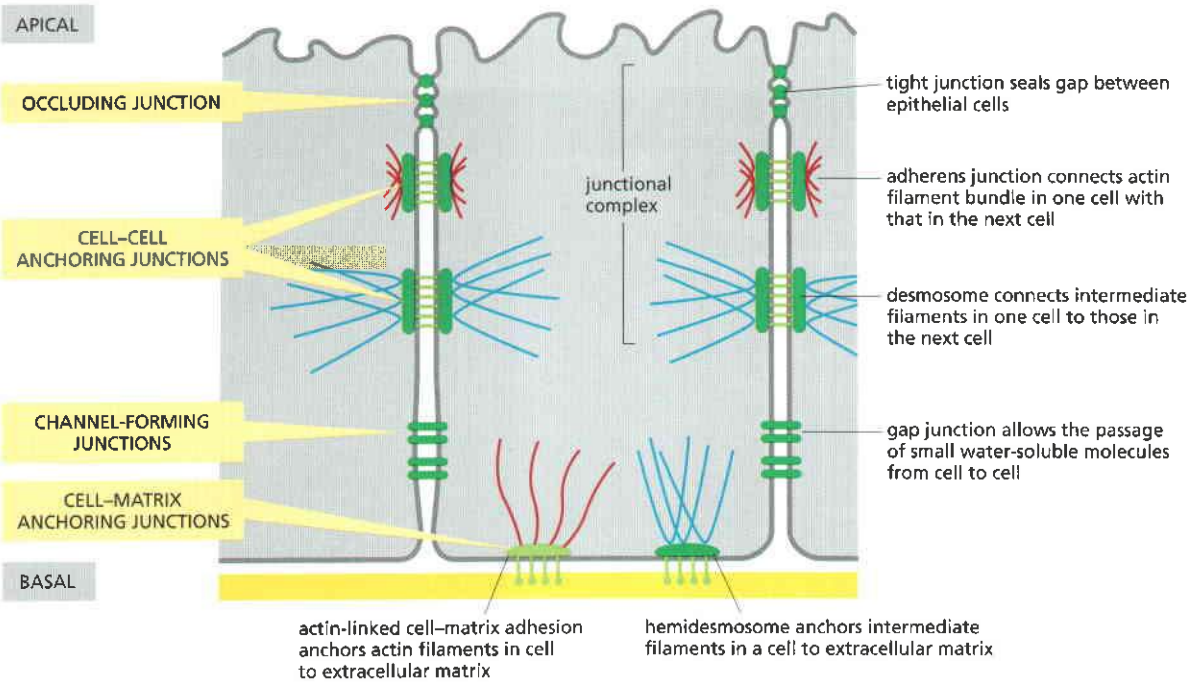


Figure 19–3 A summary of the various cell junctions found in a vertebrate epithelial cell, classified according to their primary functions. In the most apical portion of the cell, the relative positions of the junctions are the same in nearly all vertebrate epithelia. The tight junction occupies the most apical position, followed by the adherens junction (adhesion belt) and then by a special parallel row of desmosomes; together these form a structure called a junctional complex. Gap junctions and additional desmosomes are less regularly organized.The drawing is based on epithelial cells of the small intestine. Specialized signal-relaying junctions are discussed later in the chapter.

adhesions. **Adherens junctions** are anchorage sites for actin filaments; **desmosome junctions** are anchorage sites for intermediate filaments. Still lower, often mingled with additional desmosome junctions, lie channel-forming junctions, called *gap junctions.*

Additional sets of adhesions attach the epithelial cells to the basal lamina and will be discussed in a later section. We classify these cell–matrix adhesions, like the cell–cell adhesions, according to their cytoskeletal connections: *actin-linked cell–matrix adhesions* (indistinct in the small intestine, but prominent elsewhere) anchor actin filaments to the matrix, while *hemidesmosomes* anchor intermediate filaments to it.

At each of the four types of anchoring junctions, the central role is played by **transmembrane adhesion proteins** that span the membrane, with one end linking to the cytoskeleton inside the cell and the other end linking to other structures outside it (**Figure 19–4**). These cytoskeleton-linked transmembrane molecules fall neatly into two superfamilies, corresponding to the two basic kinds of external attachment (**Table 19–2**). Proteins of the **cadherin** superfamily chiefly mediate attachment of cell to cell. Proteins of the **integrin** superfamily chiefly mediate attachment of cells to matrix. Within each family, there is specialization: some cadherins link to actin and form adherens junctions, while others link to intermediate filaments and form desmosome junctions; likewise,

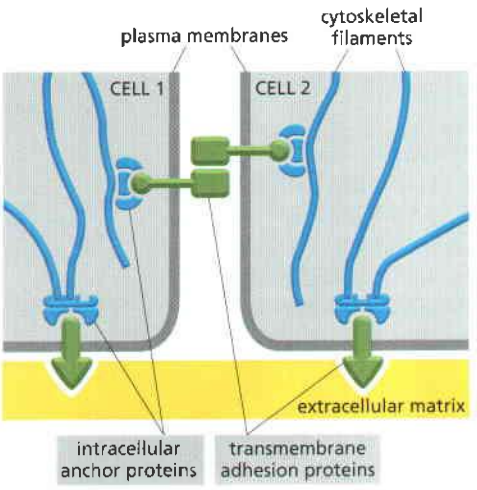


Figure 19–4 Transmembrane adhesion proteins link the cytoskeleton to extracellular structures. The external linkage may be either to parts of other cells (cell–cell anchorage, mediated typically by cadherins) or to extracellular matrix (cell–matrix anchorage, mediated typically by integrins). The internal linkage to the cytoskeleton is generally indirect, via intracellular anchor proteins, to be discussed later.

Table 19–2 Anchoring Junctions

JUNCTION	TRANSMEMBRANE ADHESION PROTEIN	EXTRACELLULAR LIGAND	INTRACELLULAR CYTOSKELETAL ATTACHMENT	INTRACELLULAR ANCHOR PROTEINS
Cell–Cell				
adherens junction	cadherin (classical cadherin)	cadherin in neighboring cell	actin filaments	α-catenin, β-catenin, plakoglobin (γ-catenin), p120-catenin, vinculin, α-actinin
desmosome	cadherin (desmoglein, desmocollin)	desmoglein and desmocollin in neighboring cell	intermediate filaments	plakoglobin (γ-catenin), plakophilin, desmoplakin
Cell–Matrix				
actin-linked cell–matrix adhesion	integrin	extracellular matrix proteins	actin filaments	talin, vinculin, α-actinin, filamin, paxillin, focal adhesion kinase (FAK)
hemidesmosome	integrin α6β4, type XVII collagen (BP180)	extracellular matrix proteins	intermediate filaments	plectin, dystonin (BP230)

some integrins link to actin and form actin-linked cell–matrix adhesions, while others link to intermediate filaments and form hemidesmosomes.

There are some exceptions to these rules. Some integrins, for example, mediate cell–cell rather than cell–matrix attachment. Moreover, there are other types of cell adhesion molecules that can provide attachments more flimsy than anchoring junctions, but sufficient to stick cells together in special circumstances. Cell–cell adhesions based on cadherins, however, seem to be the most fundamentally important class, and we begin our account of cell–cell adhesion with them. <CGAA>

Cadherins Mediate Ca^{2+}-Dependent Cell–Cell Adhesion in All Animals

Cadherins are present in all multicellular animals whose genomes have been analyzed, and in one other known group, the choanoflagellates. These creatures can exist either as free-living unicellular organisms or as multicellular colonies and are thought to be representatives of the group of protists from which all animals evolved. Other eucaryotes, including fungi and plants, lack cadherins, and they are absent from bacteria and archaea also. Cadherins therefore seem to be part of the essence of what it is to be an animal.

The cadherins take their name from their dependence on Ca^{2+} ions: removing Ca^{2+} from the extracellular medium causes adhesions mediated by cadherins to come adrift. Sometimes, especially for embryonic tissues, this is enough to let the cells be easily separated. In other cases, a more severe treatment is required, combining Ca^{2+} removal with exposure to a protease such as trypsin. The protease loosens additional connections mediated by extracellular matrix and by other cell–cell adhesion molecules that do not depend on Ca^{2+}. In either case, when the dissociated cells are put back into a normal culture medium, they will generally stick together again by reconstructing their adhesions.

This type of cell–cell association provided one of the first assays that allowed cell–cell adhesion molecules to be identified. In these experiments, monoclonal antibodies were raised against the cells of interest, and each antibody was tested for its ability to prevent the cells from sticking together again after they had been dissociated. Rare antibodies that bound to the cell–cell adhesion molecules showed this blocking effect. These antibodies then were used to isolate the adhesion molecule that they recognized.

Virtually all cells in vertebrates, and probably in other animals too, seem to express one or more proteins of the cadherin family, according to the cell type.

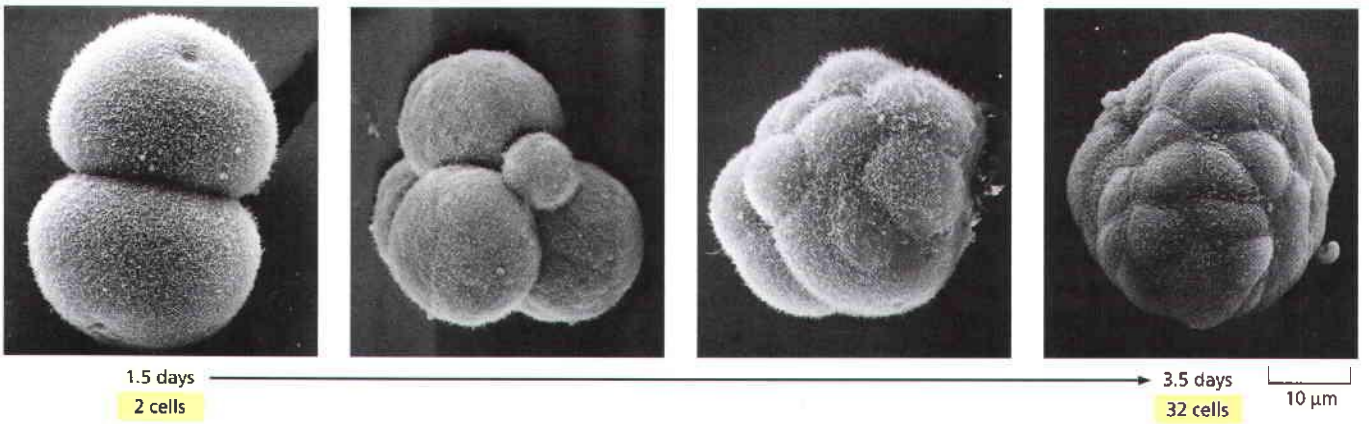


Figure 19–5 Compaction of an early mouse embryo. The cells of the early embryo at first stick together only weakly. At about the eight-cell stage, they begin to express E-cadherin and as a result become strongly and closely adherent to one another. (Scanning electron micrographs courtesy of Patricia Calarco; 16–32-cell stage is from P. Calarco and C.J. Epstein, *Dev. Biol.* 32:208–213, 1973. With permission from Academic Press.)

Several lines of evidence indicate that they are the main adhesion molecules holding cells together in early embryonic tissues. For example, embryonic tissues in culture disintegrate when treated with anti-cadherin antibodies, and if cadherin-mediated adhesion is left intact, antibodies against other adhesion molecules have little effect. Studies of the early mouse embryo illustrate the role of cadherins in development. Up to the eight-cell stage, the mouse embryo cells are only very loosely held together, remaining individually more or less spherical; then, rather suddenly, in a process called compaction, they become tightly packed together and joined by cell–cell junctions, so that the outer surface of the embryo becomes smoother (**Figure 19–5**). Antibodies against a specific cadherin, called *E-cadherin*, block compaction, whereas antibodies that react with various other cell-surface molecules on these cells do not. Mutations that inactivate E-cadherin cause the embryos to fall apart and die early in development.

The Cadherin Superfamily in Vertebrates Includes Hundreds of Different Proteins, Including Many with Signaling Functions

The first three cadherins that were discovered were named according to the main tissues in which they were found: *E-cadherin* is present on many types of epithelial cells; *N-cadherin* on nerve, muscle, and lens cells; and *P-cadherin* on cells in the placenta and epidermis. All are also found in various other tissues; N-cadherin, for example, is expressed in fibroblasts, and E-cadherin is expressed in parts of the brain (**Figure 19–6**). These and other **classical cadherins** are closely related in sequence throughout their extracellular and intracellular domains. While all of them have well-defined adhesive functions, they are also important in signaling. Through their intracellular domains, as we shall see later, they relay information into the cell interior, enabling the cell to adapt its behavior according to whether it is attached or detached from other cells.

There are also a large number of **nonclassical cadherins** more distantly related in sequence, with more than 50 expressed in the brain alone. The nonclassical cadherins include proteins with known adhesive function, such as the diverse *protocadherins* found in the brain, and the *desmocollins* and *desmogleins* that form desmosome junctions. They also include proteins that appear to be primarily involved in signaling, such as *T-cadherin*, which lacks a transmembrane domain and is attached to the plasma membrane of nerve and muscle



Figure 19–6 Cadherin diversity in the central nervous system. The diagram shows the expression patterns of three different classical cadherins in the embryonic mouse brain. More than 70 other cadherins, both classical and nonclassical, are also expressed in the brain, in complex patterns that are thought to reflect their roles in guiding and maintaining the organization of this intricate organ.

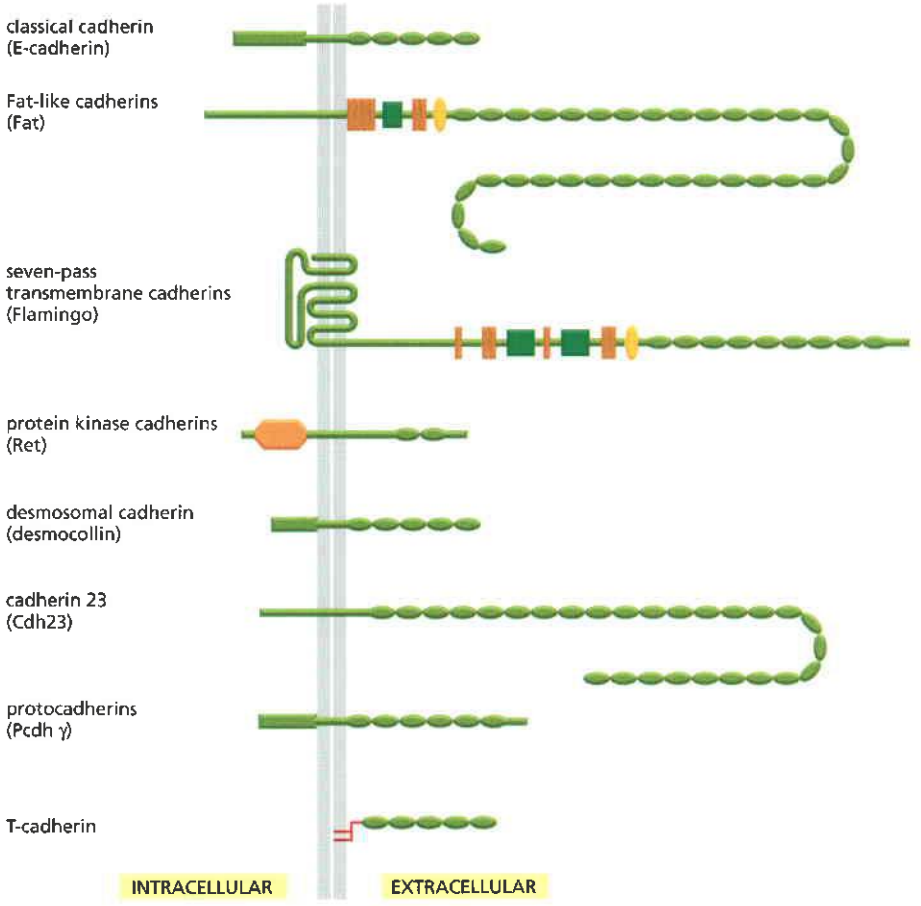


Figure 19–7 The cadherin superfamily. The diagram shows some of the diversity among cadherin superfamily members. These proteins all have extracellular portions containing multiple copies of the cadherin domain motif *(green ovals)*, but their intracellular portions are more varied, reflecting interactions with a wide variety of intracellular ligands, including signaling molecules as well as components that anchor the cadherin to the cytoskeleton. The differently colored motifs in Fat, Flamingo, and Ret represent conserved domains that are also found in other protein families.

cells by a glycosylphosphatidylinositol (GPI) anchor, and the *Fat* and *Flamingo* proteins, which were first identified as the products of genes in *Drosophila* that regulate, respectively, epithelial growth and cell polarity. Together, the classical and nonclassical cadherin proteins constitute the **cadherin superfamily** (**Figure 19–7** and **Table 19–3**), with more than 180 members in humans. How do the structures of these proteins relate to their functions, and why are there so many of them?

Cadherins Mediate Homophilic Adhesion

Anchoring junctions between cells are usually symmetrical: if the linkage is to actin, for example, in the cell on one side of the junction, it will be to actin in the cell on the other side also. In fact, the binding between cadherins is generally **homophilic** (like-to-like, **Figure 19–8**): cadherin molecules of a specific subtype on one cell bind to cadherin molecules of the same or closely related subtype on adjacent cells. According to a current model, the binding occurs at the N-terminal tips of the cadherin molecules—the ends that lie furthest from the membrane. The protein chain here forms a terminal knob and a nearby pocket, and the cadherin molecules protruding from opposite cell membranes bind by insertion of the knob of each one in the pocket of the other (**Figure 19–9A**).

The spacing between the cell membranes at an anchoring junction is precisely defined and depends on the structure of the participating cadherin molecules. All the members of the superfamily, by definition, have an extracellular portion consisting of several copies of a motif called the *cadherin domain.* In the classical cadherins of vertebrates there are 5 of these repeats, and in desmogleins and desmocollins there are 4 or 5, but some nonclassical cadherins have more than 30. Each cadherin domain forms a more or less rigid unit, joined to the next cadherin domain by a hinge (Figure 19–9B). Ca^{2+} ions bind to sites

Table 19–3 Some Members of the Cadherin Superfamily

NAME	MAIN LOCATION	JUNCTION ASSOCIATION	PHENOTYPE WHEN INACTIVATED IN MICE
Classical cadherins			
E-cadherin	many epithelia	adherens junctions	death at blastocyst stage; embryos fail to undergo compaction
N-cadherin	neurons, heart, skeletal muscle, lens, and fibroblasts	adherens junctions and chemical synapses	embryos die from heart defects
P-cadherin	placenta, epidermis, breast epithelium	adherens junctions	abnormal mammary gland development
VE-cadherin	endothelial cells	adherens junctions	abnormal vascular development (apoptosis of endothelial cells)
Nonclassical cadherins			
Desmocollin	skin	desmosomes	blistering of skin
Desmoglein	skin	desmosomes	blistering skin disease due to loss of keratinocyte cell–cell adhesion
T-cadherin	neurons, muscle, heart	none	unknown
Cadherin 23	inner ear, other epithelia	links between stereocilia in sensory hair cells	deafness
Fat (in *Drosophila*)	epithelia and central nervous system	signal-relaying junction (planar cell polarity)	enlarged imaginal discs and tumors; disrupted planar cell polarity
Fat1 (in mammals)	various epithelia and central nervous system	slit diaphragm in kidney glomerulus and other cell junctions	loss of slit diaphragm; malformation of forebrain and eye
α, β, and γ-Protocadherins	neurons	chemical synapses and nonsynaptic membranes	neuronal degeneration
Flamingo	sensory and some other epithelia	cell–cell junctions	disrupted planar cell polarity; neural tube defects

near each hinge and prevent it from flexing, so that the whole string of cadherin domains behaves as a rigid, slightly curved, rod. When Ca^{2+} is removed, the hinges can flex, and the structure becomes floppy. At the same time, the conformation at the N terminus is thought to change slightly, weakening the binding affinity for the matching cadherin molecule on the opposite cell. Cadherin molecules destabilized in this way by loss of Ca^{2+} are rapidly degraded by proteolytic enzymes.

Unlike receptors for soluble signal molecules, which bind their specific ligand with high affinity, cadherins (and most other cell–cell adhesion proteins) typically bind to their partners with relatively low affinity. Strong attachments result from the formation of many such weak bonds in parallel. When binding to oppositely oriented partners on another cell, cadherin molecules are often clustered side-to-side with many other cadherin molecules on the same cell. Many cadherin molecules packed side by side in this way collaborate to form an anchoring junction (Figure 19–9C). The strength of this junction is far greater than that of any individual intermolecular bond, and yet it can be easily disassembled by separating the molecules sequentially, just as two pieces of fabric can be strongly joined by Velcro and yet easily peeled apart. A similar "Velcro principle" also operates at cell–cell and cell–matrix adhesions formed by other

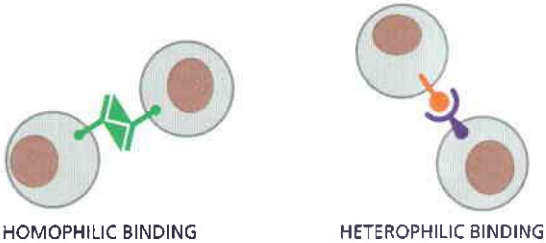


Figure 19–8 Homophilic versus heterophilic binding. Cadherins in general bind homophilically; some other cell adhesion molecules, discussed later, bind heterophilically.

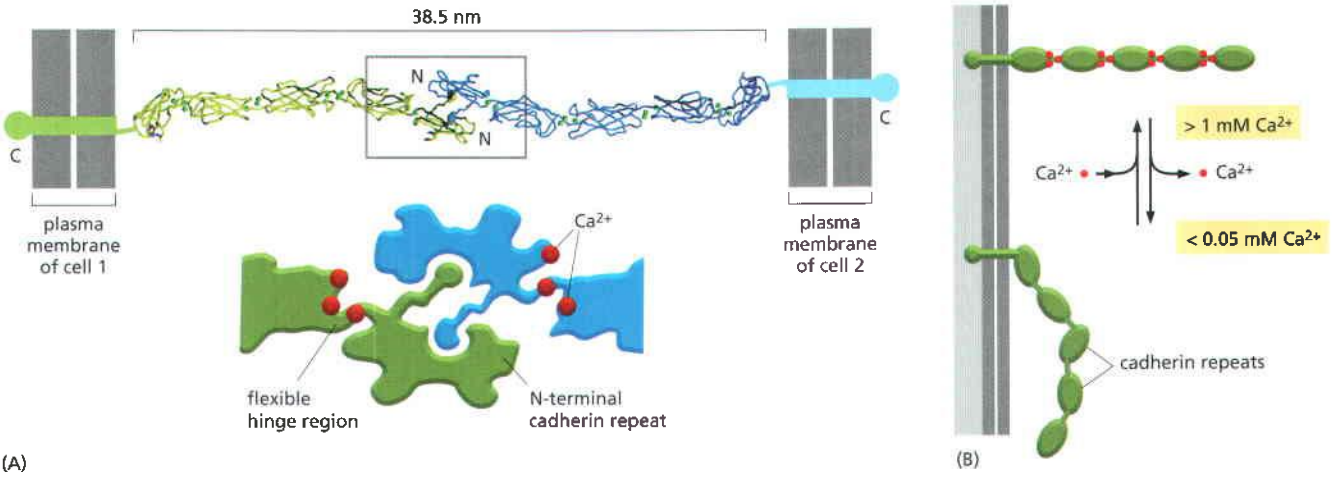


Figure 19–9 Cadherin structure and function. (A) The extracellular domain of a classical cadherin (C-cadherin) is shown here, illustrating how two such molecules on opposite cells are thought to bind homophilically, end-to-end. The structure was determined by x-ray diffraction of the crystallized C-cadherin extracellular domain. (B) The extracellular part of each polypeptide consists of a series of compact domains called cadherin repeats, joined by flexible hinge regions. Ca^{2+} binds in the neighborhood of each hinge, preventing it from flexing. In the absence of Ca^{2+}, the molecule becomes floppy and adhesion fails. (C) At a typical junction, many cadherin molecules are arrayed in parallel, functioning like Velcro to hold cells together. Cadherins on the same cell are thought to be coupled by side-to-side interactions between their N-terminal head regions, and via the attachments of their intracellular tails to a mat of other proteins (not shown here). (Based on T.J. Boggon et al., *Science* 296:1308–1313, 2002. With permission from AAAS.)

types of transmembrane adhesion proteins. The making and breaking of anchoring junctions plays a vital part in development and in the constant turnover of tissues in many parts of the mature body. <CGAA>

Selective Cell–Cell Adhesion Enables Dissociated Vertebrate Cells to Reassemble into Organized Tissues

Cadherins form specific homophilic attachments, and this explains why there are so many different family members. Cadherins are not like glue, making cell surfaces generally sticky. Rather, they mediate highly selective recognition, enabling cells of a similar type to stick together and to stay segregated from other types of cells.

This selectivity in the way that animal cells consort with one another was demonstrated more than 50 years ago, long before the discovery of cadherins, in experiments in which amphibian embryos were dissociated into single cells. These cells were then mixed up and allowed to reassociate. Remarkably, the dissociated cells often reassembled *in vitro* into structures resembling those of the original embryo (**Figure 19–10**). The same phenomenon occurs when dissociated cells from two embryonic vertebrate organs, such as the liver and the retina, are mixed together and artificially formed into a pellet: the mixed aggregates

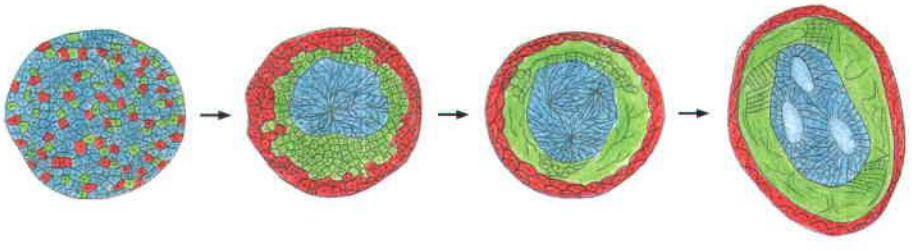


Figure 19–10 Sorting out. Cells from different parts of an early amphibian embryo will sort out according to their origins. In the classical experiment shown here, mesoderm cells (*green*), neural plate cells (*blue*), and epidermal cells (*red*) have been disaggregated and then reaggregated in a random mixture. They sort out into an arrangement reminiscent of a normal embryo, with a "neural tube" internally, epidermis externally, and mesoderm in between. (Modified from P.L. Townes and J. Holtfreter, *J. Exp. Zool.* 128:53–120, 1955. With permission from Wiley-Liss.)

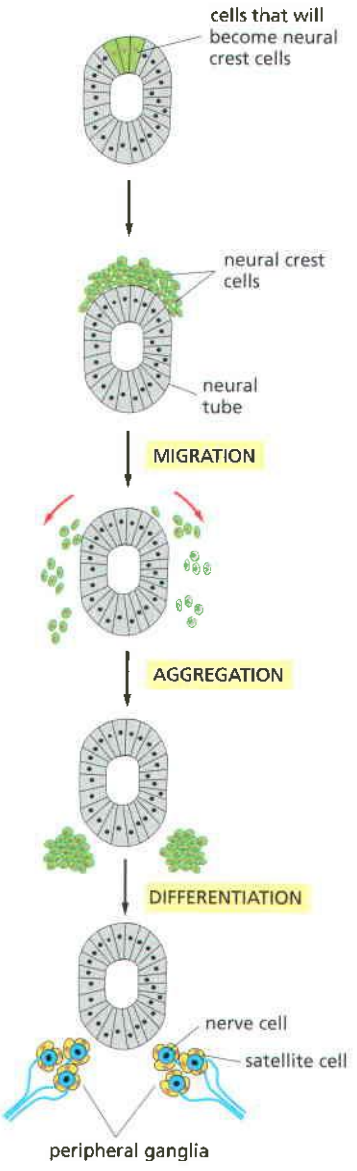


Figure 19–11 Selective dispersal and reassembly of cells to form tissues in a vertebrate embryo. Some cells that are initially part of the epithelial neural tube alter their adhesive properties and disengage from the epithelium to form the neural crest on the upper surface of the neural tube. The cells then migrate away and form a variety of cell types and tissues throughout the embryo. Here they are shown assembling and differentiating to form two clusters of nerve cells, called ganglia, in the peripheral nervous system. While some of the neural crest cells differentiate in the ganglion to become the neurons, others become satellite cells (specialized glial supporting cells) wrapped around the neurons. Changing patterns of expression of cell adhesion molecules underlie all these architectural rearrangements.

gradually sort out according to their organ of origin. More generally, disaggregated cells are found to adhere more readily to aggregates of their own organ than to aggregates of other organs. Evidently there are cell–cell recognition systems that make cells of the same differentiated tissue preferentially adhere to one another.

Such findings suggest that tissue architecture in animals is not just a product of history but is actively organized and maintained by the system of affinities that cells have for one another and for the extracellular matrix. In the developing embryo, we can indeed watch the cells as they differentiate, and see how they move and regroup to form new structures, guided by selective adhesion. Some of these movements are subtle, others more far-reaching, involving long-range migrations, as we shall describe in Chapter 22. In vertebrate embryos, for example, cells from the *neural crest* break away from the epithelial neural tube, of which they are initially a part, and migrate along specific paths to many other regions. There they reaggregate with other cells and with one another to form a variety of tissues, including those of the peripheral nervous system (**Figure 19–11**). To find their way, the cells depend on guidance from the embryonic tissues along the path. This may involve *chemotaxis* or *chemorepulsion*, that is, movement under the influence of soluble chemicals that attract or repel migrating cells. It may also involve *contact guidance*, in which the migrant cell touches other cells or extracellular matrix components, making transient adhesions that govern the track taken. Then, once the migrating cell has reached its destination, it must recognize and join other cells of the appropriate type to assemble into a tissue. In all these processes of sorting out, contact guidance, and tissue assembly, cadherins play a crucial part.

Cadherins Control the Selective Assortment of Cells

The appearance and disappearance of specific cadherins correlate with steps in embryonic development where cells regroup and change their contacts to create new tissue structures. As the neural tube forms and pinches off from the overlying ectoderm, for example, neural tube cells lose E-cadherin and acquire other cadherins, including N-cadherin, while the cells in the overlying ectoderm continue to express E-cadherin (**Figure 19–12**A, B). Then, when the neural crest cells migrate away from the neural tube, these cadherins become scarcely detectable, and another cadherin (cadherin-7) appears that helps hold the migrating cells together as loosely associated cell groups (Figure 19–12C). Finally, when the cells aggregate to form a ganglion, they switch on expression of N-cadherin again (see Figure 19–11). If N-cadherin is artificially overexpressed in the emerging neural crest cells, the cells fail to escape from the neural tube.

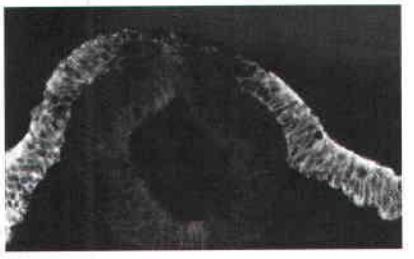
Studies with cultured cells support the suggestion that the homophilic binding of cadherins controls these processes of tissue segregation. In a line of cultured fibroblasts called *L cells*, for example, cadherins are not expressed and the cells do not adhere to one another. When these cells are transfected with DNA encoding E-cadherin, however, they become adherent to one another, and the adhesion is inhibited by anti-E-cadherin antibodies. Since the transfected cells do not stick to untransfected L cells, we can conclude that the attachment

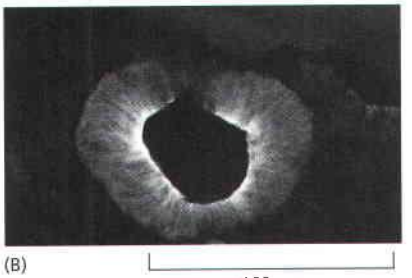
depends on E-cadherin on one cell binding to E-cadherin on another. If L cells expressing different cadherins are mixed together, they sort out and aggregate separately, indicating that different cadherins preferentially bind to their own type (**Figure 19–13**A), mimicking what happens when cells derived from tissues that express different cadherins are mixed together. A similar segregation of cells occurs if L cells expressing different amounts of the same cadherin are mixed together (Figure 19–13B). It therefore seems likely that both qualitative and quantitative differences in the expression of cadherins have a role in organizing tissues.

Twist Regulates Epithelial–Mesenchymal Transitions

The assembly of cells into an epithelium is a reversible process. By switching on expression of adhesion molecules, dispersed unattached cells—often called *mesenchymal cells*—can come together to form an epithelium. Conversely, epithelial cells can change their character, disassemble, and migrate away from their parent epithelium as separate individuals. Such *epithelial–mesenchymal transitions* play an important part in normal embryonic development; the origin of the neural crest is one example (see Figure 19–11). A control system involving a set of gene regulatory components called Slug, Snail, and Twist, with E-cadherin as a downstream component, seems to be critical for such transitions: in several tissues, both in flies and vertebrates, switching on expression of Twist, for example, converts epithelial cells to a mesenchymal character, and switching it off does the opposite.

Epithelial–mesenchymal transitions also occur as pathological events during adult life, in cancer. Most cancers originate in epithelia, but become dangerously prone to spread—that is, *malignant*—only when the cancer cells escape from the epithelium of origin and invade other tissues. Experiments with malignant breast cancer cells in culture show that blocking expression of Twist can convert them back toward a nonmalignant character. Conversely, by forcing Twist expression, one can make normal epithelial cells undergo an epithelial–mesenchymal transition and behave like malignant cells. Twist exerts its effects, in part at least, by inhibiting expression of the cadherins that hold epithelial cells together. E-cadherin, in particular, is a target. Mutations that disrupt the production or function of E-cadherin are in fact often found in cancer cells and are thought to help make them malignant, as we shall discuss in Chapter 20.





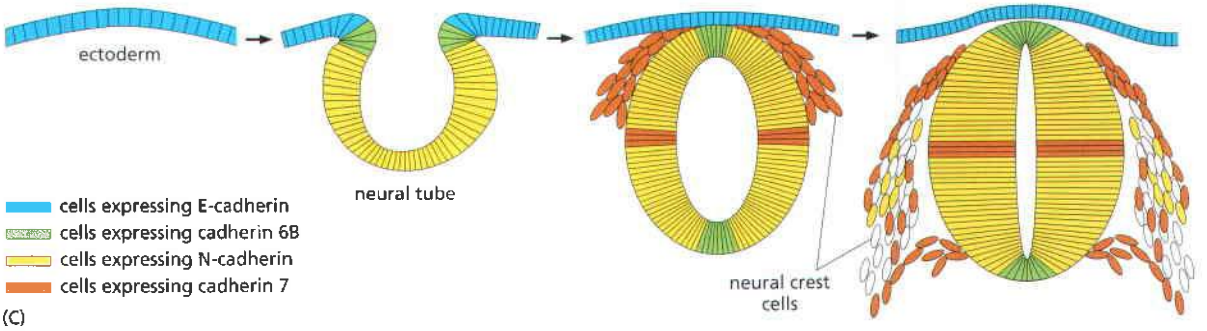


Figure 19–12 Changing patterns of cadherin expression during construction of the nervous system. The figure shows cross-sections of the early chick embryo, as the neural tube detaches from the ectoderm and then as neural crest cells detach from the neural tube. (A,B) Immunofluorescence micrographs showing the developing neural tube labeled with antibodies against (A) E-cadherin and (B) N-cadherin. (C) As the patterns of gene expression change, the different groups of cells segregate from one another according to the cadherins they express. (Micrographs courtesy of Kohei Hatta and Masatoshi Takeichi.)

Catenins Link Classical Cadherins to the Actin Cytoskeleton

The extracellular domains of cadherins mediate homophilic binding. The intracellular domains of typical cadherins, including all classical and some nonclassical ones, provide anchorage for filaments of the cytoskeleton: anchorage to actin at adherens junctions, and to intermediate filaments at desmosome junctions, as mentioned earlier (see Figure 19–3). The linkage to the cytoskeleton is indirect and depends on a cluster of accessory *intracellular anchor proteins* that assemble on the tail of the cadherin. This linkage, connecting the cadherin family member to actin or intermediate filaments, includes several different components (**Figure 19–14**). These components vary somewhat according to the type of anchorage—but in general a central part is played by *β-catenin* and/or its close relative *γ-catenin (plakoglobin)*.

At adherens junctions, a remote relative of this pair of proteins, *p120-catenin*, is also present and helps to regulate assembly of the whole complex. When p120-catenin is artificially depleted, cadherin proteins are rapidly degraded, and cell–cell adhesion is lost. An artificial increase in the level of p120-catenin has an opposite effect. It is possible that cells use changes in the level of p120-catenin or in its phosphorylation state as one way to regulate their strength of adhesion. In any case, it seems that the linkage to actin is essential for efficient cell–cell adhesion, as classical cadherins that lack their cytoplasmic domain cannot hold cells strongly together.

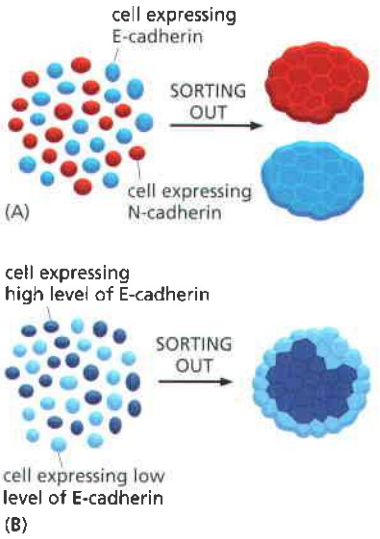


Figure 19–13 Cadherin-dependent cell sorting. Cells in culture can sort themselves out according to the type and level of cadherins they express. This can be visualized by labeling different populations of cells with dyes of different colors. (A) Cells expressing N-cadherin sort out from cells expressing E-cadherin. (B) Cells expressing high levels of E-cadherin sort out from cells expressing low levels of E-cadherin.

Adherens Junctions Coordinate the Actin-Based Motility of Adjacent Cells

Adherens junctions are an essential part of the machinery for modeling the shapes of multicellular structures in the animal body. By indirectly linking the actin filaments in one cell to those in its neighbors, they enable the cells in the tissue to use their actin cytoskeletons in a coordinated way.

Adherens junctions occur in various forms. In many nonepithelial tissues, they appear as small punctate or streaklike attachments that indirectly connect the cortical actin filaments beneath the plasma membranes of two interacting cells. In heart muscle (discussed in Chapter 23), they anchor the actin bundles of the contractile apparatus and act in parallel with desmosome junctions to link the contractile cells end-to-end. (The cell–cell interfaces in the muscle where these adhesions occur are so substantial that they show up clearly in stained light-microscope sections as so-called *intercalated discs*.) But the prototypical examples of adherens junctions occur in epithelia, where they often form a continuous **adhesion belt** (or *zonula adherens)* close beneath the apical face of the epithelium, encircling each of the interacting cells in the sheet (**Figure 19–15**). Within each cell, a contractile bundle of actin filaments lies adjacent to the adhesion belt, oriented parallel to the plasma membrane and tethered to it by the cadherins and their associated intracellular anchor proteins. The actin bundles are thus linked, via the cadherins and anchor proteins, into an extensive transcellular network. This network can contract with the help of myosin motor proteins (discussed in Chapter 16), providing the motile force for a fundamental process in animal morphogenesis—the folding of epithelial cell sheets into tubes, vesicles, and other related structures (**Figure 19–16**).

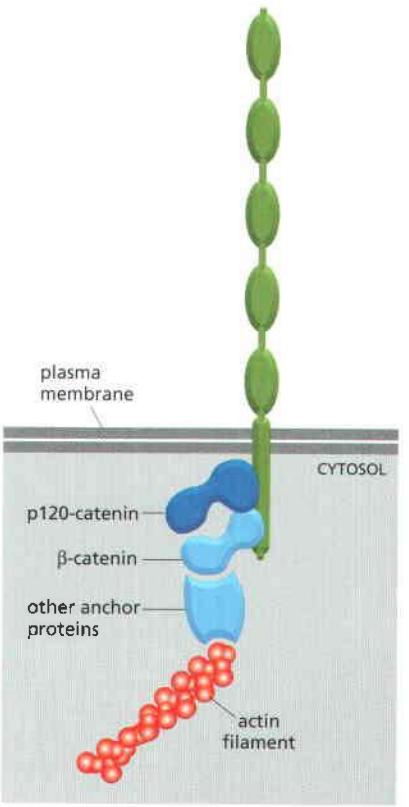


Figure 19–14 The linkage of classical cadherins to actin filaments. The cadherins are coupled indirectly to actin filaments via β-catenin and other anchor proteins. α-Catenin, vinculin, and plakoglobin (a relative of β-catenin, also called γ-catenin) are probably also present in the linkage or involved in control of its assembly, but the details of the anchorage are not well understood. Another intracellular protein, called p120-catenin, also binds to the cadherin cytoplasmic tail and regulates cadherin function. β-Catenin has a second, and very important, function in intracellular signaling, as we discuss in Chapter 15 (see Figure 15–77).

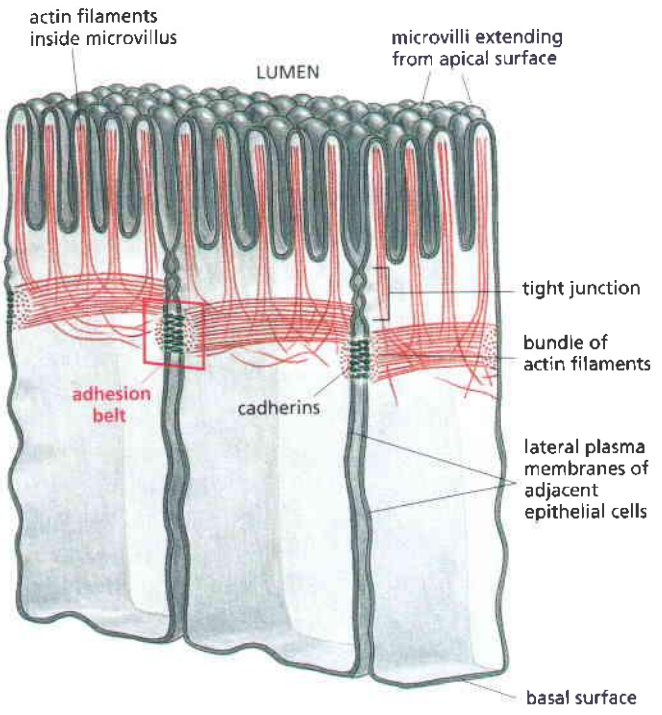


Figure 19–15 Adherens junctions between epithelial cells in the small intestine. These cells are specialized for absorption of nutrients; at their apex, facing the lumen of the gut, they have many microvilli (protrusions that serve to increase the absorptive surface area). The adherens junction takes the form of an *adhesion belt*, encircling each of the interacting cells. Its most obvious feature is a contractile bundle of actin filaments running along the cytoplasmic surface of the junctional plasma membrane. The actin filament bundles are tethered by intracellular anchor proteins to cadherins. The cadherins span the plasma membrane, and their extracellular domains bind homophilically to those of the cadherins on the adjacent cell. In this way, the actin filament bundles in adjacent cells are tied together.

Desmosome Junctions Give Epithelia Mechanical Strength

Desmosome junctions are structurally similar to adherens junctions but link to intermediate filaments instead of actin. Their main function is to provide mechanical strength. Desmosome junctions are important in vertebrates but are not found, for example, in *Drosophila*. They are present in most mature vertebrate epithelia, and are extremely plentiful in the epidermis, the epithelium that forms the outer layer of the skin; a favorite source for biochemical studies is the epidermis of the snout of cows, which has to withstand constant battering as the animal grazes.

Figure 19–17A shows the general structure of a desmosome, and Figure 19–17B shows some of the proteins that form it. Desmosomes typically appear as buttonlike spots of intercellular adhesion, riveting the cells together (Figure 19–17C). Inside the cell, the bundles of ropelike intermediate filaments that are

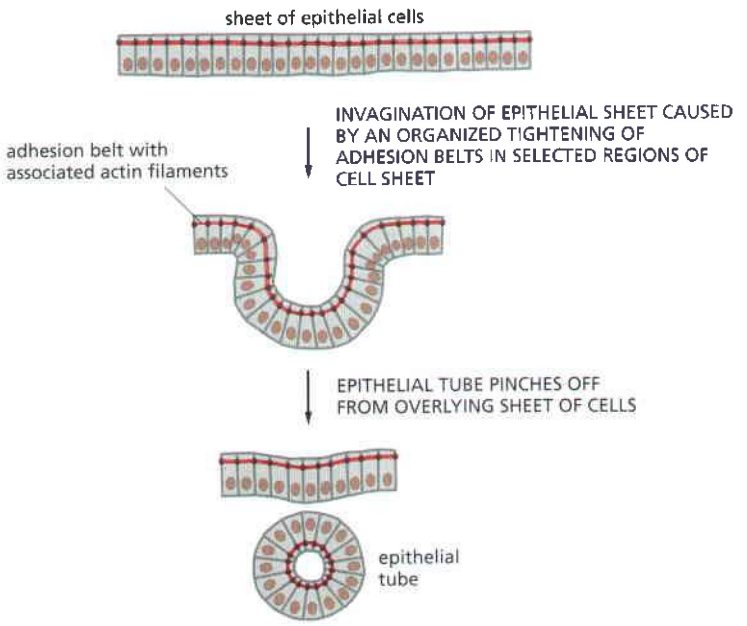


Figure 19–16 The folding of an epithelial sheet to form an epithelial tube. The oriented contraction of the bundles of actin filaments running along adhesion belts causes the epithelial cells to narrow at their apex and helps the epithelial sheet to roll up into a tube. An example is the formation of the neural tube in early vertebrate development (see Figure 19–12 and Chapter 22). Although not shown here, rearrangements of the cells within the epithelial sheet are also thought to have an important role in the process.

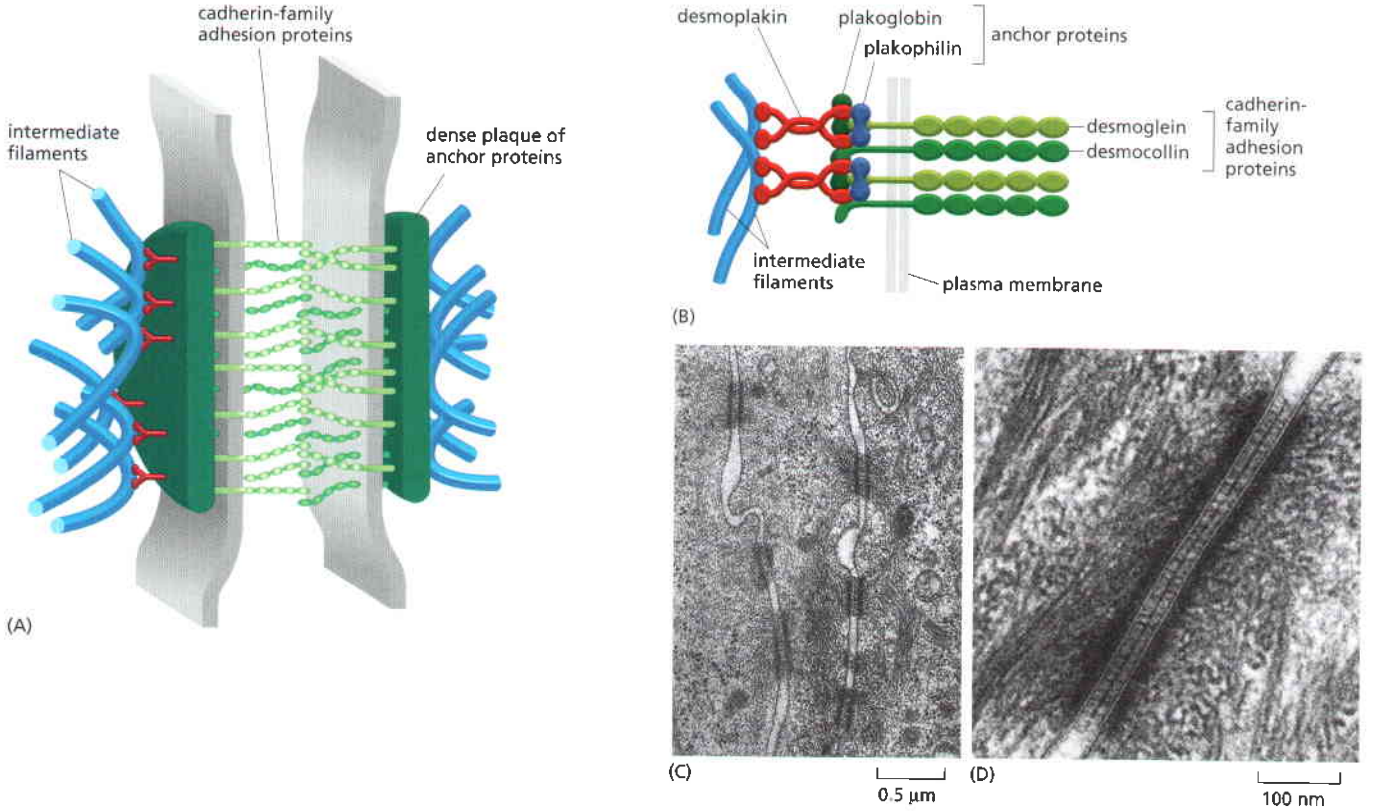


Figure 19–17 Desmosomes. (A) The structural components of a desmosome. On the cytoplasmic surface of each interacting plasma membrane is a dense plaque composed of a mixture of intracellular anchor proteins. A bundle of keratin intermediate filaments is attached to the surface of each plaque. Transmembrane adhesion proteins of the cadherin family bind to the plaques and interact through their extracellular domains to hold the adjacent membranes together by a Ca^{2+}-dependent mechanism. (B) Some of the molecular components of a desmosome. Desmoglein and desmocollin are members of the cadherin family of adhesion proteins. Their cytoplasmic tails bind *plakoglobin* (γ-catenin) and *plakophilin* (a distant relative of p120-catenin), which in turn bind to *desmoplakin*. Desmoplakin binds to the sides of intermediate filaments, thereby tying the desmosome to these filaments. (C) An electron micrograph of desmosome junctions between epidermal cells in the skin of a baby mouse. (D) Part of the same tissue at higher magnification, showing a single desmosome, with intermediate filaments attached to it. (C and D, from W. He, P. Cowin and D.L. Stokes, *Science* 302:109–113, 2003. With permission from AAAS.)

anchored to the desmosomes form a structural framework of great tensile strength (Figure 19–17D), with linkage to similar bundles in adjacent cells, creating a network that extends throughout the tissue (**Figure 19–18**). The particular type of intermediate filaments attached to the desmosomes depends on the cell type: they are *keratin filaments* in most epithelial cells, for example, and *desmin filaments* in heart muscle cells.

The importance of desmosome junctions is demonstrated by some forms of the potentially fatal skin disease *pemphigus*. Affected individuals make antibodies against one of their own desmosomal cadherin proteins. These antibodies bind to and disrupt the desmosomes that hold their epidermal cells (keratinocytes) together. This results in a severe blistering of the skin, with leakage of body fluids into the loosened epithelium.

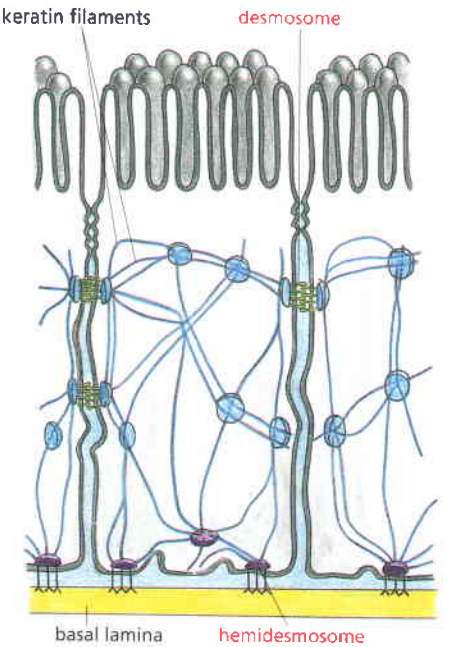


Figure 19–18 Desmosomes, hemidesmosomes, and the intermediate filament network. The keratin intermediate filament networks of adjacent cells—in this example, epithelial cells of the small intestine—are indirectly connected to one another through desmosomes, and to the basal lamina through hemidesmosomes.

Cell–Cell Junctions Send Signals into the Cell Interior

The making and breaking of attachments are important events in the lives of cells and provoke large changes in their internal affairs. Conversely, changes in the internal state of a cell must be able to trigger the making or breaking of attachments. Thus there is a complex cross-talk between the adhesion machinery and chemical signaling pathways. We have described, for example, how changes in p120-catenin may regulate the formation of adherens junctions, and several intracellular signaling pathways can control junction formation by phosphorylating this and other junctional proteins. Later, we shall discuss how the making and breaking of adhesions can send signals into the cell interior through mechanisms involving *scaffold proteins* on the intracellular side of the junction.

Another of the central players in the two-way interaction between adhesion and signaling is thought to be β-catenin. In this chapter, we have mentioned it as an essential intracellular anchor protein at adherens junctions, linking cadherins to actin filaments. In Chapter 15, we encountered it in another guise, as a component of the Wnt cell–cell signaling pathway, moving from the cytoplasm to the nucleus to activate the transcription of target genes. Separate parts of the molecule are responsible for the adhesive and gene-regulatory functions, but an individual molecule cannot do both things at once. Disintegration of an adherens junction can set β-catenin molecules free to move from the cell surface into the nucleus as signaling molecules, and, conversely, the activities of components of the Wnt signaling pathway (which regulate the phosphorylation and degradation of β-catenin) may control the availability of β-catenin to form adherens junctions.

Some nonclassical cadherins transmit signals into the cell interior in yet other ways. Members of the Flamingo subfamily, for example, have a seven-pass transmembrane domain suggesting that they might function as G-protein-coupled receptors. Vascular endothelial cadherin (VE-cadherin) provides another example. This protein not only mediates adhesion between endothelial cells but also is required for endothelial cell survival. Although endothelial cells that do not express VE-cadherin still adhere to one another via N-cadherin, they fail to survive, because they are unable to respond to an extracellular protein called *vascular endothelial growth factor (VEGF)* that acts as a survival signal. VEGF binds to a receptor tyrosine kinase (discussed in Chapter 15) that requires VE-cadherin as a co-receptor.

Selectins Mediate Transient Cell–Cell Adhesions in the Bloodstream

The cadherin superfamily is central to cell–cell adhesion in animals, but at least three other superfamilies of cell–cell adhesion proteins are also important: the *integrins*, the *selectins*, and the adhesive *immunoglobulin (Ig)-superfamily* members. We shall discuss integrins in more detail later: their main function is in cell–matrix adhesion, but a few of them mediate cell–cell adhesion in specialized circumstances. Ca^{2+} dependence provides one simple way to distinguish among these classes of proteins experimentally. Selectins, like cadherins and integrins, require Ca^{2+} for their adhesive function; Ig-superfamily members do not.

Selectins are cell-surface carbohydrate-binding proteins *(lectins)* that mediate a variety of transient, cell–cell adhesion interactions in the bloodstream. Their main role, in vertebrates at least, is in inflammatory responses and in governing the traffic of white blood cells. White blood cells lead a nomadic life, roving between the bloodstream and the tissues, and this necessitates special adhesive behavior. The selectins control the binding of white blood cells to the endothelial cells lining blood vessels, thereby enabling the blood cells to migrate out of the bloodstream into a tissue.

Each selectin is a transmembrane protein with a conserved lectin domain that binds to a specific oligosaccharide on another cell (**Figure 19–19A**). There are at least three types: *L-selectin* on white blood cells, *P-selectin* on blood platelets and on endothelial cells that have been locally activated by an inflammatory

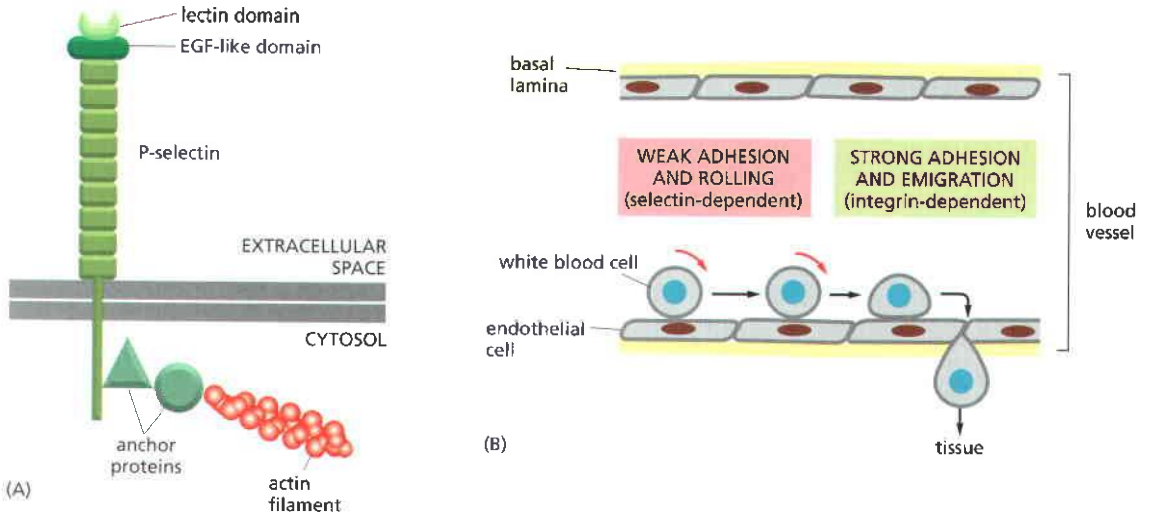


Figure 19–19 The structure and function of selectins. (A) The structure of P-selectin. The selectin attaches to the actin cytoskeleton through anchor proteins that are still poorly characterized. (B) How selectins and integrins mediate the cell–cell adhesions required for a white blood cell to migrate out of the bloodstream into a tissue. First, selectins on endothelial cells bind to oligosaccharides on the white blood cell, so that it becomes loosely attached to the vessel wall. Then the white blood cell activates an integrin (usually one called LFA1) in its plasma membrane, enabling this integrin to bind to a protein called ICAM1, belonging to the immunoglobulin superfamily, in the membrane of the endothelial cell. This creates a stronger attachment that allows the white blood cell to crawl out of the vessel.

response, and *E-selectin* on activated endothelial cells. In a lymphoid organ, such as a lymph node or a tonsil, the endothelial cells express oligosaccharides that are recognized by L-selectin on lymphocytes, causing the lymphocytes to loiter and become trapped. At sites of inflammation, the roles are reversed: the endothelial cells switch on expression of selectins that recognize the oligosaccharides on white blood cells and platelets, flagging the cells down to help deal with the local emergency. Selectins do not act alone, however; they collaborate with integrins, which strengthen the binding of the blood cells to the endothelium. The cell–cell adhesions mediated by both selectins and integrins are *heterophilic*—that is, the binding is to a molecule of a different type: selectins bind to specific oligosaccharides on glycoproteins and glycolipids, while integrins bind to other specific proteins.

Selectins and integrins act in sequence to let white blood cells leave the bloodstream and enter tissues (Figure 19–19B). The selectins mediate a weak adhesion because the binding of the lectin domain of the selectin to its carbohydrate ligand is of low affinity. This allows the white blood cell to adhere weakly and reversibly to the endothelium, rolling along the surface of the blood vessel, propelled by the flow of blood. The rolling continues until the blood cell activates its integrins. As we discuss later, these transmembrane molecules can be switched into an adhesive conformation that enables them to latch onto other molecules external to the cell—in the present case, proteins on the surfaces of the endothelial cells. Once it has attached in this way, the white blood cell escapes from the blood stream into the tissue by crawling out of the blood vessel between adjacent endothelial cells.

Members of the Immunoglobulin Superfamily of Proteins Mediate Ca^{2+}-Independent Cell–Cell Adhesion

The chief endothelial cell proteins that are recognized by the white blood cell integrins are called *ICAMs (intercellular cell adhesion molecules)* or *VCAMs (vascular cell adhesion molecules)*. They are members of another large and ancient family of cell surface molecules—the **immunoglobulin (Ig) superfamily**. These contain one or more of the extracellular Ig-like domains that are characteristic of antibody molecules (discussed in Chapter 25). They have many functions outside the immune system that are unrelated to immune defenses.

While ICAMs and VCAMs on endothelial cells both mediate heterophilic binding to integrins, many other Ig superfamily members appear to mediate homophilic binding. An example is the *neural cell adhesion molecule (NCAM)*, which is expressed by various cell types, including most nerve cells, and can take different forms, generated by alternative splicing of an RNA transcript produced from a single gene (**Figure 19–20**). Some forms of NCAM carry an unusually large quantity of sialic acid (with chains containing hundreds of repeating sialic

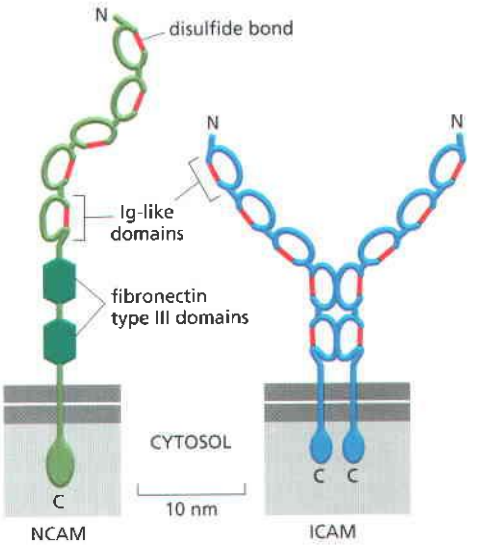


Figure 19–20 Two members of the Ig superfamily of cell–cell adhesion molecules. NCAM is expressed on neurons and many other cell types, and mediates homophilic binding. Only the protein backbone of NCAM is shown here; this often has side chains of sialic acid (a polysaccharide) covalently attached to it, hindering adhesion. ICAM is expressed on endothelial cells and some other cell types and binds heterophilically to an integrin on white blood cells.

acid units). By virtue of their negative charge, the long polysialic acid chains can interfere with cell adhesion (because like charges repel one another); NCAM heavily loaded with sialic acid may even serve to inhibit adhesion, rather than cause it.

A cell of a given type generally uses an assortment of different adhesion proteins to interact with other cells, just as each cell uses an assortment of different receptors to respond to the many soluble extracellular signal molecules, such as hormones and growth factors, in its environment. Although cadherins and Ig family members are frequently expressed on the same cells, the adhesions mediated by cadherins are much stronger, and they are largely responsible for holding cells together, segregating cell collectives into discrete tissues, and maintaining tissue integrity. Molecules such as NCAM seem to contribute more to the fine-tuning of these adhesive interactions during development and regeneration, playing a part in various specialized adhesive phenomena, such as that discussed for blood and endothelial cells. Thus, while mutant mice that lack N-cadherin die early in development, those that lack NCAM develop relatively normally but show some mild abnormalities in the development of certain specific tissues, including parts of the nervous system.

Many Types of Cell Adhesion Molecules Act in Parallel to Create a Synapse

Cells of the nervous system, especially, rely on complex systems of adhesion molecules, as well as chemotaxis and soluble signal factors, to guide axon outgrowth along precise pathways and to direct the formation of specific nerve connections (discussed in Chapter 22). Adhesion proteins of the Ig superfamily, along with many other classes of adhesion and signaling molecules, have important roles in these processes. Thus, for example, in flies with a mutation of *Fasciclin2*, related to NCAM, some axons follow aberrant pathways and fail to reach their proper targets.

Another member of the Ig superfamily, *Fasciclin3*, enables the neuronal growth cones to recognize their proper targets when they meet them. This protein is expressed transiently on some motor neurons in *Drosophila*, as well as on the muscle cells they normally innervate. If Fasciclin3 is genetically removed from these motor neurons, they fail to recognize their muscle targets and do not make synapses with them. Conversely, if motor neurons that normally do not express Fasciclin3 are made to express this protein, they will synapse with Fasciclin3-expressing muscle cells to which they normally do not connect. It seems that Fasciclin3 mediates these synaptic connections by a homophilic "matchmaking" mechanism. Ig superfamily proteins have similar roles in vertebrates. Proteins of the *Sidekicks* subfamily, for example, mediate homophilic adhesion, and different Sidekicks proteins are expressed in different layers of the retina, with synapses forming between sets of retinal neurons that share expression of the same family member. When the pattern of expression of the proteins is artificially altered, the pattern of synaptic connections changes accordingly.

These Ig superfamily members are by no means the only adhesion molecules involved in initiating synapse formation. Misexpression of certain other synaptic adhesion proteins, unrelated to any of the types we have mentioned so far, can even trick growth cones into synapsing on non-neuronal cells that would never normally be innervated. Thus, if non-neuronal cells are forced to express *neuroligin*, a transmembrane protein evolutionarily related to the

enzyme acetylcholinesterase, neurons will synapse on them, as a consequence of binding of neuroligin to a protein called *neurexin* in the membrane of the presynaptic neuron.

Scaffold Proteins Organize Junctional Complexes

To make a synapse, the pre- and postsynaptic cells have to do more than recognize one another and adhere: they have to assemble a complex system of signal receptors, ion channels, synaptic vesicles, docking proteins, and other components, as described in Chapter 11. This apparatus for synaptic signaling could not exist without cell adhesion molecules to join the pre- and postsynaptic membranes firmly together and to help hold all the components of the signaling machinery in their proper positions. Thus, cadherins are generally present, concentrated at spots around the periphery of the synapse and within it, as well as Ig superfamily members and various other types of adhesion molecules. In fact, about 20 different classical cadherins are expressed in the vertebrate nervous system, in different combinations in different subsets of neurons, and it is likely that selective binding of these molecules also plays a part in ensuring that neurons synapse with their correct partners.

But how does the array of adhesion molecules recruit the other components of the synapse and hold them in place? **Scaffold proteins** are thought to have a central role here. These intracellular molecules consist of strings of protein-binding domains, typically including several **PDZ domains**—segments about 70 amino acids long that can recognize and bind the C-terminal intracellular tails of specific transmembrane molecules (**Figure 19–21**). One domain of a scaffold protein may attach to a cell–cell adhesion protein, for example, while another latches onto a ligand-gated ion channel, and yet another binds a protein that regulates exocytosis or endocytosis or provides attachment to the cytoskeleton. Moreover, one molecule of scaffold protein can bind to another. In this way, the cell can assemble a mat of proteins, with all the components that are needed at the synapse woven into its fabric (**Figure 19–22**). Several hundred different types of proteins participate in this complex structure. Mutations in synaptic scaffold proteins alter the size and structure of synapses and can have severe consequences for the function of the nervous system. Among other things, such mutations can damage the molecular machinery underlying learning and memory, which depend on the ability of electrical activity to leave a long-lasting trace in the form of alterations of synaptic architecture.

The scaffold proteins, with their many potential binding partners, are involved in organizing other structures and functions beside synapses and synaptic signaling. The *Discs large (Dlg)* protein of *Drosophila* is an example (see Figure 19–21). Dlg is needed for the construction of normal synapses; but we shall see that it, along with a set of other related scaffold proteins, also plays an

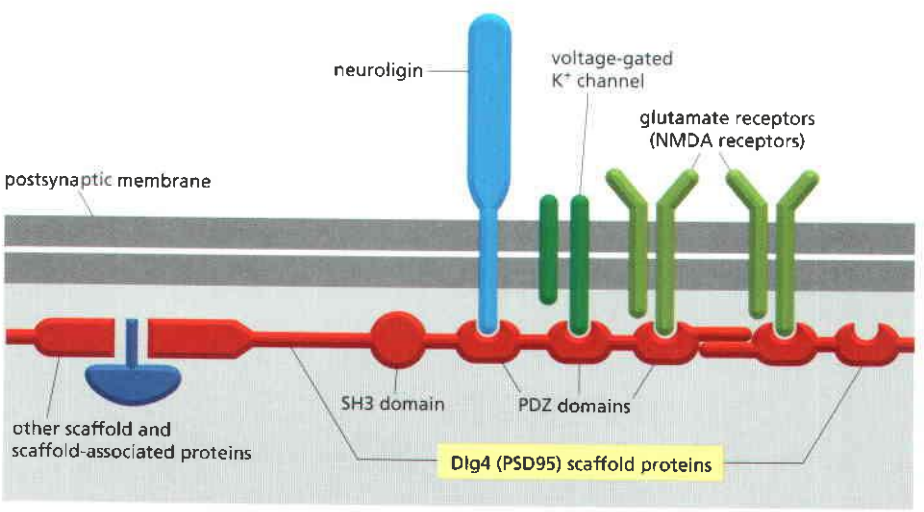
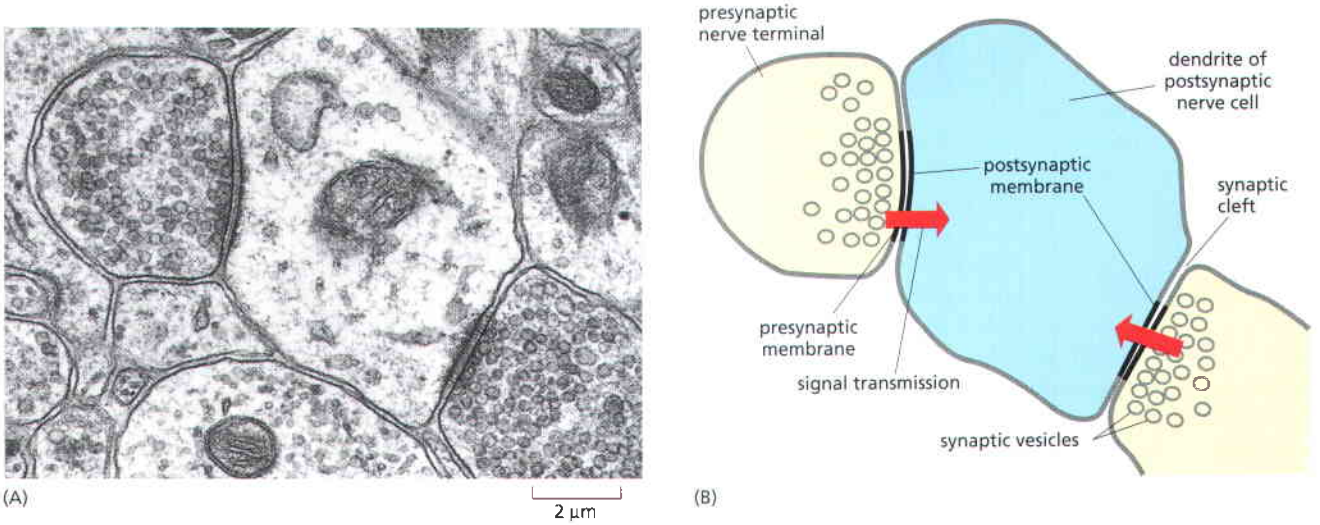


Figure 19–21 A scaffold protein. The diagram shows the domain structure of Dlg4, a mammalian homolog of the *Drosophila* protein Discs-large, along with some of its binding partners. Dlg4 is concentrated beneath the postsynaptic membrane at synapses, and is also known as postsynaptic density protein 95, or PSD95. With its multiple protein-binding domains, it can link together different components of the synapse. One molecule of Dlg4 can also bind to another or to scaffolding molecules of other types, thereby creating an extensive framework that holds together all the components of the synapse. Scaffold proteins also have important roles at other types of cell junctions.



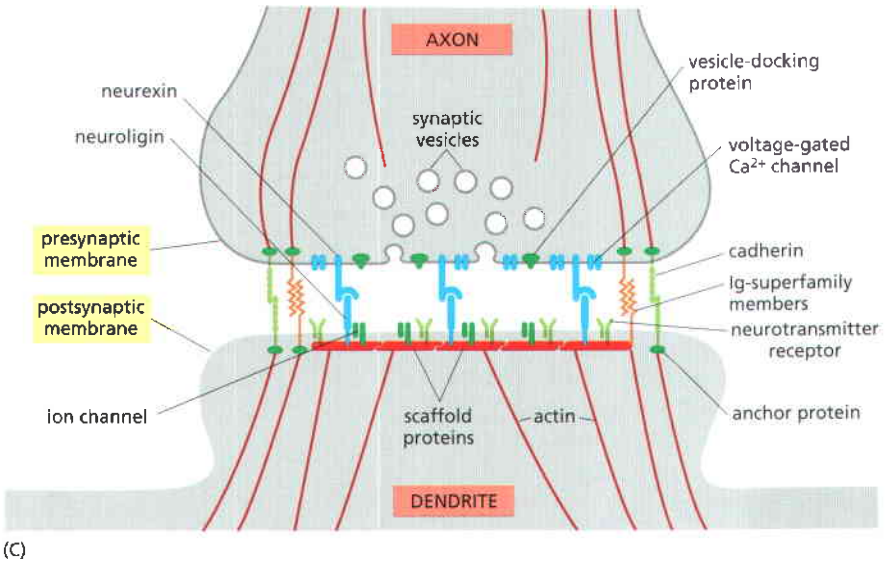


Figure 19–22 Organization of a synapse. (A) Electron micrograph and (B) line drawing of a cross section of two nerve terminals synapsing on a dendrite in the mammalian brain. Note the synaptic vesicles in the two nerve terminals and the dark-staining material associated with the pre- and postsynaptic membranes. (C) Schematic diagram showing some of the synaptic components that are assembled at these sites. Cell–cell adhesion molecules, including cadherins and neuroligins and neurexins, hold the pre- and postsynaptic membranes together. Scaffolding proteins help to form a mat (corresponding to the dark-staining material seen in (A)) that links the adhesion molecules by their intracellular tails to the components of the synaptic signal-transmission machinery, such as ion channels and neurotransmitter receptors. The structure of this large, complex multiprotein assembly is not yet known in detail. It includes anchorage sites for hundreds of additional components, not shown here, including cytoskeletal molecules and various regulatory kinases and phosphatases. (A, courtesy of Cedric Raine.)

essential part in almost every aspect of the organization of epithelia, including the formation of occluding junctions between the cells, the control of cell polarity, and even the control of cell proliferation. All these processes have a shared dependence on the same machinery, not only in flies, but also in vertebrates.

Summary

In epithelia, as well as in some other types of tissue, cells are directly attached to one another through strong cell–cell adhesions, mediated by transmembrane proteins that are anchored intracellularly to the cytoskeleton. At adherens junctions, the anchorage is to actin filaments; at desmosome junctions, it is to intermediate filaments. In both these structures, and in many less conspicuous cell–cell junctions, the adhesive transmembrane proteins are members of the cadherin superfamily. Cadherins generally bind to one another homophilically: the head of one cadherin molecule binds to the head of a similar cadherin on an opposite cell. This selectivity enables mixed populations of cells of different types to sort out from one another according to the specific cadherins they express, and it helps to control cell rearrangements during development, where many different cadherins are expressed in complex, changing patterns. Changes in cadherin expression can cause cells to undergo transitions between a cohesive epithelial state and a detached mesenchymal state—a phenomenon important in cancer as well as in embryonic development.

The "classical" cadherins are linked to the actin cytoskeleton by intracellular proteins called catenins. These form an anchoring complex on the intracellular tail of the

cadherin molecule, and are involved not only in physical anchorage but also in the genesis of intracellular signals. Conversely, intracellular signals can regulate the formation of cadherin-mediated adhesions. β-Catenin, for example, is also a key component of the Wnt cell signaling pathway.

In addition to cadherins, at least three other classes of transmembrane molecules are also important mediators of cell–cell adhesion: selectins, immunoglobulin (Ig) superfamily members, and integrins. Selectins are expressed on white blood cells, blood platelets, and endothelial cells, and they bind heterophilically to carbohydrate groups on cell surfaces. They help to trap circulating white blood cells at sites of inflammation. Ig-superfamily proteins also play a part in this trapping, as well as in many other adhesive processes; some of them bind homophilically, some heterophilically. Integrins, though they mainly serve to attach cells to the extracellular matrix, can also mediate cell–cell adhesion by binding to the Ig-superfamily members.

Many different Ig-superfamily members, cadherins, and other cell–cell adhesion molecules guide the formation of nerve connections and hold neuronal membranes together at synapses. In these complicated structures, as well as at other types of cell–cell junctions, intracellular scaffold proteins containing multiple PDZ protein-binding domains have an important role in holding the many different adhesive and signaling molecules in their proper arrangements.

TIGHT JUNCTIONS AND THE ORGANIZATION OF EPITHELIA

An epithelial sheet, with its cells joined side by side and standing on a basal lamina, may seem a specialized type of structure, but it is central to the construction of multicellular animals. In fact, more than 60% of the cell types in the vertebrate body are epithelial. Just as cell membranes enclose and partition the interior of the eucaryotic cell, so epithelia enclose and partition the animal body, lining all its surfaces and cavities, and creating internal compartments where specialized processes occur. The epithelial sheet seems to be one of the inventions that lie at the origin of animal evolution, diversifying in a huge variety of ways (as we see in Chapter 23), but retaining an organization based on a set of conserved molecular mechanisms that practically all epithelia have in common.

Essentially all epithelia are anchored to other tissue on one side—the **basal** side—and free of such attachment on their opposite side—the **apical** side. A basal lamina lies at the interface with the underlying tissue, mediating the attachment, while the apical surface of the epithelium is generally bathed by extracellular fluid (but sometimes covered by material that the cells have secreted at their apices). Thus all epithelia are structurally **polarized**, and so are their individual cells: the basal end of a cell, adherent to the basal lamina below, differs from the apical end, exposed to the medium above.

Correspondingly, all epithelia have at least one function in common: they serve as selective permeability barriers, separating the fluid that permeates the tissue on their basal side from fluid with a different chemical composition on their apical side. This barrier function requires that the adjacent cells be sealed together by **occluding junctions**, so that molecules cannot leak freely across the cell sheet. In this section we consider how the occluding junctions are formed, and how the polarized architecture of the epithelium is maintained. These two fundamental aspects of epithelia are closely linked: the junctions play a key part in organizing and maintaining the polarity of the cells in the sheet.

Tight Junctions Form a Seal Between Cells and a Fence Between Membrane Domains

The occluding junctions found in vertebrate epithelia are called **tight junctions**. The epithelium of the small intestine provides a good illustration of their structure and function (see Figure 19–3). This epithelium has a *simple columnar* structure; that is, it consists of a single layer of tall (columnar) cells. These are of

several differentiated types, but the majority are absorptive cells, specialized for uptake of nutrients from the internal cavity, or *lumen*, of the gut.

The absorptive cells have to transport selected nutrients across the epithelium from the lumen into the extracellular fluid that permeates the connective tissue on the other side. From there, these nutrients diffuse into small blood vessels to provide nourishment to the organism. This *transcellular transport* depends on two sets of transport proteins in the plasma membrane of the absorptive cell. One set is confined to the apical surface of the cell (facing the lumen) and actively transports selected molecules into the cell from the gut. The other set is confined to the *basolateral* (basal and lateral) surfaces of the cell, and it allows the same molecules to leave the cell by facilitated diffusion into the extracellular fluid on the other side of the epithelium. For this transport activity to be effective, the spaces between the epithelial cells must be tightly sealed, so that the transported molecules cannot leak back into the gut lumen through these spaces (**Figure 19–23**). Moreover, the proteins that form the pumps and channels must be correctly distributed in the cell membranes: the apical set of active transport proteins must be delivered to the cell apex (as discussed in Chapter 13) and must not be allowed to drift to the basolateral surface, and the basolateral set of channel proteins must be delivered to the basolateral surface and must not be allowed to drift to the apical surface. The tight junctions between epithelial cells, besides sealing the gaps between the cells, may also function as "fences" helping to separate domains within the plasma membrane of each cell, so as to hinder apical proteins (and lipids) from diffusing into the basal region, and vice versa (see Figure 19–23).

The sealing function of tight junctions is easy to demonstrate experimentally: a low-molecular-weight tracer added to one side of an epithelium will generally not pass beyond the tight junction (**Figure 19–24**). This seal is not absolute, however. Although all tight junctions are impermeable to macromolecules, their permeability to small molecules varies. Tight junctions in the epithelium lining the small intestine, for example, are 10,000 times more permeable to inorganic ions,

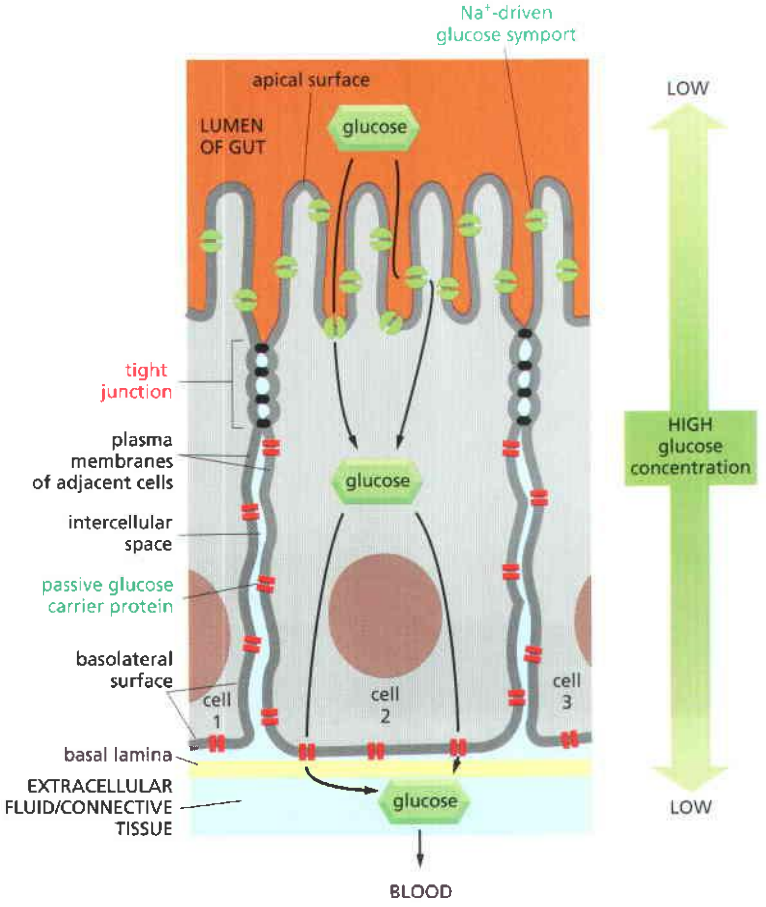


Figure 19–23 The role of tight junctions in transcellular transport. Transport proteins are confined to different regions of the plasma membrane in epithelial cells of the small intestine. This segregation permits a vectorial transfer of nutrients across the epithelium from the gut lumen to the blood. In the example shown, glucose is actively transported into the cell by Na^+-driven glucose symports at its apical surface, and it diffuses out of the cell by facilitated diffusion mediated by glucose carriers in its basolateral membrane. Tight junctions are thought to confine the transport proteins to their appropriate membrane domains by acting as diffusion barriers or "fences" within the lipid bilayer of the plasma membrane; these junctions also block the backflow of glucose from the basal side of the epithelium into the gut lumen.

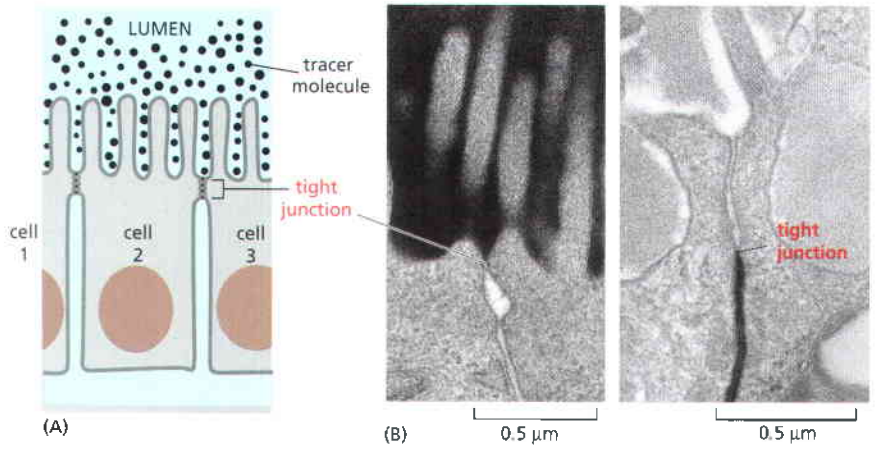


Figure 19–24 The role of tight junctions in allowing epithelia to serve as barriers to solute diffusion. (A) The drawing shows how a small extracellular tracer molecule added on one side of an epithelium is prevented from crossing the epithelium by the tight junctions that seal adjacent cells together. (B) Electron micrographs of cells in an epithelium in which a small, extracellular, electron-dense tracer molecule has been added to either the apical side (on the *left*) or the basolateral side (on the *right*). In both cases, the tight junction blocks passage of the tracer. (B, courtesy of Daniel Friend.)

such as Na^+, than the tight junctions in the epithelium lining the urinary bladder. These differences reflect differences in the proteins that form the junctions.

Epithelial cells can also alter their tight junctions transiently to permit an increased flow of solutes and water through breaches in the junctional barriers. Such *paracellular transport* is especially important in the absorption of amino acids and monosaccharides from the lumen of the intestine, where the concentration of these nutrients can increase enough after a meal to drive passive transport in the proper direction.

When tight junctions are visualized by freeze-fracture electron microscopy, they seem to consist of a branching network of *sealing strands* that completely encircles the apical end of each cell in the epithelial sheet (**Figure 19–25**A and B). In conventional electron micrographs, the outer leaflets of the two interacting

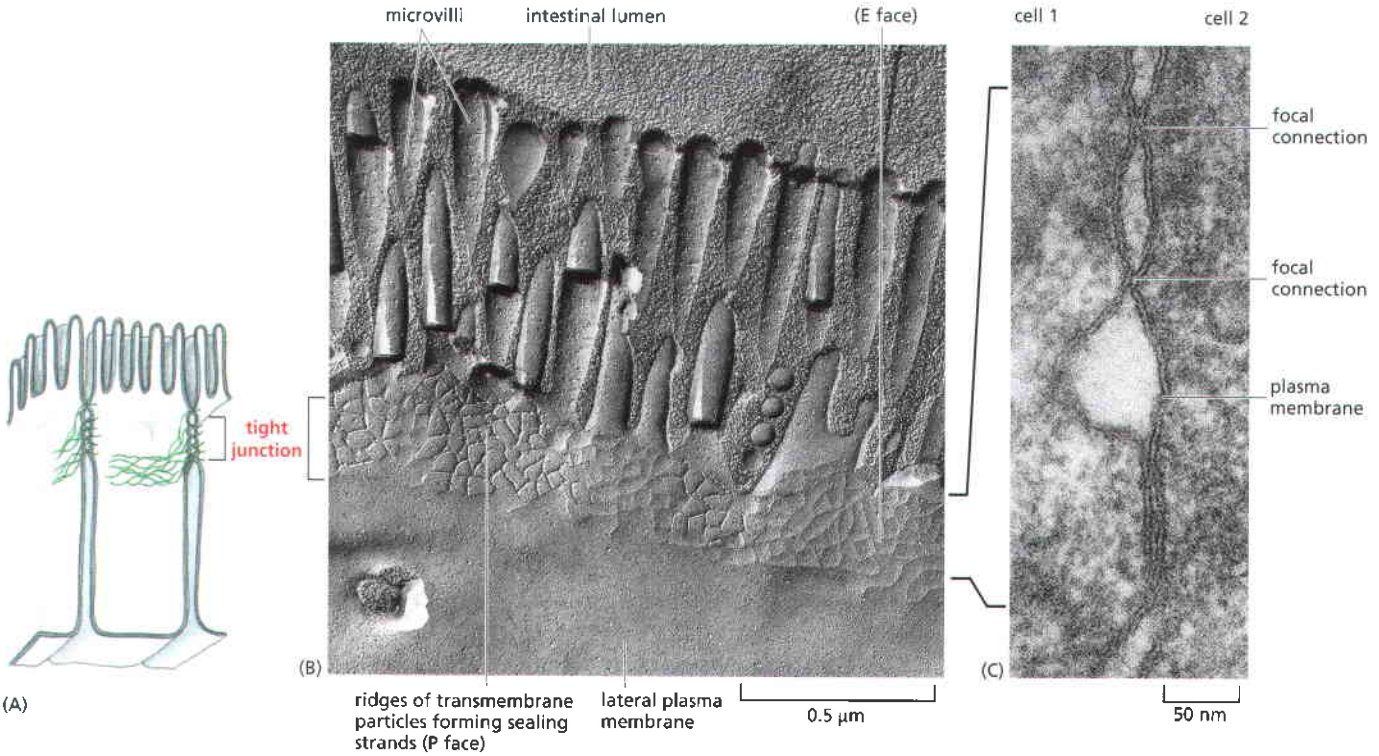


Figure 19–25 The structure of a tight junction between epithelial cells of the small intestine. The junctions are shown (A) schematically, (B) in a freeze-fracture electron micrograph, and (C) in a conventional electron micrograph. In (B), the plane of the micrograph is parallel to the plane of the membrane, and the tight junction appears as a band of branching sealing strands that encircle each cell in the epithelium. The sealing strands are seen as ridges of intramembrane particles on the cytoplasmic fracture face of the membrane (the P face) or as complementary grooves on the external face of the membrane (the E face) (see Figure 19–26A). In (C), the junction is seen in cross section as a series of focal connections between the outer leaflets of the two interacting plasma membranes, each connection corresponding to a sealing strand in cross section. (B and C, from N.B. Gilula, in Cell Communication [R.P. Cox, ed.], pp. 1–29. New York: Wiley, 1974.)

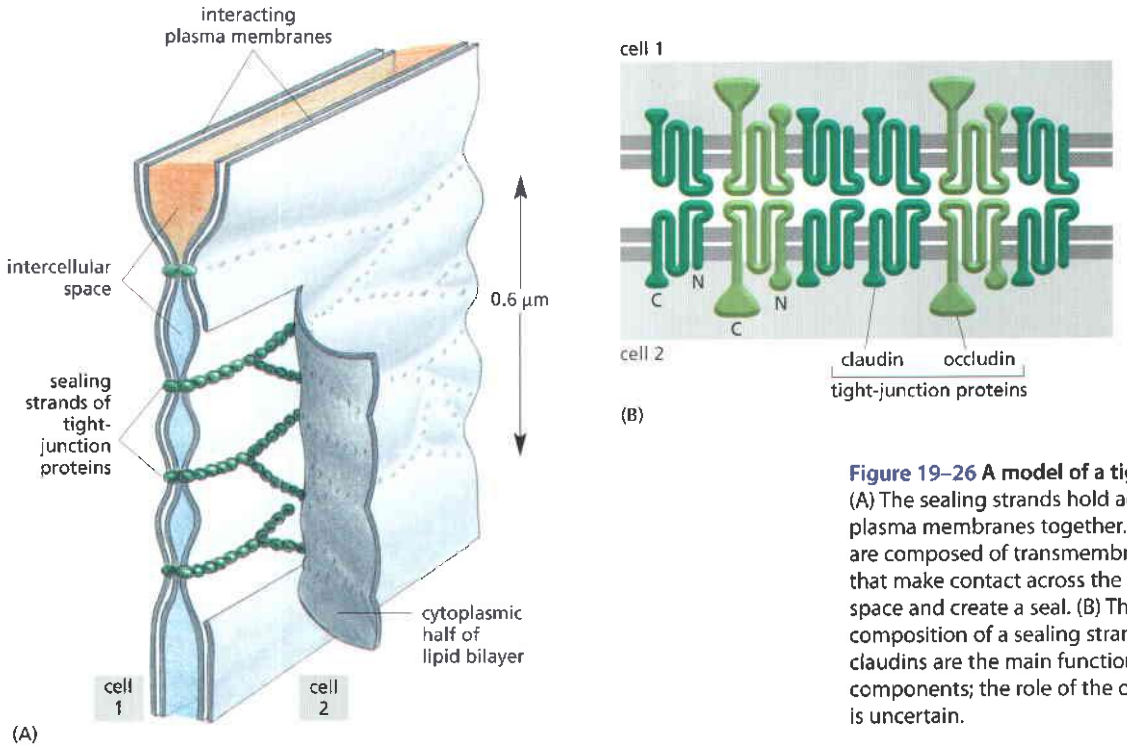


Figure 19–26 A model of a tight junction. (A) The sealing strands hold adjacent plasma membranes together. The strands are composed of transmembrane proteins that make contact across the intercellular space and create a seal. (B) The molecular composition of a sealing strand. The claudins are the main functional components; the role of the occludins is uncertain.

plasma membranes are seen to be tightly apposed where sealing strands are present (Figure 19–25C). Each tight junction sealing strand is composed of a long row of transmembrane adhesion proteins embedded in each of the two interacting plasma membranes. The extracellular domains of these proteins adhere directly to one another to occlude the intercellular space (**Figure 19–26**).

The main transmembrane proteins forming these strands are the *claudins,* which are essential for tight junction formation and function. Mice that lack the *claudin-1* gene, for example, fail to make tight junctions between the cells in the epidermal layer of the skin; as a result, the baby mice lose water rapidly by evaporation through the skin and die within a day after birth. Conversely, if nonepithelial cells such as fibroblasts are artificially caused to express claudin genes, they will form tight-junctional connections with one another. Normal tight junctions also contain a second major transmembrane protein called *occludin,* but the function of this protein is uncertain, and it does not seem to be as essential as the claudins. A third transmembrane protein, *tricellulin* (related to occludin), is required to seal cell membranes together and prevent transepithelial leakage at the points where three cells meet.

The claudin protein family has many members (24 in humans), and these are expressed in different combinations in different epithelia to confer particular permeability properties on the epithelial sheet. They are thought to form *paracellular pores*—selective channels allowing specific ions to cross the tight-junctional barrier, from one extracellular space to another. A specific claudin found in kidney epithelial cells, for example, is needed to let Mg^{2+} pass between the cells of the sheet so that this ion can be resorbed from the urine into the blood. A mutation in the gene encoding this claudin results in excessive loss of Mg^{2+} in the urine.

Scaffold Proteins in Junctional Complexes Play a Key Part in the Control of Cell Proliferation

The claudins and occludins have to be held in the right position in the cell, so as to form the tight-junctional network of sealing strands. This network usually lies just apical to the adherens and desmosome junctions that bond the cells together mechanically, and the whole assembly is called a *junctional complex* (**Figure 19–27**). The parts of this junctional complex depend on each other for

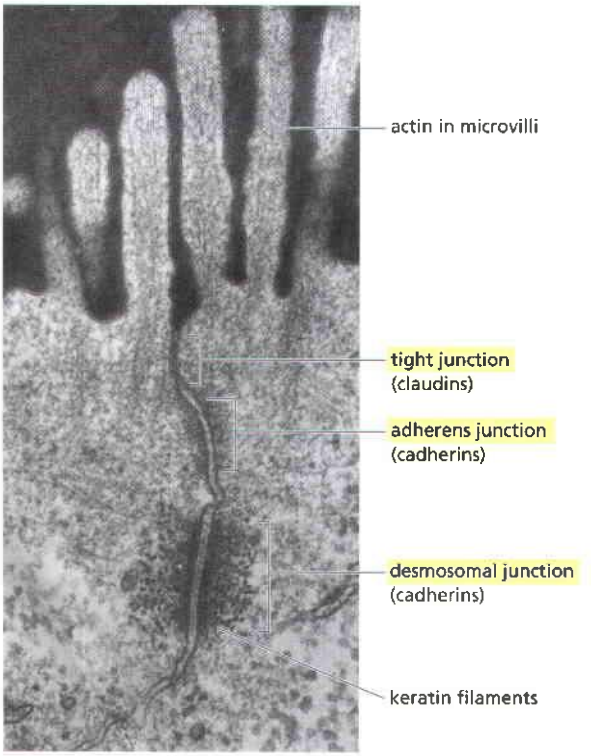


Figure 19–27 A junctional complex between two epithelial cells in the lining of the gut. Most apically, there is a tight junction; beneath this, an adherens junction; and beneath the adherens junction, a desmosomal junction. This example is from a vertebrate; in insects, the arrangement is different. (Courtesy of Daniel S. Friend.)

their formation. For example, anti-cadherin antibodies that block the formation of adherens junctions also block the formation of tight junctions. The positioning and organization of tight junctions in relation to these other structures is thought to depend on association with intracellular scaffold proteins of the *Tjp (Tight junction protein)* family, also called *ZO proteins* (a tight junction is also known as a *zonula occludens*). The vertebrate Tjp proteins belong to the same family as the Discs-large proteins that we mentioned earlier for their role at synapses, and they anchor the tight-junctional strands to other components including the actin cytoskeleton.

In invertebrates such as insects and mollusks, occluding junctions have a different appearance and are called **septate junctions**. Like tight junctions, these form a continuous band around each epithelial cell, but the structure is more regular, and the interacting plasma membranes are joined by proteins that are arranged in parallel rows with a regular periodicity (**Figure 19–28**). Septate junctions are nevertheless based on proteins homologous to the vertebrate claudins, and they depend on scaffold proteins in a similar way, including in particular the same Discs-large protein that is present at synapses. Mutant flies that are deficient in Discs-large have defective septate junctions.

Strikingly, these mutants also develop epithelial tumors, in the form of large overgrowths of the imaginal discs—the structures in the fly larva from which most of the adult body derives (as described in Chapter 22). The gene takes its name from this remarkable effect, which depends on the presence of binding sites for growth regulators on the Discs-large protein. But why should the apparatus of cell–cell adhesion be linked in this way with the control of cell proliferation? The relationship seems to be fundamental: in vertebrates also, genes

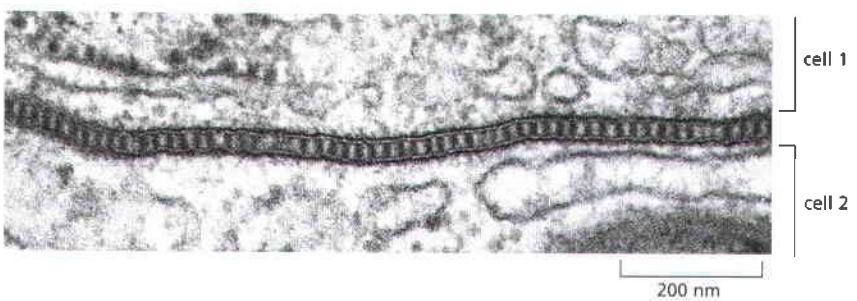


Figure 19–28 A septate junction. A conventional electron micrograph of a septate junction between two epithelial cells in a mollusk. The interacting plasma membranes, seen in cross section, are connected by parallel rows of junctional proteins. The rows, which have a regular periodicity, are seen as dense bars, or septa. (From N.B. Gilula, in Cell Communication [R.P. Cox, ed.], pp. 1–29. New York: Wiley, 1974.)

homologous to *Discs large* have this dual involvement. One possibility is that it reflects a basic mechanism for repair and maintenance of epithelia. If an epithelial cell is deprived of adhesive contacts with neighbors, its program of growth and proliferation is activated, thereby creating new cells to reconstruct a continuous multicellular sheet. In fact, a large body of evidence indicates that junctional complexes are important sites of cell–cell signaling not only via Discs-large but also through other components of these structures, including cadherins as we have seen.

Cell–Cell Junctions and the Basal Lamina Govern Apico-Basal Polarity in Epithelia

Most cells in animal tissues are strongly polarized: they have a front that differs from the back, or a top that differs from the bottom. Examples include virtually all epithelial cells, as we have discussed, as well as neurons with their dendrite–axon polarity, migrating fibroblasts and white blood cells, with their locomotor leading edge and trailing rear end, and many other cells in embryos as they prepare to divide asymmetrically to create daughter cells that are different. A core set of components is critical for cell polarity in all these cases, throughout the animal kingdom, from worms and flies to mammals.

In the case of epithelial cells, these fundamental generators of cell polarity have to establish the difference between the apical and basal poles, and they have to do so in a properly oriented way, in accordance with the cell's surroundings. The basic phenomenon is nicely illustrated by experiments with a cultured line of epithelial cells, called MDCK cells (**Figure 19–29**A). These can be separated from one another and cultured in suspension in a collagen gel. A single isolated cell in these circumstances does not show any obvious polarity, but if it is allowed to divide to form a small colony of cells, these cells will organize themselves into a hollow epithelial vesicle where the polarity of each cell is clearly apparent. The vesicle becomes surrounded by a basal lamina, and all the cells orient themselves in the same way, with apex-specific marker molecules facing the lumen. Evidently, the MDCK cells have a spontaneous tendency to become polarized, but the mechanism is cooperative and depends on contacts with neighbors.

To discover how the underlying molecular mechanism works, the first step is to identify its components. Studies in the worm *C. elegans* and in *Drosophila* have been most informative here. In the worm, a screen for mutations upsetting the organization of the early embryo has revealed a set of genes essential for normal cell polarity and asymmetry of cell division (as discussed in Chapter 22). There are at least six of these genes, called *Par (partitioning defective)* genes. In all animal species studied, they and their homologs (along with other genes discovered through studies in *Drosophila* and vertebrates) have a fundamental role not just in asymmetric cell division in the early embryo, but in many other processes of cell polarization, including the polarization of epithelial cells. The *Par4* gene of *C. elegans*, for example, is homologous to a gene called *Lkb1* in mammals and *Drosophila*, coding for a serine/threonine kinase. In the fly, mutations of

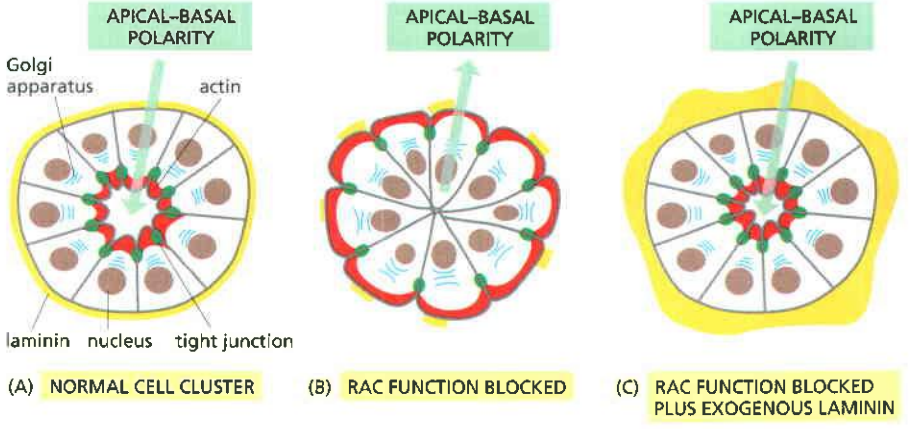


Figure 19–29 Cooperative polarization of a cluster of epithelial cells in culture and its dependence on Rac and laminin. Cells of the MDCK line, derived from dog kidney epithelium, were dissociated, embedded in a collagen matrix, and allowed to proliferate, creating small isolated colonies, shown here schematically in cross section. (A) The cells in such a colony will normally organize themselves spontaneously into an epithelium surrounding a central cavity. Staining for actin (which marks apical microvilli), ZO1 protein (a tight-junction protein), Golgi apparatus, and laminin (a basal lamina component) shows that the cells have all cooperatively become polarized, with apical components facing the lumen of the cavity and basal components facing the surrounding collagen gel. (B) When Rac function is blocked by expression of a dominant-negative form of the protein, the cells show inverted polarity, fail to form a cyst with a central cavity, and cease to deposit laminin in the normal manner around the periphery of the cell cluster. (C) When the cyst is embedded in a matrix rich in exogenous laminin, near-normal polarity is restored even though Rac function is still blocked. (Based on L.E. O'Brien et al., *Nat. Cell Biol.* 3:831–838, 2001. With permission from Macmillan Publishers Ltd.)

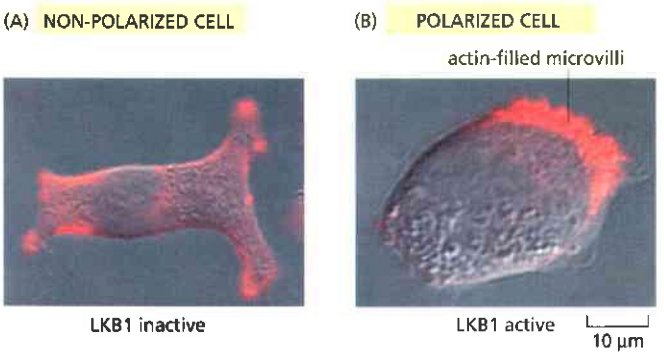


Figure 19–30 Development of polarity in single isolated epithelial cells. Cells of a line derived from intestinal epithelium were transfected with DNA constructs coding for regulatory components through which the activity of the LKB1 protein could be switched on or off by a change in the composition of the culture medium. When LKB1 activity is low, the cells appear unpolarized; when it is high, they become individually polarized. Their polarity is manifest in the distribution of tight-junction proteins (ZO1) and adherens-junction proteins (p120-catenin), which accumulate on one side of the cell, around a cap of actin-filled microvilli, even though the cells are isolated from one another and make no cell–cell junctions. This cell-autonomous polarization occurs even when the cells are cultured in suspension, without contact with any substratum that could tell them which way was up. (From A.F. Baas et al., *Cell* 116:457–466, 2004. With permission from Elsevier.)

this gene disrupt the polarity of the egg cell and of cells in epithelia. In humans, such mutations give rise to *Peutz–Jeghers syndrome*, involving disorderly abnormal growths of the lining of the gut and a predisposition to certain rare types of cancer. When cultured human colon epithelial cells are prevented from expressing LKB1, they fail to polarize normally. Moreover, when such cells in culture are artificially driven to express abnormally high levels of LKB1 activity, they can become individually polarized, even when isolated from other cells, and surrounded on all sides by a uniform medium (**Figure 19–30**). This suggests that normal epithelial polarity depends on two interlocking mechanisms: one that endows individual cells with a tendency to become polarized cell-autonomously, and another that orients their polarity axis in relation to their neighbors and the basal lamina. The latter mechanism would be peculiar to epithelia; the former could be much more general, operating also in other polarized cell types.

The molecules known to be needed for epithelial polarity can be classified in relation to these two mechanisms. Central to the polarity of individual animal cells in general is a set of three membrane-associated proteins: **Par3**, **Par6**, and **atypical protein kinase C** (**aPKC**). Par3 and Par6 are both scaffold proteins containing PDZ domains, and they bind to one another and to aPKC. The complex of these three components also has binding sites for various other molecules, including the small GTPases Rac and Cdc42. These latter molecules play a crucial part. Thus, for example, when Rac function is blocked in a cluster of MDCK cells, the cells develop with inverted polarity (see Figure 19–29B). Rac and Cdc42 are key regulators of actin assembly, as explained in Chapter 16; through them, it seems, assembly of a Par3-Par6-aPKC complex in a specific region of the cell cortex is associated with polarization of the cytoskeleton towards that region. The assembly process is evidently cooperative and involves some positive feedback and spatial signaling, so that a small initial cluster of these components is able to recruit more of them and to inhibit the development of clusters of the same type elsewhere in the cell. One source of positive feedback may lie in the behavior of Cdc42 and Rac: a high activity of these molecules at a particular site, by organizing the cytoskeleton, may direct intracellular transport so as to bring still more Cdc42 or Rac, or more of their activators, to the same site. This is suspected to be an essential part of the polarization mechanism in budding yeast cells, and it may be the way in which cells such as migrating fibroblasts establish the difference between their leading edge and the rest of their periphery. It could be the core of the eucaryotic cell polarization machinery, at least in evolutionary terms.

The Par3-Par6-aPKC complex, combined with Cdc42 or Rac, seems to control the organization of other protein complexes associated with the internal face of the cell membrane. In particular, in epithelial cells, it causes the *Crumbs complex*, held together by the PDZ-domain scaffold proteins Discs-lost and Stardust, to become localized toward the apex of the cell, while a third such complex, called the *Scribble complex*, held together by the scaffold proteins Scribble and Discs-large (the same protein that we encountered previously) is localized more basally (**Figure 19–31**). These various protein assemblies interact with one another and with other cell components in ways that are only beginning to be understood.

But how is this whole elaborate system oriented correctly in relation to neighboring cells? In an epithelium, the Par3-Par6-aPKC complex assembles at

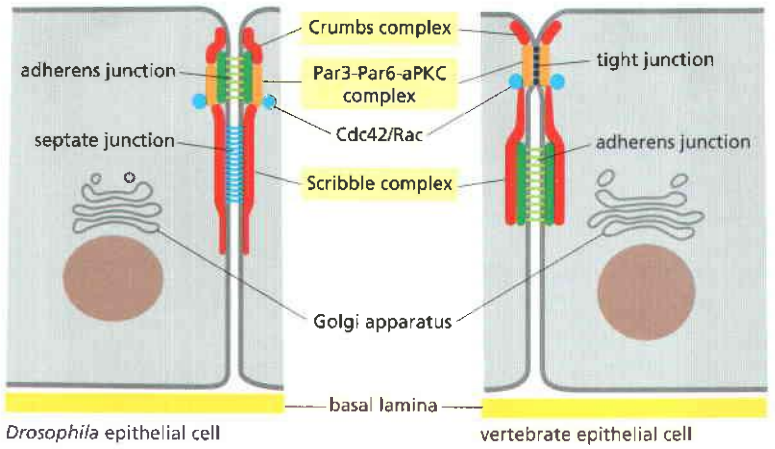


Figure 19–31 The coordinated arrangement of three membrane-associated protein complexes thought to be critical for epithelial polarity. A *Drosophila* epithelial cell is shown schematically on the left, and a vertebrate epithelial cell on the right. All three complexes—the Par3-Par6-aPkc complex, the Crumbs complex, and the Scribble complex—are organized around scaffold proteins containing PDZ domains. The detailed distribution of the complexes varies somewhat according to cell type.

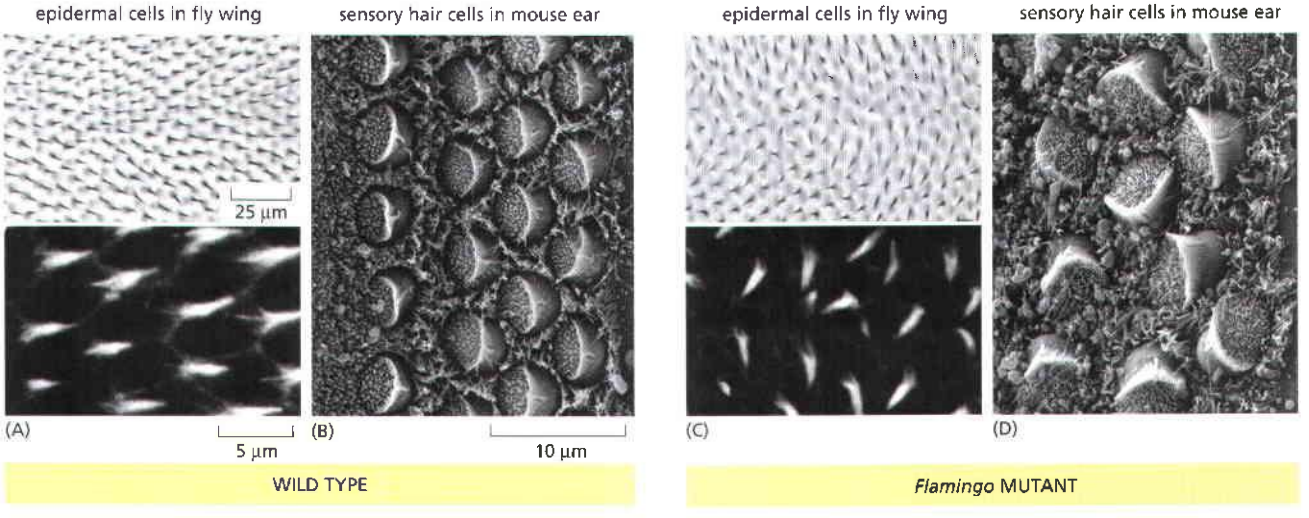
cell–cell junctions—tight junctions in vertebrates, adherens junctions in *Drosophila*—because the scaffold proteins in the complex bind to the tails of certain of the junctional transmembrane adhesion proteins. Meanwhile, the cytoskeleton, under the influence of Rac or its relatives, directs the delivery of basal lamina components to the opposite end of the cell. These extracellular matrix molecules then act back on the cell to give that region a basal character (see Figure 19–29C). In this way, the polarity of the cell is coupled to its orientation in the epithelial sheet and its relation to the basal lamina.

A Separate Signaling System Controls Planar Cell Polarity

Apico-basal polarity is a universal feature of epithelia, but the cells of some epithelia show an additional polarity at right angles to this axis: it is as if they had an arrow written on them, pointing in a specific direction in the plane of the epithelium. This type of polarity is called **planar cell polarity** (Figure 19–32A and B). In the wing of a fly, for example, each epithelial cell has a tiny asymmetrical projection, called a wing-hair, on its surface, and the hairs all point toward the tip of the wing. Similarly, in the inner ear of a vertebrate, each mechanosensory hair cell has an asymmetric bundle of stereocilia (actin-filled rod-like protrusions) sticking up from its apical surface: tilting the bundle in one direction causes ion channels to open, stimulating the cell electrically; tilting in the opposite direction has the contrary effect. For the ear to function correctly, the hair cells must be correctly oriented. Planar cell polarity is important also in the respiratory tract, for example, where every ciliated cell must orient its beating so as to sweep mucus up out of the lungs, and not down into them (see Chapter 23).

Screens for mutants with disorderly wing hairs in *Drosophila* have identified a set of genes that are critical for planar cell polarity in the fly. Some of these,

Figure 19–32 Planar cell polarity. (A) Wing hairs on the wing of a fly. Each cell in the wing epithelium forms one of these little spiky protrusions or "hairs" at its apex, and all the hairs point the same way, toward the tip of the wing. This reflects a planar polarity in the structure of each cell. (B) Sensory hair cells in the inner ear of a mouse similarly have a well-defined planar polarity, manifest in the oriented pattern of stereocilia (actin-filled protrusions) on their surface. The detection of sound depends on the correct, coordinated orientation of the hair cells. (C) A mutation in the gene *Flamingo* in the fly, coding for a non-classical cadherin, disrupts the pattern of planar cell polarity in the wing. (D) A mutation in a homologous *Flamingo* gene in the mouse randomizes the orientation of the planar cell polarity vector of the hair cells in the ear. The mutant mice are deaf. (A and C, from J. Chae et al., *Development* 126:5421–5429, 1999. With permission from The Company of Biologists; B and D, from J.A. Curtin et al., *Curr. Biol.* 13:1129–1133, 2003. With permission from Elsevier.)



such as *Frizzled*, for example, and *Dishevelled*, code for proteins that have since been shown to be components of the Wnt signaling pathway (discussed in Chapter 15). Two others, *Flamingo* (see Figure 19–32C) and *Dachsous*, code for members of the cadherin superfamily. Still others are less easily classified functionally, but it is clear that planar cell polarity is organized by machinery formed from these components and assembled at cell–cell junctions in such a way that a polarizing influence can propagate from cell to cell. Essentially the same system of proteins controls planar cell polarity in vertebrates. Mice with mutations in a *Flamingo* homolog, for example, have incorrectly oriented hair cells in their ears (among other defects) and thus are deaf (see Figure 19–32D).

Summary

Occluding junctions—tight junctions in vertebrates, septate junctions in insects and molluscs—seal the gaps between cells in epithelia, creating a barrier to the diffusion of molecules across the cell sheet. They also form a bar to the diffusion of proteins in the plane of the membrane, and so help to maintain a difference between the populations of proteins in the apical and basolateral membrane domains of the epithelial cell. The major transmembrane proteins forming occluding junctions are called claudins; different members of the family are expressed in different tissues, conferring different permeability properties on the various epithelial sheets.

Intracellular scaffold proteins bind to the transmembrane components at occluding junctions and coordinate these junctions with cadherin-based anchoring junctions, so as to create junctional complexes. The junctional scaffold proteins have at least two other crucial functions. They play a part in the control of epithelial cell proliferation; and, in conjunction with other regulatory molecules such as Rac and Cdc42, they govern cell polarity. Epithelial cells have an intrinsic tendency to develop a polarized apico-basal axis. The orientation of this axis in relation to the cell's neighbors in an epithelial sheet depends on protein complexes involving scaffold proteins that assemble at cell–cell junctions, as well as on cytoskeletal polarization controlled by Rac/Cdc42 and on influences from the basal lamina.

The cells of some epithelia have an additional polarity in the plane of the epithelium, at right angles to the apico-basal axis. A separate set of conserved proteins, operating in a similar way in vertebrates and in insects, governs this planar cell polarity through poorly understood signaling processes that are likewise based on cell–cell junctions.

PASSAGEWAYS FROM CELL TO CELL: GAP JUNCTIONS AND PLASMODESMATA

Tight junctions block the passageways through the gaps between cells, preventing extracellular molecules from leaking from one side of an epithelium to the other. Another type of junctional structure has a radically different function: it bridges gaps between adjacent cells so as to create direct passageways from the cytoplasm of one into that of the other. These passageways take quite different forms in animal tissues, where they are called *gap junctions*, and in plants, where they are called *plasmodesmata* (singular *plasmodesma*). In both cases, however, the function is similar: the connections allow neighboring cells to exchange small molecules but not macromolecules (with some exceptions for plasmodesmata). Many of the implications of this cell coupling are only beginning to be understood.

Gap Junctions Couple Cells Both Electrically and Metabolically

Gap junctions are present in most animal tissues, including connective tissues as well as epithelia, allowing the cells to communicate with their neighbors. Each gap junction appears in conventional electron micrographs as a patch where the membranes of two adjacent cells are separated by a uniform narrow

gap of about 2–4 nm. The gap is spanned by channel-forming proteins, of which there are two distinct families, called the *connexins* and the *innexins*. These are unrelated in sequence but similar in shape and function: in vertebrates, both families are present, but connexins predominate, with 21 members in humans. In *Drosophila* and *C. elegans*, only innexins are present, with 15 family members in the fly and 25 in the worm.

The channels formed by the gap-junction proteins allow inorganic ions and other small water-soluble molecules to pass directly from the cytoplasm of one cell to the cytoplasm of the other, thereby coupling the cells both electrically and metabolically. Thus, when a suitable dye is injected into one cell, it diffuses readily into the other, without escaping into the extracellular space. Similarly, an electric current injected into one cell through a microelectrode causes an almost instantaneous electrical disturbance in the neighboring cell, due to the flow of ions carrying electric charge through gap junctions. With microelectrodes inserted into both cells, one can easily monitor this effect and measure properties of the gap junctions, such as their electrical resistance and the ways in which the coupling changes as conditions change. In fact, some of the earliest evidence of gap-junctional communication came from electrophysiological studies that demonstrated this type of rapid, direct electrical coupling between some types of neurons. Similar methods were used to identify connexins as the proteins that mediate the gap-junctional communication: when connexin mRNA is injected into either frog oocytes or gap-junction-deficient cultured cells, channels with the properties expected of gap-junction channels can be demonstrated electrophysiologically where pairs of injected cells make contact.

From experiments with injected dye molecules of different sizes, it seems that the largest functional pore size for gap-junctional channels is about 1.5 nm. Thus, the coupled cells share their small molecules (such as inorganic ions, sugars, amino acids, nucleotides, vitamins, and the intracellular mediators cyclic AMP and inositol trisphosphate) but not their macromolecules (proteins, nucleic acids, and polysaccharides) (**Figure 19–33**).

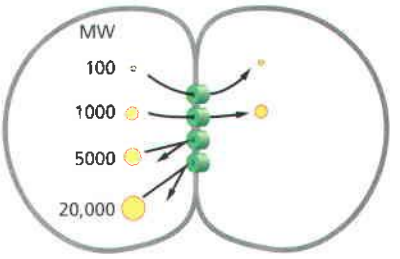


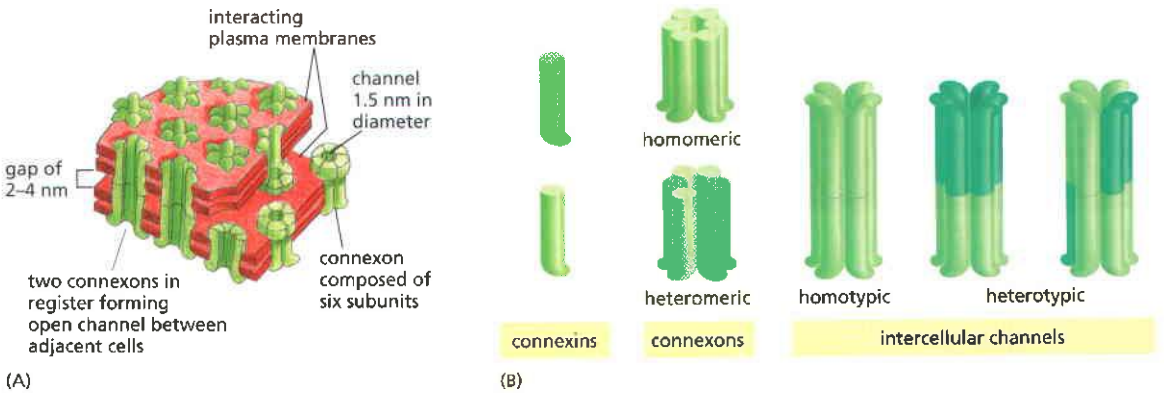
Figure 19–33 Determining the size of a gap-junction channel. When fluorescent molecules of various sizes are injected into one of two cells coupled by gap junctions, molecules with a mass of less than about 1000 daltons can pass into the other cell, but larger molecules cannot.

A Gap-Junction Connexon Is Made Up of Six Transmembrane Connexin Subunits

Connexins are four-pass transmembrane proteins, six of which assemble to form a *hemichannel*, or **connexon**. When the connexons in the plasma membranes of two cells in contact are aligned, they form a continuous aqueous channel that connects the two cell interiors (**Figure 19–34A** and **Figure 19–35**). A gap junction consists of many such connexon pairs in parallel, forming a sort of molecular sieve. The connexons hold the interacting plasma membranes a fixed distance apart—hence the gap.

Gap junctions in different tissues can have different properties because they are formed from different combinations of connexins, creating channels that differ in permeability. Most cell types express more than one type of connexin, and two different connexin proteins can assemble into a heteromeric connexon, with its own distinct properties. Moreover, adjacent cells expressing different

Figure 19–34 Gap junctions. (A) A three-dimensional drawing showing the interacting plasma membranes of two adjacent cells connected by gap junctions. Each lipid bilayer is shown as a pair of *red* sheets. Protein assemblies called connexons (*green*), each of which is formed by six connexin subunits, penetrate the apposed lipid bilayers (*red*). Two connexons join across the intercellular gap to form a continuous aqueous channel connecting the two cells. (B) The organization of connexins into connexons and connexons into intercellular channels. The connexons can be homomeric or heteromeric, and the intercellular channels can be homotypic or heterotypic.



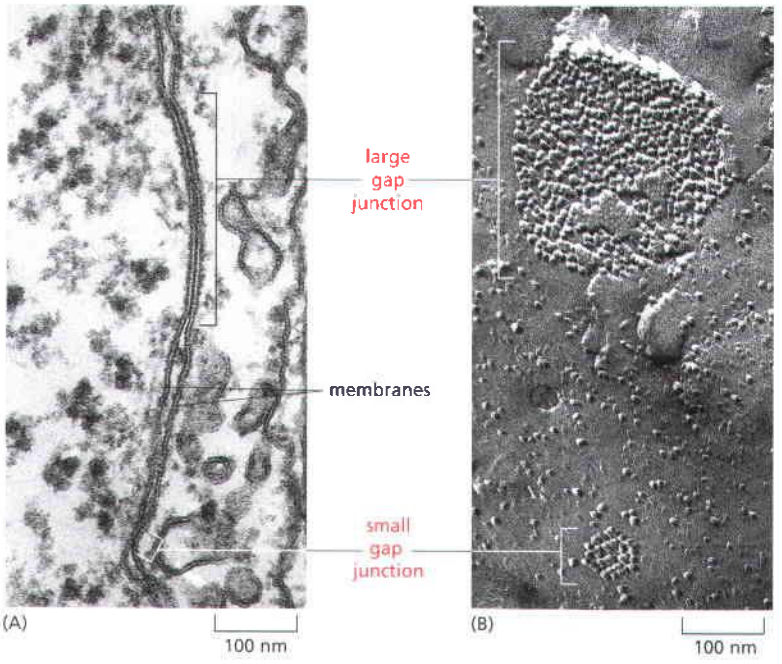


Figure 19–35 Gap junctions as seen in the electron microscope. (A) Thin-section and (B) freeze-fracture electron micrographs of a large and a small gap junction between fibroblasts in culture. In (B), each gap junction is seen as a cluster of homogeneous intramembrane particles. Each intramembrane particle corresponds to a connexon. (From N.B. Gilula, in Cell Communication [R.P. Cox, ed.], pp. 1–29. New York: Wiley, 1974.)

connexins can form intercellular channels in which the two aligned half-channels are different (Figure 19–34B).

Each gap-junctional plaque is a dynamic structure that can readily assemble, disassemble, or be remodelled, and it can contain a cluster of a few to many thousands of connexons (see Figure 19–35B). Studies with fluorescently labeled connexins in living cells show that new connexons are continually added around the periphery of an existing junctional plaque, while old connexons are removed from the middle of it and destroyed (**Figure 19–36**). This turnover is rapid: the connexin molecules have a half-life of a few hours.

The mechanism of removal of old connexons from the middle of the plaque is not known, but the route of delivery of new connexons to its periphery seems clear: they are inserted into the plasma membrane by exocytosis, like other integral membrane proteins, and then diffuse in the plane of the membrane until they bump into the periphery of a plaque and become trapped. This has a corollary: the plasma membrane away from the gap junction should contain connexons—hemichannels—that have not yet paired with their counterparts on another cell. It is thought that these unpaired hemichannels are normally held

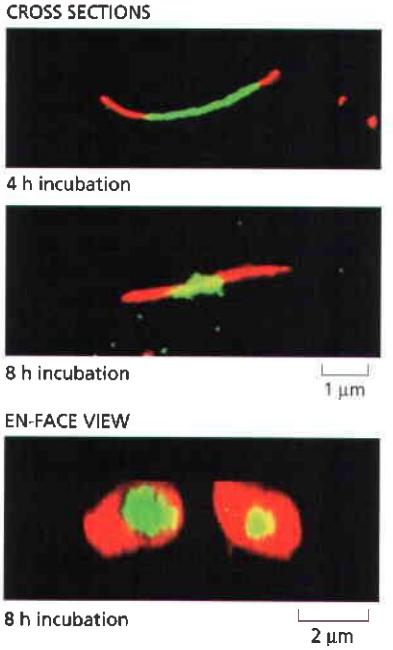


Figure 19–36 Connexin turnover at a gap junction. Cells were transfected with a slightly modified connexin gene, coding for a connexin with a short amino-acid tag containing four cysteines in the sequence ...Cys-Cys-X-X-Cys-Cys (where X denotes an arbitrary amino acid). This *tetracysteine tag* can bind strongly, and in effect irreversibly, to certain small fluorescent dye molecules that can be added to the culture medium and will readily enter cells by diffusing across the plasma membrane. In the experiment shown, a green dye was added first, and the cells were then washed and incubated for 4 or 8 hours. At the end of this time, a red dye was added to the medium and the cells were washed again and fixed. Connexin molecules already present at the beginning of the experiment are labeled green (and take up no red dye because their tetracysteine tags are already saturated with green dye), while connexins synthesized subsequently, during the 4- or 8-hour incubation, are labeled red. The fluorescence images show optical sections of gap junctions between pairs of cells prepared in this way. The central part of the gap-junction plaque is *green*, indicating that it consists of old connexin molecules, while the periphery is *red*, indicating that it consists of connexins synthesized during the past 4 or 8 hours. The longer the time of incubation, the smaller the green central patch of old molecules, and the larger the peripheral ring of new molecules that have been recruited to replace them. (From G. Gaietta et al., *Science* 296:503–507, 2002. With permission from AAAS.)

in a closed conformation, preventing the cell from losing its small molecules by leakage through them. But there is also evidence that in some physiological circumstances they can open and serve as channels for the release of small molecules, such as the neurotransmitter glutamate, to the exterior, or for the entry of small molecules into the cell.

Gap Junctions Have Diverse Functions

In tissues containing electrically excitable cells, cell–cell coupling via gap junctions serves an obvious purpose. Some nerve cells, for example, are electrically coupled, allowing action potentials to spread rapidly from cell to cell, without the delay that occurs at chemical synapses. This is advantageous when speed and reliability are crucial, as in certain escape responses in fish and insects, or where a set of neurons need to act in synchrony. Similarly, in vertebrates, electrical coupling through gap junctions synchronizes the contractions of heart muscle cells as well as those of the smooth muscle cells responsible for the peristaltic movements of the intestine.

Gap junctions also occur in many tissues whose cells are not electrically excitable. In principle, the sharing of small metabolites and ions provides a mechanism for coordinating the activities of individual cells in such tissues and for smoothing out random fluctuations in small-molecule concentrations in different cells. Gap junctions are required in the liver, for example, to coordinate the response of the liver cells to signals from nerve terminals that contact only a part of the cell population (see Figure 15–7). The normal development of ovarian follicles also depends on gap-junction-mediated communication—in this case, between the oocyte and the surrounding granulosa cells. A mutation in the gene that encodes the connexin that normally couples these two cell types causes infertility.

Mutations in connexins, especially connexin-26, are the commonest of all genetic causes of congenital deafness: they result in the death of cells in the organ of Corti, probably because they disrupt functionally important pathways for the flow of ions from cell to cell in this electrically active sensory epithelium. Connexin mutations are responsible for many other disorders besides deafness, ranging from cataracts in the lens of the eye to a form of demyelinating disease in peripheral nerves.

Cell coupling via gap junctions also seems to play a part in embryogenesis. In early vertebrate embryos (beginning with the late eight-cell stage in mouse embryos), most cells are electrically coupled to one another. As specific groups of cells in the embryo develop their distinct identities and begin to differentiate, they commonly uncouple from surrounding tissue. As the neural plate folds up and pinches off to form the neural tube, for instance (see Figure 19–16), its cells uncouple from the overlying ectoderm. Meanwhile, the cells within each group remain coupled with one another and therefore tend to behave as a cooperative assembly, all following a similar developmental pathway in a coordinated fashion.

Cells Can Regulate the Permeability of Their Gap Junctions

Like conventional ion channels (discussed in Chapter 11), individual gap-junction channels do not remain continuously open; instead, they flip between open and closed states. Moreover, the permeability of gap junctions is rapidly (within seconds) and reversibly reduced by experimental manipulations that decrease the cytosolic pH or increase the cytosolic concentration of free Ca^{2+} to very high levels.

The purpose of the pH regulation of gap-junction permeability is unknown. In one case, however, the purpose of Ca^{2+} control seems clear. When a cell is damaged, its plasma membrane can become leaky. Ions present at high concentration in the extracellular fluid, such as Ca^{2+} and Na^{+}, then move into the cell, and valuable metabolites leak out. If the cell were to remain coupled to its

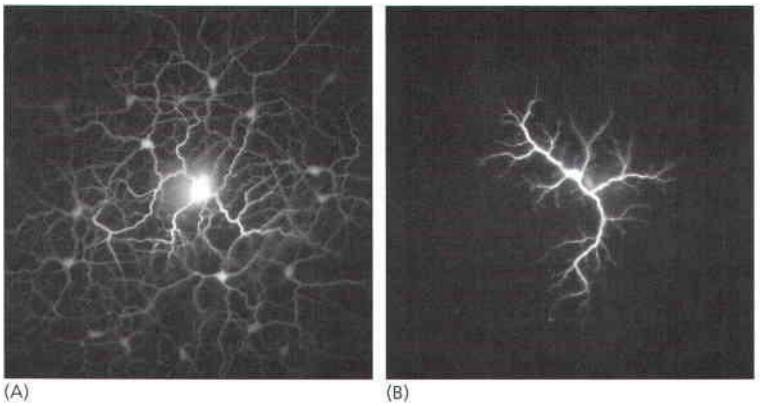


Figure 19–37 The regulation of gap-junction coupling by a neurotransmitter. (A) A neuron in a rabbit retina was injected with the dye Lucifer yellow, which passes readily through gap junctions and labels other neurons of the same type that are connected to the injected cell by gap junctions. (B) The retina was first treated with the neurotransmitter dopamine, before the neuron was injected with dye. As can be seen, the dopamine treatment greatly decreased the permeability of the gap junctions. Dopamine acts by increasing intracellular cyclic AMP levels. (Courtesy of David Vaney.)

healthy neighbors, these too would suffer a dangerous disturbance of their internal chemistry. But the large influx of Ca^{2+} into the damaged cell causes its gap-junction channels to close immediately, effectively isolating the cell and preventing the damage from spreading to other cells.

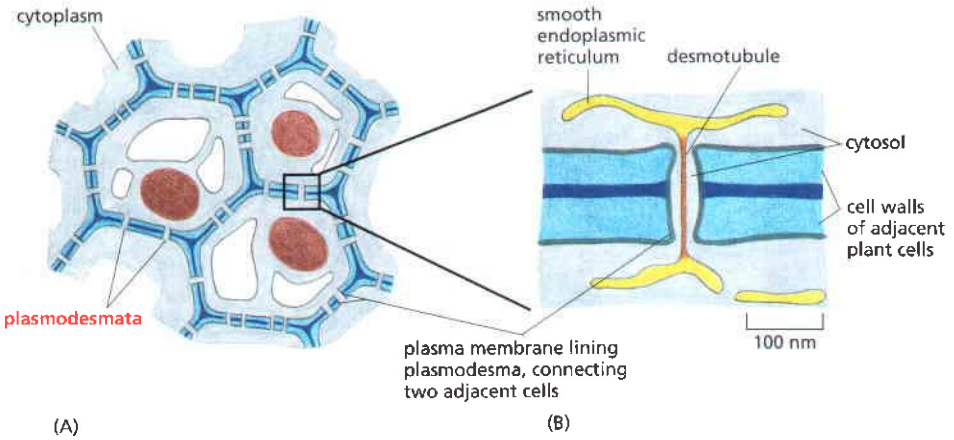
Gap-junction communication can also be regulated by extracellular signals. The neurotransmitter *dopamine*, for example, reduces gap-junction communication between a class of neurons in the retina in response to an increase in light intensity (**Figure 19–37**). This reduction in gap-junction permeability helps the retina switch from using rod photoreceptors, which are good detectors of low light, to cone photoreceptors, which detect color and fine detail in bright light.

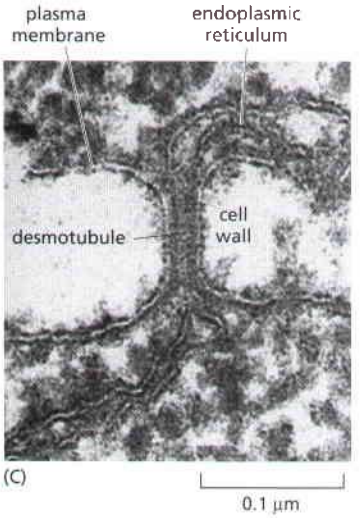
In Plants, Plasmodesmata Perform Many of the Same Functions as Gap Junctions

The tissues of a plant are organized on different principles from those of an animal. This is because plant cells are imprisoned within tough *cell walls* composed of an extracellular matrix rich in cellulose and other polysacharides, as we discuss later. The cell walls of adjacent cells are firmly cemented to those of their neighbors, which eliminates the need for anchoring junctions to hold the cells in place. But a need for direct cell–cell communication remains. Thus, plant cells have only one class of intercellular junctions, **plasmodesmata**. Like gap junctions, they directly connect the cytoplasms of adjacent cells.

In plants, the cell wall between a typical pair of adjacent cells is at least 0.1 μm thick, and so a structure very different from a gap junction is required to mediate communication across it. Plasmodesmata solve the problem. With a few specialized exceptions, every living cell in a higher plant is connected to its living neighbors by these structures, which form fine cytoplasmic channels through the intervening cell walls. As shown in **Figure 19–38**A, the plasma membrane of one cell is continuous with that of its neighbor at each plasmodesma, which connects the cytoplasms of the two cells by a roughly cylindrical channel with a diameter of 20–40 nm.

Running through the center of the channel in most plasmodesmata is a narrower cylindrical structure, the *desmotubule*, which is continuous with elements of the smooth endoplasmic reticulum in each of the connected cells (Figure 19–38B–D). Between the outside of the desmotubule and the inner face of the cylindrical channel formed by plasma membrane is an annulus of cytosol through which small molecules can pass from cell to cell. As each new cell wall is assembled during the cytokinesis phase of cell division, plasmodesmata are created within it. They form around elements of smooth ER that become trapped across the developing cell plate (discussed in Chapter 17). They can also be inserted *de novo* through preexisting cell walls, where they are commonly found in dense clusters called *pit fields*. When no longer required, plasmodesmata can be readily removed.





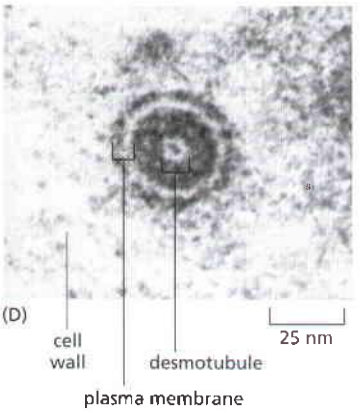


Figure 19–38 Plasmodesmata. (A) The cytoplasmic channels of plasmodesmata pierce the plant cell wall and connect all cells in a plant together. (B) Each plasmodesma is lined with plasma membrane that is common to two connected cells. It usually also contains a fine tubular structure, the desmotubule, derived from smooth endoplasmic reticulum. (C) Electron micrograph of a longitudinal section of a plasmodesma from a water fern. The plasma membrane lines the pore and is continuous from one cell to the next. Endoplasmic reticulum and its association with the central desmotubule can also be seen. (D) A similar plasmodesma seen in cross section. (C and D, from R. Overall, J. Wolfe and B.E.S. Gunning, in Protoplasma 9, pp. 137 and 140. Heidelberg: Springer-Verlag, 1982.)

In spite of the radical difference in structure between plasmodesmata and gap junctions, they seem to function in remarkably similar ways. Evidence obtained by injecting tracer molecules of different sizes suggests that plasmodesmata allow the passage of molecules with a molecular weight of less than about 800, which is similar to the molecular-weight cutoff for gap junctions. As with gap junctions, transport through plasmodesmata is regulated. Dye-injection experiments, for example, show that there can be barriers to the movement of even low-molecular-weight molecules between certain cells, or groups of cells, that are connected by apparently normal plasmodesmata; the mechanisms that restrict communication in these cases are not understood.

During plant development, groups of cells within the shoot and root meristems signal to one another in the process of defining their future fates (discussed in Chapter 22). Some gene regulatory proteins involved in this process of cell fate determination pass from cell to cell through plasmodesmata. They bind to components of the plasmodesmata and override the size exclusion mechanism that would otherwise prevent their passage. In some cases, the mRNA that encodes the protein can also pass through. Some plant viruses also exploit this route: infectious viral RNA, or even intact virus particles, can pass from cell to cell in this way. These viruses produce proteins that bind to components of the plasmodesmata in ways that increase the effective pore size of the channel dramatically. The functional components of plasmodesmata that mediate this effect are unknown, and it is still a mystery how endogenous or viral macromolecules regulate the transport properties of the channel so as to pass through it.

Summary

The cells of many animal tissues, both epithelial and nonepithelial, are coupled by channel-forming junctions called gap junctions. These take the form of plaques of clustered connexons that usually allow molecules smaller than about 1000 daltons to pass directly from the inside of one cell to the inside of the next. The cells can regulate the permeability of these junctions. Gap junctions are dynamic structures: new connexons are

continually recruited to the periphery of the plaque, while old connexons are continually removed from its center.

Cells connected by gap junctions share many of their inorganic ions and other small molecules and are therefore chemically and electrically coupled. Gap junctions are important in coordinating the activities of electrically active cells, and they have a coordinating role in other groups of cells as well. In plants, cells are linked by communicating junctions called plasmodesmata. Although their structure is entirely different from that of gap junctions, and they can sometimes transport informational macromolecules, they function remarkably like gap junctions in permitting small molecules to pass from cell to cell while blocking the passage of most large molecules.

THE BASAL LAMINA

Tissues are not made up solely of cells. A part of their volume—sometimes a very large part—is *extracellular space*, which is occupied by an intricate network of macromolecules constituting the *extracellular matrix*. This matrix is composed of various proteins and polysaccharides that are secreted locally and assembled into an organized meshwork in close association with the surfaces of the cells that produced them.

In our own bodies, the most plentiful forms of extracellular matrix are found in bulky connective tissues such as bone, tendon, and the dermal layer of the skin. For animals in general, however, from an evolutionary point of view, pride of place goes to the extracellular matrix that forms a much less obvious structure—the **basal lamina** (also referred to as the **basement membrane**). This exceedingly thin, tough, flexible sheet of matrix molecules is an essential underpinning of all epithelia. Small as it is in volume, it has a critical role in the architecture of the body. Like the cadherins, it seems to be one of the defining features common to all multicellular animals. Other forms of extracellular matrix are more variable from one animal phylum to another, in both composition and quantity.

In this section, we discuss the basal lamina itself. In the next section we shall consider how epithelial cells and basal lamina interact with one another, through *integrin proteins* in the epithelial cell membranes, and we shall see that integrins are also present in other cell types, mediating their interactions with the varied types of extracellular matrix found in connective tissues. These other forms of extracellular matrix will be discussed in detail later.

Basal Laminae Underlie All Epithelia and Surround Some Nonepithelial Cell Types

Basal laminae are typically 40–120 nm thick. A sheet of basal lamina not only underlies all epithelia but also surrounds individual muscle cells, fat cells, and Schwann cells (which wrap around peripheral nerve cell axons to form myelin). The basal lamina thus separates these cells and epithelia from the underlying or surrounding connective tissue and forms the mechanical connection between them. In other locations, such as the kidney glomerulus, a basal lamina lies between two cell sheets and functions as a selective filter (**Figure 19–39**). Basal laminae have more than simple structural and filtering roles, however. They are able to determine cell polarity, influence cell metabolism, organize the proteins

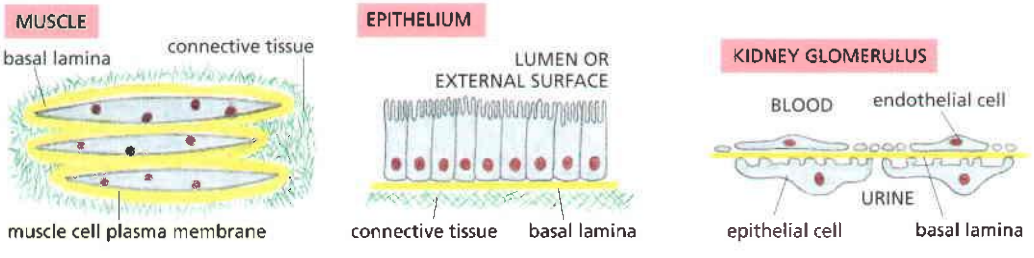


Figure 19–39 Three ways in which basal laminae are organized. Basal laminae *(yellow)* surround certain cells (such as skeletal muscle cells), underlie epithelia, and are interposed between two cell sheets (as in the kidney glomerulus). Note that, in the kidney glomerulus, both cell sheets have gaps in them, and the basal lamina has a filtering as well as a supportive function, helping to determine which molecules will pass into the urine from the blood. The filtration also depends on other protein-based structures, called *slit diaphragms*, that span the intercellular gaps in the epithelial sheet.

in adjacent plasma membranes, promote cell survival, proliferation, or differentiation, and serve as highways for cell migration.

The mechanical role is nevertheless essential. In the skin, for example, the epithelial outer layer—the epidermis—depends on the strength of the basal lamina to keep it attached to the underlying connective tissue—the dermis. In people with genetic defects in certain basal lamina proteins or in a special type of collagen that anchors the basal lamina to the underlying connective tissue, the epidermis becomes detached from the dermis. This causes a blistering disease called *junctional epidermolysis bullosa*, a severe and sometimes lethal condition.

Laminin Is a Primary Component of the Basal Lamina

The basal lamina is synthesized by the cells on each side of it: the epithelial cells contribute one set of basal lamina components, while cells of the underlying bed of connective tissue (called the *stroma*, Greek for "bedding") contribute another set (**Figure 19–40**). Like other extracellular matrices in animal tissues, the basal lamina consists of two main classes of extracellular macromolecules: (1) fibrous proteins (usually glycoproteins, which have short oligosaccharide side chains attached) and (2) polysaccharide chains of the type called *glycosaminoglycans (GAGs)*, which are usually found covalently linked to specific *core proteins* to form *proteoglycans* (**Figure 19–41**). In a later section, we shall discuss these two large and varied classes of matrix molecules in greater detail. We introduce them here through the special subset that are found in basal laminae.

Although the precise composition of the mature basal lamina varies from tissue to tissue and even from region to region in the same lamina, it typically contains the glycoproteins *laminin*, *type IV collagen*, and *nidogen*, along with the proteoglycan *perlecan*. Together with these key components, present in the basal laminae of virtually all animals from jellyfish to mammals, it holds in its meshes, or is closely associated with, various other molecules. These include *collagen XVIII* (an atypical member of the collagen family, forming the core protein of a proteoglycan) and *fibronectin*, a fibrous protein important in the adhesion of connective-tissue cells to matrix.

The laminin is thought to be the primary organizer of the sheet structure, and early in development, basal laminae consist mainly of laminin molecules. **Laminin-1** (classical laminin) is a large, flexible protein composed of three very

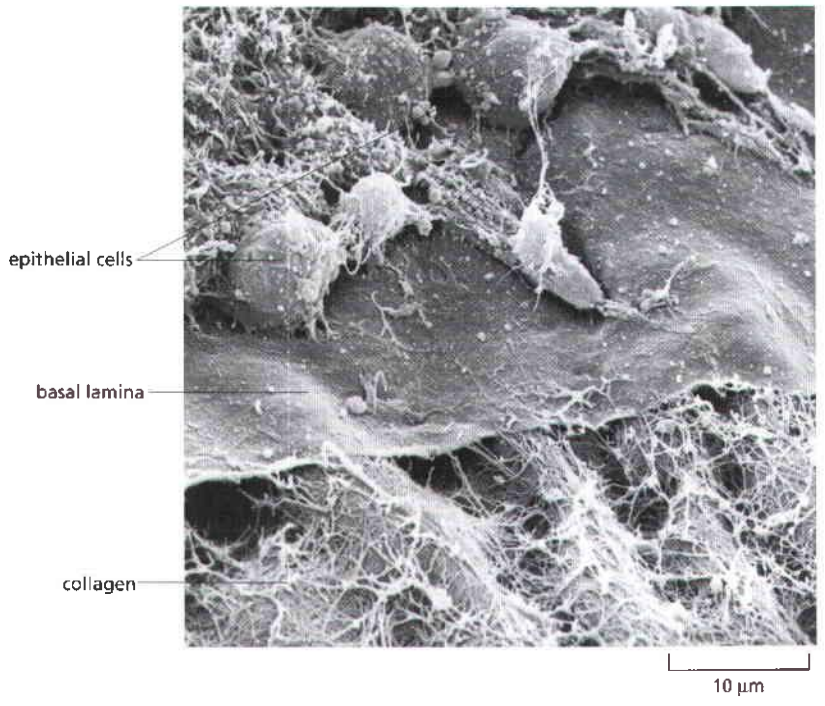


Figure 19–40 The basal lamina in the cornea of a chick embryo. In this scanning electron micrograph, some of the epithelial cells have been removed to expose the upper surface of the matlike basal lamina. A network of collagen fibrils in the underlying connective tissue interacts with the lower face of the lamina. (Courtesy of Robert Trelstad.)

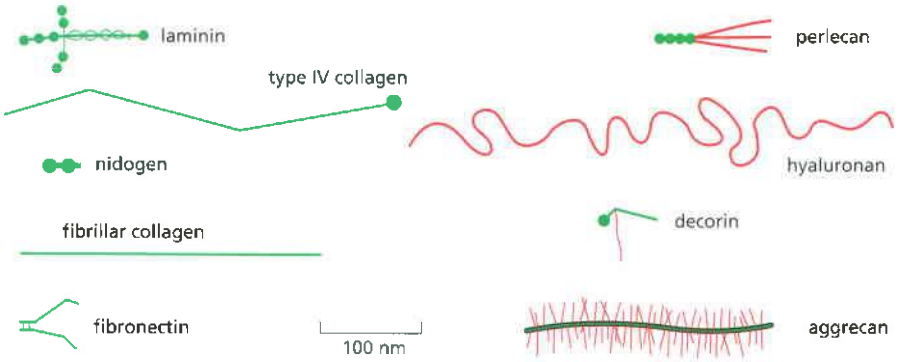


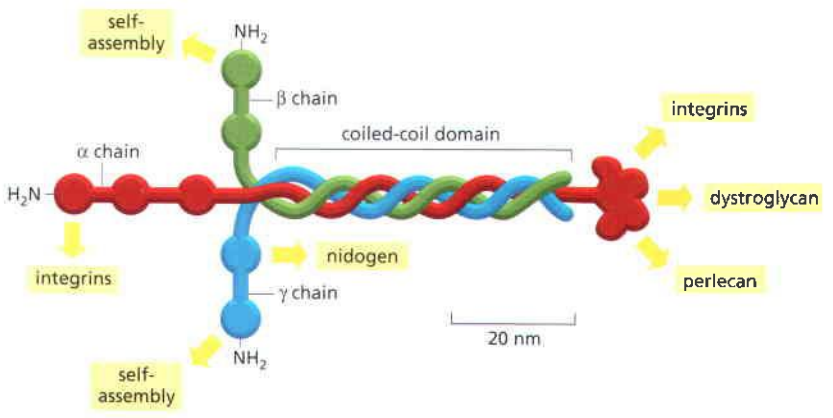
Figure 19–41 The comparative shapes and sizes of some of the major extracellular matrix macromolecules. Protein is shown in *green*, and glycosaminoglycan in *red*.

long polypeptide chains (α, β, and γ) held together by disulfide bonds and arranged in the shape of an asymmetric bouquet, like a bunch of three flowers whose stems are twisted together at the foot but whose heads remain separate (**Figure 19–42**). These heterotrimers can self-assemble *in vitro* into a network, largely through interactions between their heads, although interaction with cells is needed to organize the the network into an orderly sheet. Since there are several isoforms of each type of chain, and these can associate in different combinations, many different laminins can be produced, creating basal laminae with distinctive properties. The laminin γ-1 chain is, however, a component of most laminin heterotrimers; mice lacking it die during embryogenesis because they are unable to make basal lamina.

Type IV Collagen Gives the Basal Lamina Tensile Strength

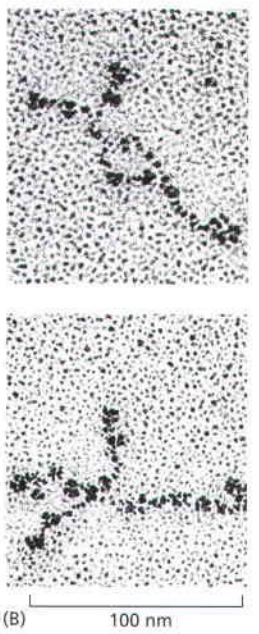
Type IV collagen is a second essential component of mature basal laminae, and it, too, exists in several isoforms. Like the *fibrillar collagens* that constitute the bulk of the protein in connective tissues such as bone or tendon (discussed later), type IV collagen molecules consist of three separately synthesized long protein chains that twist together to form a ropelike superhelix; but they differ from the fibrillar collagens in that the triple-stranded helical structure is interrupted in more than 20 regions, allowing multiple bends. The type IV collagen molecules interact via their terminal domains to assemble extracellularly into a flexible, felt-like network. In this way, type IV collagen gives the basal lamina tensile strength.

But how do the networks of laminin and type IV collagen bond to one another and to the surfaces of the cells that sit on the basal lamina? Why do they form a two-dimensional sheet, rather than a three-dimensional gel? The molecules of laminin have several functional domains, including one that binds to the perlecan proteoglycan, one that binds to the nidogen protein, and two or more that bind to laminin receptor proteins on the surface of cells. Type IV collagen also has domains that bind nidogen and perlecan. It is thought, therefore,

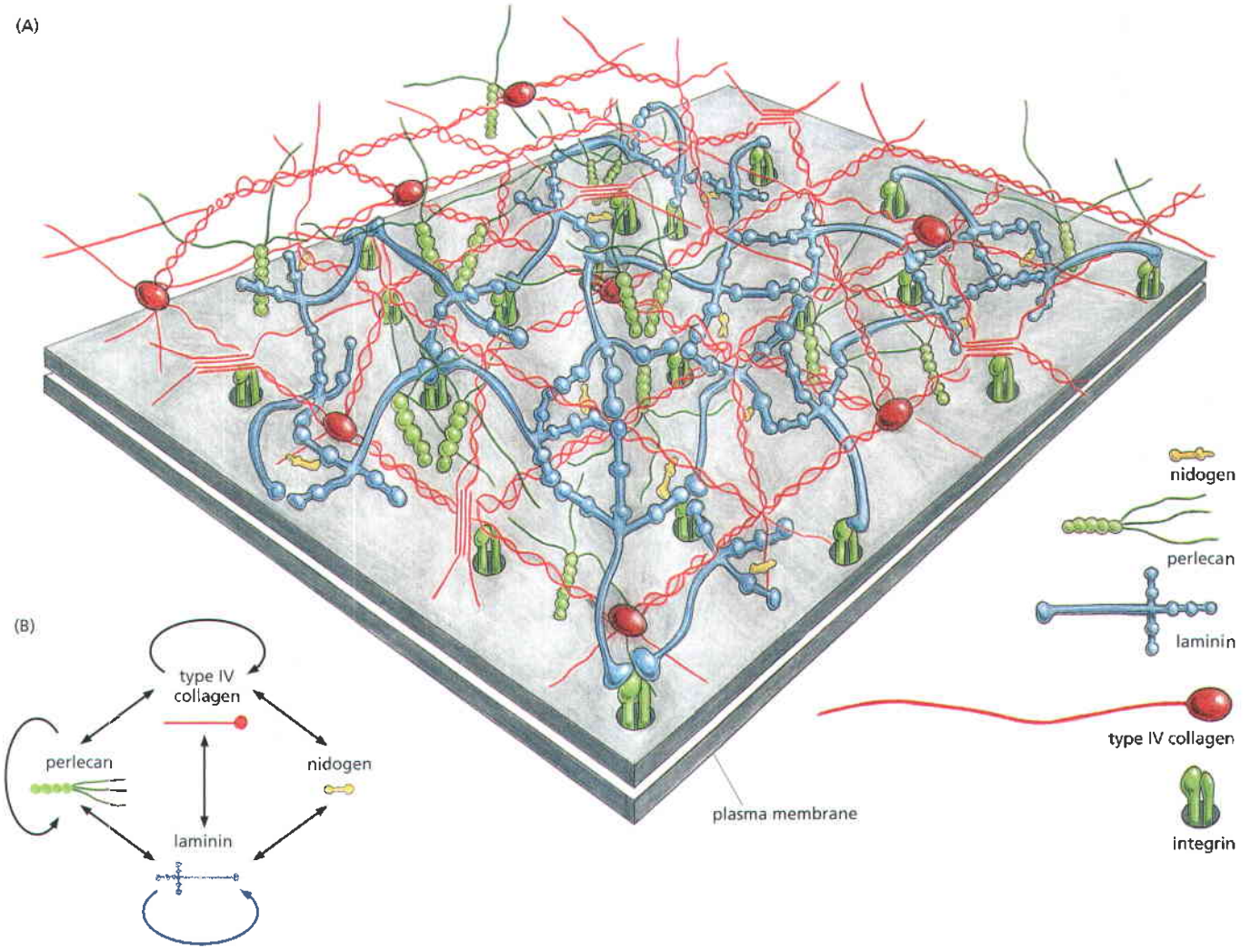


(A)

Figure 19–42 The structure of laminin. (A) The subunits of a laminin-1 molecule, and some of their binding sites for other molecules *(yellow boxes)*. Laminin is a multidomain glycoprotein composed of three polypeptides (α, β, and γ) that are disulfide-bonded into an asymmetric crosslike structure. Each of the polypeptide chains is more than 1500 amino acids long. Five types of α chains, three types of β chains, and three types of γ chains are known; in principle, they can assemble to form 45 (5 × 3 × 3) laminin isoforms. Several such isoforms have been found, each with a characteristic tissue distribution. Through their binding sites for other proteins, laminin molecules play a central part in organizing the assembly of basal laminas and anchoring them to cells. (B) Electron micrographs of laminin molecules shadowed with platinum. (B, from J. Engel et al., *J. Mol. Biol.* 150:97–120, 1981. With permission from Academic Press.)



(B)



that nidogen and perlecan serve as linkers to connect the laminin and type IV collagen networks once the laminin is in place (**Figure 19–43**).

The laminin molecules that generate the initial sheet structure first join to each other while bound to receptors on the surface of the cells that produce them. The cell-surface receptors are of several sorts. Many of them are members of the integrin family; another important type of laminin receptor is *dystroglycan*, a proteoglycan with a core protein that spans the cell membrane, dangling its glycosaminoglycan polysaccharide chains in the extracellular space. Together, these receptors organize basal lamina assembly: they hold the laminin molecules by their feet, leaving the laminin heads positioned to interact so as to form a two-dimensional network. This laminin network then presumably coordinates the assembly of the other basal lamina components.

Figure 19–43 A model of the molecular structure of a basal lamina. (A) The basal lamina is formed by specific interactions (B) between the proteins laminin, type IV collagen, and nidogen, and the proteoglycan perlecan. *Arrows* in (B) connect molecules that can bind directly to each other. There are various isoforms of type IV collagen and laminin, each with a distinctive tissue distribution. Transmembrane laminin receptors (integrins and dystroglycan) in the plasma membrane are thought to organize the assembly of the basal lamina; only the integrins are shown. (Based on H. Colognato and P.D. Yurchenco, *Dev. Dyn.* 218:213–234, 2000. With permission from Wiley-Liss.)

Basal Laminae Have Diverse Functions

As we have mentioned, in the kidney glomerulus, an unusually thick basal lamina acts as one of the layers of a molecular filter, helping to prevent the passage of macromolecules from the blood into the urine as urine is formed (see Figure 19–39). The proteoglycan in the basal lamina seems to be important for this function: when its GAG chains are removed by specific enzymes, the filtering properties of the lamina are destroyed. Type IV collagen also has a role: in a human hereditary kidney disorder *(Alport syndrome)*, mutations in type IV collagen genes result in an irregularly thickened and dysfunctional glomerular filter. Laminin mutations, too, can disrupt the function of the kidney filter, but in a different way—by interfering with the differentiation of the cells that contact it and support it.

The basal lamina can act as a selective barrier to the movement of cells, as well as a filter for molecules. The lamina beneath an epithelium, for example, usually prevents fibroblasts in the underlying connective tissue from making contact with the epithelial cells. It does not, however, stop macrophages, lymphocytes, or nerve processes from passing through it, using specialized protease enzymes to cut a hole for their transit. The basal lamina is also important in tissue regeneration after injury. When cells in tissues such as muscles, nerves, and epithelia are damaged or killed, the basal lamina often survives and provides a scaffold along which regenerating cells can migrate. In this way, the original tissue architecture is readily reconstructed.

A particularly striking example of the role of the basal lamina in regeneration comes from studies on the *neuromuscular junction*, the site where the nerve terminals of a motor neuron form a chemical synapse with a skeletal muscle cell (discussed in Chapter 11). In vertebrates, the basal lamina that surrounds the muscle cell separates the nerve and muscle cell plasma membranes at the synapse, and the synaptic region of the lamina has a distinctive chemical character, with special isoforms of type IV collagen and laminin and a proteoglycan called *agrin*.

This basal lamina at the synapse has a central role in reconstructing the synapse after nerve or muscle injury. If a frog muscle and its motor nerve are destroyed, the basal lamina around each muscle cell remains intact and the sites of the old neuromuscular junctions are still recognizable. If the motor nerve, but not the muscle, is allowed to regenerate, the nerve axons seek out the original synaptic sites on the empty basal lamina and differentiate there to form normal-looking nerve terminals. Thus, the junctional basal lamina by itself can guide the regeneration of motor nerve terminals.

Similar experiments show that the basal lamina also controls the localization of the acetylcholine receptors that cluster in the muscle cell plasma membrane at a neuromuscular junction. If the muscle and nerve are both destroyed, but now the muscle is allowed to regenerate while the nerve is prevented from doing so, the acetylcholine receptors synthesized by the regenerated muscle localize predominantly in the region of the old junctions, even though the nerve is absent (**Figure 19–44**). Thus, the junctional basal lamina apparently coordinates the local spatial organization of the components in each of the two cells that form a neuromuscular junction. Some of the molecules responsible for these effects have been identified. Motor neuron axons, for example, deposit agrin in the junctional basal lamina, where it regulates the assembly of acetyl-

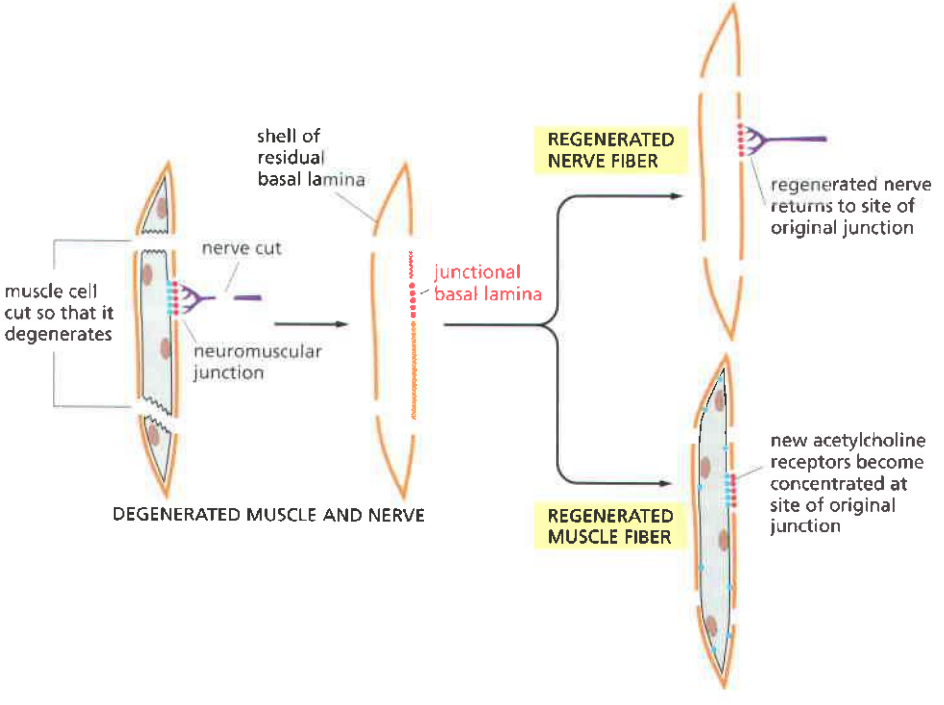


Figure 19–44 Regeneration experiments demonstrating the special character of the junctional basal lamina at a neuromuscular junction. When the nerve, but not the muscle, is allowed to regenerate after both the nerve and muscle have been damaged (*upper part* of figure), the junctional basal lamina directs the regenerating nerve to the original synaptic site. When the muscle, but not the nerve, is allowed to regenerate (*lower part* of figure), the junctional basal lamina causes newly made acetylcholine receptors *(blue)* to accumulate at the original synaptic site. The muscle regenerates from satellite cells (discussed in Chapter 23) located between the basal lamina and the original muscle cell (not shown). These experiments show that the junctional basal lamina controls the localization of synaptic components on both sides of the lamina.

choline receptors and other proteins in the junctional plasma membrane of the muscle cell. Reciprocally, muscle cells deposit a particular isoform of laminin in the junctional basal lamina, and some evidence suggests that this binds directly to the extracellular domain of voltage-gated Ca^{2+} channels in the presynaptic membrane of the nerve cell, helping to hold them at the synapse where they are needed. Both agrin and the synaptic isoform of laminin are essential for the formation of normal neuromuscular junctions. Defects in components of the basal lamina or in proteins that tether muscle cell components to it at the synapse are responsible for many of the forms of muscular dystrophy, in which muscles at first develop normally but then degenerate in later years of life.

Summary

The basal lamina is a thin tough sheet of extracellular matrix that closely underlies epithelia in all multicellular animals. It also wraps around certain other cell types, such as muscle cells. All basal laminae are organized on a framework of laminin molecules, linked together by their side-arms and held close beneath the basal ends of the epithelial cells by attachment to integrins and other receptors in the basal plasma membrane. Type IV collagen molecules are recruited into this structure, assembling into a sheetlike mesh that is an essential component of all mature basal laminae. The collagen and laminin networks in mature basal laminae are bridged by the protein nidogen and the large heparan sulfate proteoglycan perlecan.

Basal laminae provide mechanical support for epithelia; they form the interface and the attachment between epithelia and connective tissue; they serve as filters in the kidney; they act as barriers to keep cells in their proper compartments; they influence cell polarity and cell differentiation; they guide cell migrations; and molecules embedded in them help to organize elaborate structures such as neuromuscular synapses. When cells are damaged or killed, basal laminae often survive and can help guide tissue regeneration.

INTEGRINS AND CELL–MATRIX ADHESION

Cells make extracellular matrix, organize it, and degrade it. The matrix in its turn exerts powerful influences on the cells. The influences are exerted chiefly through transmembrane cell adhesion proteins that act as *matrix receptors*. These tie the matrix outside the cell to the cytoskeleton inside it, but their role goes far beyond simple passive mechanical attachment. Through them, components of the matrix can affect almost any aspect of a cell's behavior. The matrix receptors have a crucial role in epithelial cells, mediating their interactions with the basal lamina beneath them; and they are no less important in connective-tissue cells, for their interactions with the matrix that surrounds them.

Several types of molecules can function as matrix receptors or co-receptors, including the transmembrane proteoglycans. But the principal receptors on animal cells for binding most extracellular matrix proteins are the **integrins**. Like the cadherins and the key components of the basal lamina, integrins are part of the fundamental architectural toolkit that is characteristic of multicellular animals. The members of this large family of homologous transmembrane adhesion molecules have a remarkable ability to transmit signals in both directions across the cell membrane. The binding of a matrix component to an integrin can send a message into the interior of the cell, and conditions in the cell interior can send a signal outward to control binding of the integrin to matrix (or, in some cases, to a cell-surface molecule on another cell, as we saw in the case of white blood cells binding to endothelial cells). Tension applied to an integrin can cause it to tighten its grip on intracellular and extracellular structures, and loss of tension can loosen its hold, so that molecular signaling complexes fall apart on either side of the membrane. In this way, integrins can also serve not only to transmit mechanical and molecular signals, but also to convert the one type of signal into the other. Studies of the structure of integrin molecules have begun to reveal how they perform these tasks.

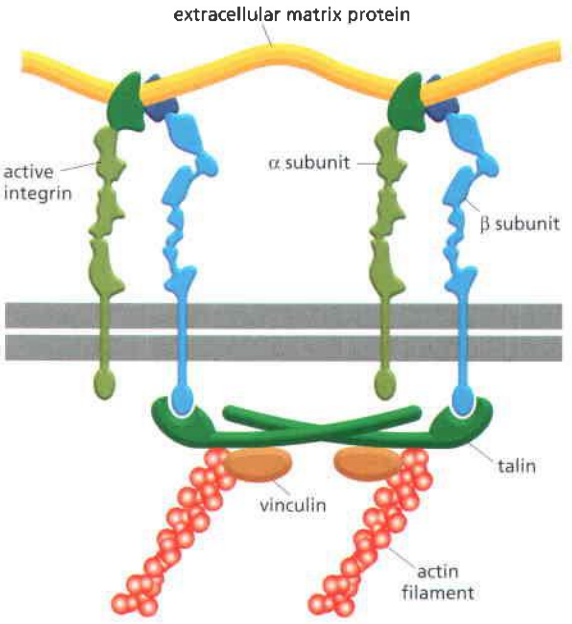


Figure 19–45 The subunit structure of an active integrin molecule, linking extracellular matrix to the actin cytoskeleton. The head of the integrin molecule attaches directly to an extracellular protein such as fibronectin; the intracellular tail of the integrin binds to talin, which in turn binds to filamentous actin. A set of other intracellular anchor proteins, including α-actinin, filamin, and vinculin, help to reinforce the linkage.

Integrins Are Transmembrane Heterodimers That Link to the Cytoskeleton

There are many varieties of integrins—at least 24 in humans—but they all conform to a common plan. An integrin molecule is composed of two noncovalently associated glycoprotein subunits called α and β. Both subunits span the cell membrane, with short intracellular C-terminal tails and large N-terminal extracellular domains. The extracellular portion of the integrin dimer binds to specific amino acid sequences in extracellular matrix proteins such as laminin or fibronectin or, in some cases, to ligands on the surfaces of other cells. The intracellular portion binds to a complex of proteins that form a linkage to the cytoskeleton.

For all but one of the 24 varieties of human integrins, this intracellular linkage is to actin filaments, via *talin* and a set of other intracellular anchorage proteins (**Figure 19–45**); talin, as we shall see later, seems to be the key component of the linkage. Like the actin-linked cell–cell junctions formed by cadherins, the actin-linked cell–matrix junctions formed by integrins may be small, inconspicuous and transient, or large, prominent, and durable. Examples of the latter are the *focal adhesions* that form when fibroblasts have sufficient time to form strong attachments to the rigid surface of a culture dish, and the *myotendinous junctions* that attach muscle cells to their tendons.

In epithelia, the most prominent cell–matrix attachment sites are the hemidesmosomes, where a specific type of integrin (α6β4) anchors the cells to laminin in the basal lamina. Here, uniquely, the intracellular attachment is to keratin filaments, via the intracellular anchor proteins plectin and dystonin (**Figure 19–46**).

Integrins Can Switch Between an Active and an Inactive Conformation

A cell crawling through a tissue—a fibroblast or a macrophage, for example, or an epithelial cell migrating along a basal lamina—has to be able both to make and to break attachments to the matrix, and to do so rapidly if it is to travel quickly. <TGAT> Similarly, a circulating white blood cell has to be able to switch on or off its tendency to bind to endothelial cells in order to crawl out of a blood vessel at a site of inflammation under the appropriate circumstances. Furthermore, if

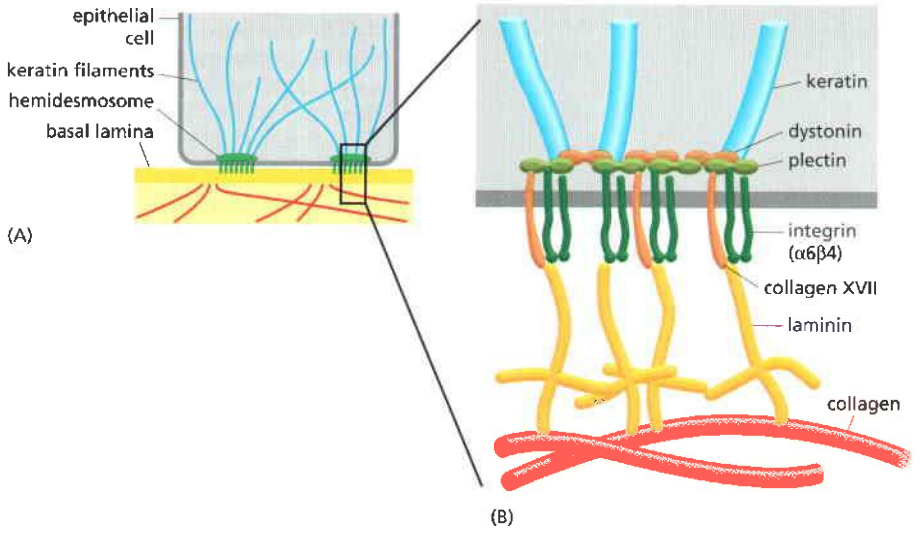


Figure 19–46 Hemidesmosomes. (A) Hemidesmosomes spot-weld epithelial cells to the basal lamina, linking laminin outside the cell to keratin filaments inside it. (B) Molecular components of a hemisdesmosome. A specialized integrin (α6β4 integrin) spans the membrane, attaching to keratin filaments intracellularly via anchor proteins called plectin and dystonin, and to laminin extracellularly. The adhesive complex also contains, in parallel with the integrin, an unusual collagen family member, collagen type XVII; this has a membrane-spanning domain attached to its extracellular collagenous portion. Defects in any of these components can give rise to a blistering disease of the skin.

force is to be applied where it is needed, the making and breaking of the extracellular attachments in all these cases has to be coupled to the prompt assembly and disassembly of cytoskeletal attachments inside the cell. The integrin molecules that span the membrane and mediate the attachments cannot simply be passive, rigid objects with sticky patches at their two ends. They must be able to switch between an active state, where they readily form attachments, and an inactive state, where they do not; and the binding of their ligands on one side of the membrane must alter their propensity to bind to a different set of ligands on the opposite side.

The basis for these dynamic phenomena is allosteric regulation: as an integrin binds to or detaches from its ligands, it undergoes conformational changes that affect both the intracellular and the extracellular ends of the molecule. Structural change at one end is coupled to structural change at the other, so that influences can be transmitted in either direction across the cell membrane. The timber tongs that lumberjacks use to grab hold of logs of wood provide a simple mechanical analogy (**Figure 19–47**).

The structural changes in integrins can be demonstrated by taking a purified preparation of integrin molecules and examining them at high resolution by electron microscopy. If the integrins are kept in a calcium-rich medium similar to normal extracellular fluid, but without any extracellular ligand, and then rapidly prepared for microscopy, they appear as tightly folded V-shaped objects. But if a small synthetic peptide containing a sequence that mimics the integrin-binding domain of a natural extracellular matrix protein is added to the medium, the integrins bind this molecule and extend into a different shape, with two legs that are no longer tightly bent, but are now straightened and separated from each other, supporting a head region high above them (**Figure 19–48A**). This pair of structures can be compared with more detailed data from x-ray crystallography, which reveals that the two legs correspond to the integrin α and β chains. The head region, where they meet, contains the binding site for the extracellular ligand. Binding of the ligand distorts this region so as to favor adoption of the extended, "active" conformation; conversely, adoption of the extended conformation creates a more favorable binding site, with a higher affinity for ligand (Figure 19–48B).

But how do these changes in the extracellular region of the integrin relate to events at the intracellular end of the integrin molecule? In its folded, inactive



Figure 19–47 Timber tongs. Holding the handles together causes the claws to grip the log; and closing the claws on the log causes the handles to come together. Moreover, the greater the pull on the tongs, the tighter the grip at both ends. In an integrin molecule, the details of the linkage are different, but the mechanical principles are similar: conformational changes at opposite ends of the molecule are coupled, and pulling tightens the grip.

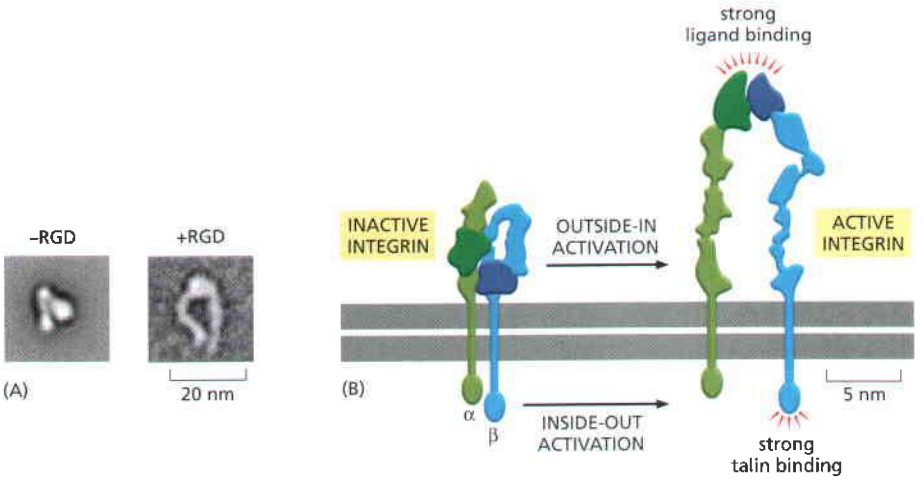


Figure 19–48 Change in conformation of an integrin molecule when it binds its ligand. (A) Images were produced by averaging many similarly aligned electron micrographs of individual integrin molecules. In the absence of extracellular ligand, the integrin molecules appear small and tightly folded. When incubated with an RGD peptide, the integrins unfold into an extended structure with two distinct legs. (B) Active (extended) and inactive (folded) structures of an integrin molecule, based on data from x-ray crystallography. Although it is difficult to crystallize the intact molecule in its natural conformations, with and without ligand bound, the complete structure can be inferred with reasonable confidence from x-ray crystallography of defined molecular fragments. (A, From J. Takagi et al., *Cell* 110:599–611, 2002. With permission from Elsevier; B, based on T. Xiao at al., *Nature* 432:59–67, 2004. With permission from Macmillan Publishers Ltd.)

state, the intracellular portions of its α and β chains lie close together and adhere to one another. When the extracellular domain unfolds, this contact is broken and the intracellular (and transmembrane) portions of these chains move apart. As a result, a binding site for talin on the tail of the β chain is exposed. The binding of talin then leads to assembly of actin filaments anchored to the intracellular end of the integrin molecule (see Figure 19–45). In this way, when an integrin catches hold of its ligand outside the cell, the cell reacts by tying its cytoskeleton to the integrin molecule, so that force can be applied at the point of attachment. This is referred to as "outside-in activation".

The chain of cause and effect can also operate in reverse, from inside to outside instead of outside to inside. Talin competes with the integrin α chain for its binding site on the tail of the β chain. Thus when talin binds to the β chain it undoes the intracellular α–β linkage, allowing the two legs of the integrin molecule to spring apart. This drives the extracellular portion of the integrin into its extended, active conformation.

This "inside-out activation" is triggered by intracellular regulatory molecules. These include the phosphoinositide PIP_2 (discussed in Chapter 15), which is thought to be capable of activating talin so that it binds to the integrin β chain strongly. In this way, a signal generated inside the cell can trigger its integrin molecules to reach out and grab hold of their extracellular ligands.

Intracellular signal molecules such as PIP_2 are themselves produced in response to signals received from outside the cell via other types of cell-surface receptors, such as G-protein-coupled receptors and receptor tyrosine kinases, which can thus control integrin activation (**Figure 19–49**). Conversely, the activation of integrins by attachment to matrix can influence the reception of signals by other pathways. The cross-talk between all these communication pathways, transmitting signals in both directions across the cell membrane, allows for some complex interactions between the cell and its physical and chemical environment.

Integrin Defects Are Responsible for Many Different Genetic Diseases

The 24 types of integrins found in a human are formed from the products of 8 different β-chain genes and 18 different α-chain genes, dimerized in different combinations. Each integrin has distinctive properties and functions. Moreover, because the same integrin molecule in different cell types can have different ligand-binding specificities, it seems that additional cell-type-specific factors can interact with integrins to modulate their binding activity. The binding of integrins to their ligands is also affected by the concentration of Ca^{2+} and Mg^{2+} in the extracellular medium, reflecting the presence of divalent-cation-binding domains in the α and β subunits. The divalent cations can influence both the affinity and the specificity of the binding of an integrin to its ligands.

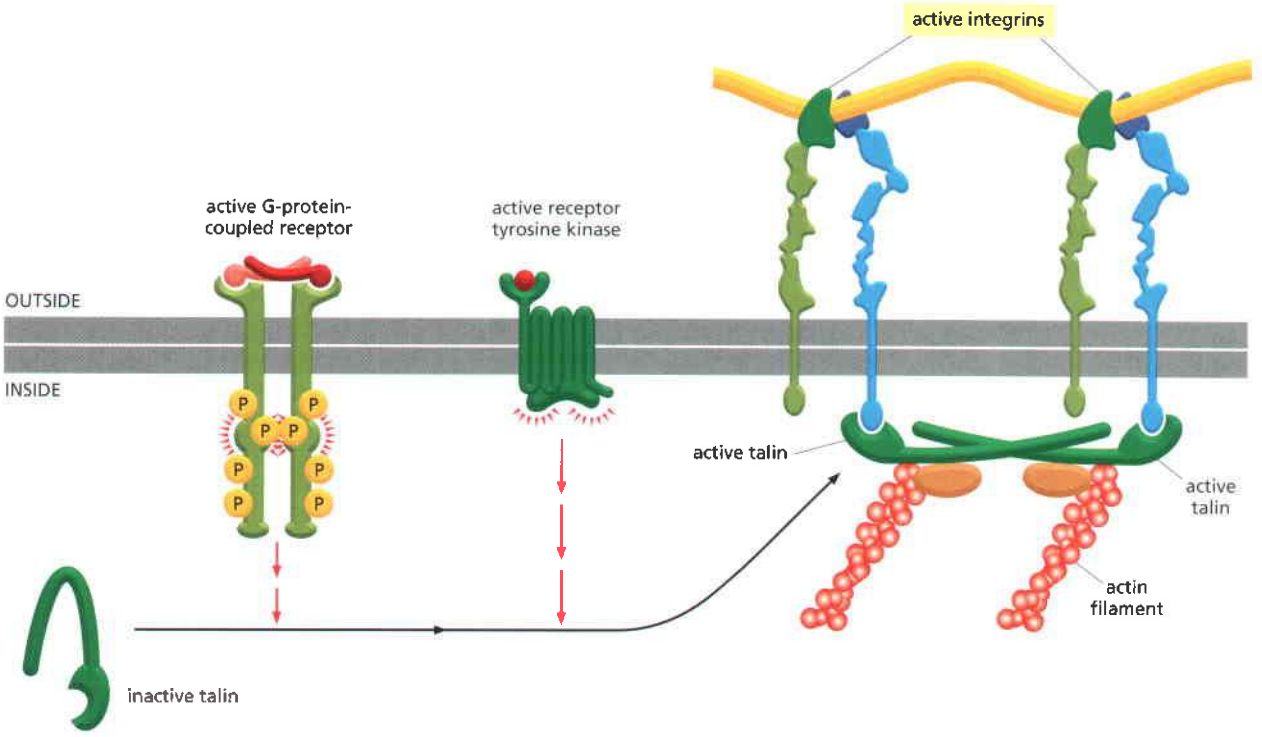


Figure 19–49 Activation of integrins by cross-talk from other signaling pathways. Signals received from outside the cell via other types of cell surface receptors, such as G-protein-coupled receptors and receptor tyrosine kinases, can alter the conformation of talin and thereby activate the cell's integrins.

Although there is some overlap in the activities of the different integrins—at least five bind laminin, for example—it is the diversity of integrin functions that is more remarkable. **Table 19–4** lists some of varieties of integrins and the problems that result when individual integrin α or β chains are defective.

The β1 subunits form dimers with at least 12 distinct α subunits and are found on almost all vertebrate cells: α5β1 is a fibronectin receptor and α6β1 a laminin receptor on many types of cells. Mutant mice that cannot make any β1 integrins die at implantation (very early in embryonic development). Mice that are only unable to make the α7 subunit (the partner for β1 in muscle) survive but develop muscular dystrophy (as do mice that cannot make the laminin ligand for the α7β1 integrin).

Table 19–4 Some Types of Integrins

INTEGRIN	LIGAND*	DISTRIBUTION	PHENOTYPE WHEN α SUBINUT IS MUTATED	PHENOTYPE WHEN β SUBUNIT IS MUTATED
α5β1	fibronectin	ubiquitous	death of embryo; defects in blood vessels, somites, neural crest	early death of embryo (at implantation)
α6β1	laminin	ubiquitous	severe skin blistering; defects in other epithelia also	early death of embryo (at implantation)
α7β1	laminin	muscle	muscular dystrophy; defective myotendinous junctions	early death of embryo (at implantation)
αLβ2 (LFA1)	Ig superfamily counterreceptors (ICAM)	white blood cells	impaired recruitment of leucocytes	leucocyte adhesion deficiency (LAD) impaired inflammatory responses; recurrent life-threatening infections
αIIbβ3	fibrinogen	platelets	bleeding; no platelet aggregation (Glanzmann's disease)	bleeding; no platelet aggregation (Glanzmann's disease); mild osteopetrosis
α6β4	laminin	hemidesmosomes in epithelia	severe skin blistering; defects in other epithelia also	severe skin blistering; defects in other epithelia also

*Not all ligands are listed.

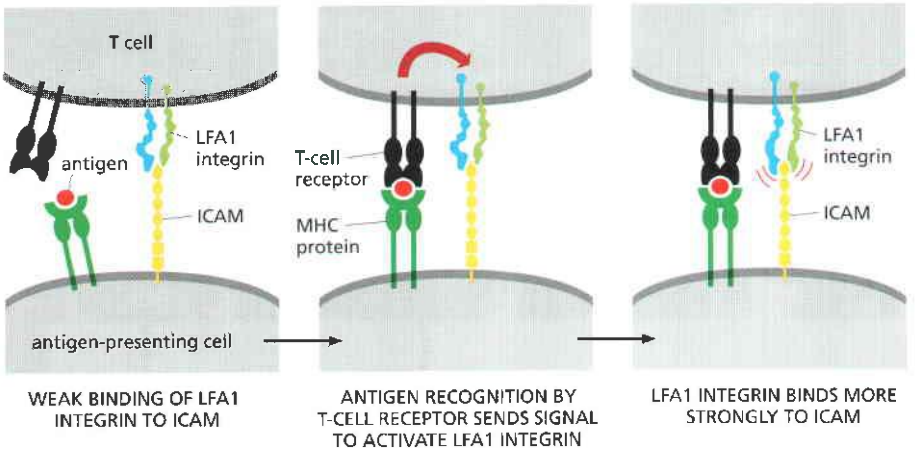


Figure 19–50 Integrin activation in the encounter of a T lymphocyte with an antigen-presenting cell. The two cells at first adhere weakly through binding of the LFA1 integrin in the T cell to the Ig-superfamily molecule ICAM in the membrane of the antigen-presenting cell. If the T-cell receptor at the same time recognizes its specific antigen, presented to it by the MHC molecule on the antigen-presenting cell, an intracellular signal is generated from the T cell receptor to activate the LFA1 integrin. As a result, LFA1 binds more strongly and persistently to ICAM. This gives the antigen-presenting cell time to activate the T cell and thereby elicit a specific immune response. (Adapted from K. Murphy et al., Janeway's Immunobiology, 7th ed. New York: Garland Science, 2008.)

The β2 subunits form dimers with at least four types of α subunit and are expressed exclusively on the surface of white blood cells, where they have an essential role in enabling these cells to fight infection. The β2 integrins mainly mediate cell–cell rather than cell–matrix interactions, binding to specific ligands on another cell, such as an endothelial cell. The ligands, sometimes referred to as *counterreceptors*, are members of the Ig superfamily of cell–cell adhesion molecules. We have already described an example earlier in the chapter: an integrin of this class (αLβ2, also known as LFA1) on white blood cells enables them to attach firmly to the Ig-family protein ICAM on endothelial cells at sites of infection and, through this attachment, to migrate out of the bloodstream into the infected tissue (see Figure 19–19B). People with the genetic disease called *leucocyte adhesion deficiency* fail to synthesize functional β2 subunits. As a consequence, their white blood cells lack the entire family of β2 receptors, and they suffer repeated bacterial infections.

The β3 integrins are found on blood platelets (as well as various other cells), and they bind several matrix proteins, including the blood clotting factor *fibrinogen*. Platelets have to interact with fibrinogen to mediate normal blood clotting, and humans with *Glanzmann's disease*, who are genetically deficient in β3 integrins, suffer from defective clotting and bleed excessively.

In both white blood cells and platelets, the ability to regulate integrin activity via inside-out signaling is particularly important. Regulated adhesion allows the cells to circulate unimpeded until they are activated by an appropriate stimulus. Because the integrins do not need to be synthesized *de novo*, the signaled adhesion response can be rapid. Platelets, for example, respond to contact with the wall of a damaged blood vessel and to various soluble signaling molecules, triggering activation of the β3 integrin in the platelet membrane. The resulting interaction of platelets with fibrinogen leads to formation of a platelet plug, which helps to stop the bleeding at just the site where it is needed. Similarly, the binding of a T lymphocyte to its specific antigen on the surface of an antigen-presenting cell (discussed in Chapter 25) switches on intracellular signaling pathways in the T cell that activate its β2 integrins (**Figure 19–50**). The activated integrins then enable the T cell to adhere strongly to the antigen-presenting cell so that it remains in contact long enough to become stimulated fully. The integrins may then return to an inactive state, allowing the T cell to disengage.

Integrins Cluster to Form Strong Adhesions

Integrins, like other cell adhesion molecules, differ from cell-surface receptors for hormones and for other extracellular soluble signal molecules in that they usually bind their ligand with lower affinity and are usually present at a 10- to 100-fold higher concentration on the cell surface. The Velcro principle, mentioned earlier, operates here too. Strong adhesion depends on clustering of

integrins, creating a plaque in which many cytoskeletal filaments are anchored, as at a hemidesmosome in the epidermis or at a focal adhesion made by a fibroblast on a culture dish. At focal adhesions, and probably also in the less prominent actin-linked cell–matrix adhesions that cells mainly make in normal tissues, activation of the small GTPase Rho plays a part in the maturation of the adhesive complex, by promoting recruitment of actin filaments and integrins to the contact site. Artificially mutated integrins that lack an intracellular tail fail to connect with cytoskeletal filaments, fail to cluster, and are unable to form strong adhesions.

Extracellular Matrix Attachments Act Through Integrins to Control Cell Proliferation and Survival

Like other transmembrane cell adhesion proteins, integrins do more than just create attachments. They also activate intracellular signaling pathways and thereby allow control of almost any aspect of the cell's behavior according to the nature of the surrounding matrix and the state of the cell's attachments to it.

Studies in culture show that many cells will not grow or proliferate unless they are attached to extracellular matrix; nutrients and soluble growth factors in the culture medium are not enough. For some cell types, including epithelial, endothelial, and muscle cells, even cell survival depends on such attachments. When these cells lose contact with the extracellular matrix, they undergo programmed cell death, or apoptosis. This dependence of cell growth, proliferation, and survival on attachment to a substratum is known as **anchorage dependence**, and it is mediated mainly by integrins and the intracellular signals they generate. Anchorage dependence is thought to help ensure that each type of cell survives and proliferates only when it is in an appropriate situation. Mutations that disrupt or override this form of control, allowing cells to escape from anchorage dependence, occur in cancer cells and play a major part in their invasive behavior.

The physical spreading of a cell on the matrix also has a strong influence on intracellular events. Cells that are forced to spread over a large surface area by the formation of multiple adhesions at widely separate sites survive better and proliferate faster than cells that are not so spread out (**Figure 19–51**). The stimulatory effect of cell spreading presumably helps tissues to regenerate after injury. If cells are lost from an epithelium, for example, the spreading of the remaining cells into the vacated space will help stimulate these survivors to proliferate until they fill the gap. It is uncertain how a cell senses its extent of spreading so as to adjust its behavior accordingly, but the ability to spread depends on integrins, and signals generated by integrins at the sites of adhesion must play a part in providing the spread cells with stimulation.

Our understanding of anchorage dependence and of the effects of cell spreading has come mainly from studies of cells living on the surface of matrix-coated culture dishes. For connective-tissue cells that are normally surrounded

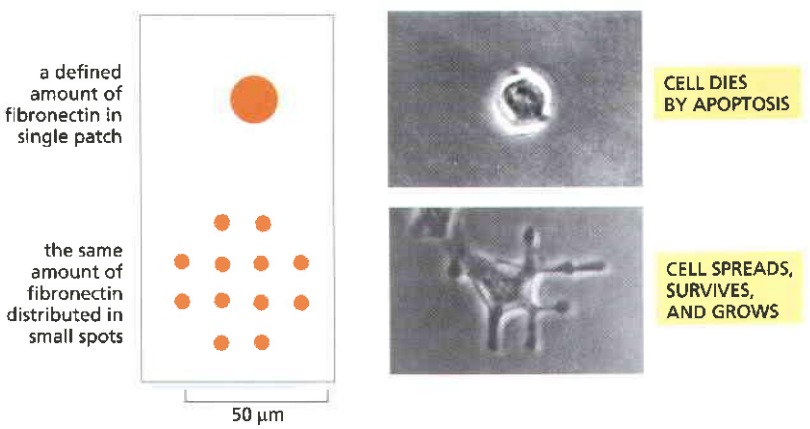


Figure 19–51 The importance of cell spreading. In this experiment, cell growth and survival are shown to depend on the extent of cell spreading on a substratum, rather than the mere fact of attachment or the number of matrix molecules the cell contacts. (Based on C.S. Chen et al., *Science* 276:1425–1428, 1997. With permission from AAAS.)

by matrix on all sides, this is a far cry from the natural environment. Walking over a plain is very different from clambering through a jungle. The types of contacts that cells make with a rigid substratum are not the same as those, much less well studied, that they make with the deformable web of fibers of the extracellular matrix, and there are substantial differences of cell behavior between the two contexts. Nevertheless, it is likely that the same basic principles apply. Both *in vitro* and *in vivo*, intracellular signals generated at cell–matrix adhesion sites, by molecular complexes organized around integrins, are crucial for cell proliferation and survival.

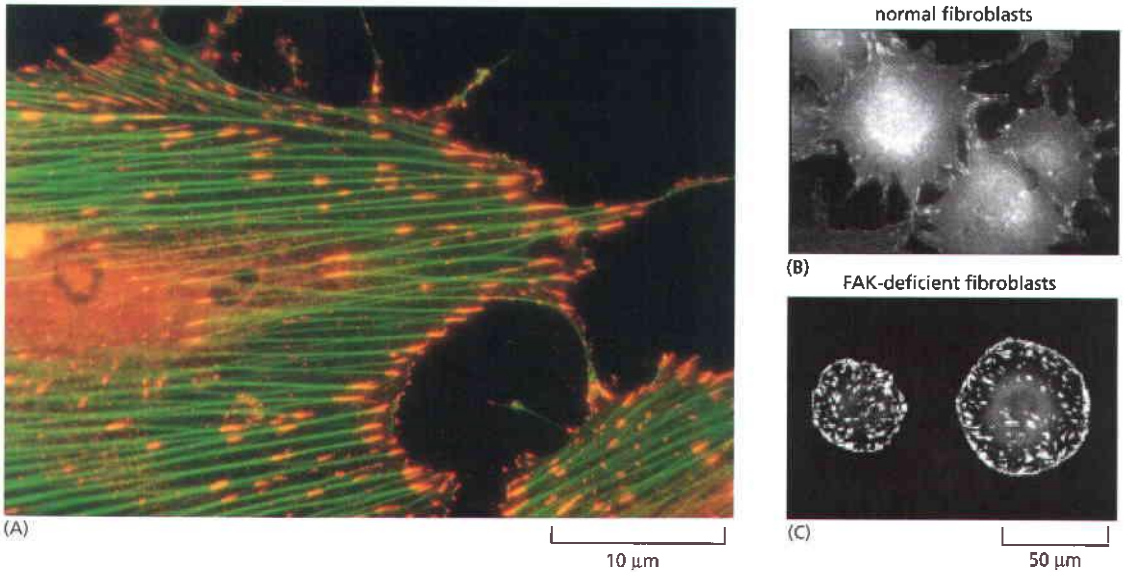
Integrins Recruit Intracellular Signaling Proteins at Sites of Cell–Substratum Adhesion

The mechanisms by which integrins signal into the cell interior are complex, involving several different pathways, and integrins and conventional signaling receptors often influence one another and work together to regulate cell behavior, as we have already emphasized. The Ras/MAP kinase pathway (see Figure 15–61), for example, can be activated both by conventional signaling receptors and by integrins, but cells often need both kinds of stimulation of this pathway at the same time to give sufficient activation to induce cell proliferation. Integrins and conventional signaling receptors also cooperate in activating similar pathways to promote cell survival (discussed in Chapters 15 and 17).

One of the best-studied modes of integrin signaling depends on a cytoplasmic protein tyrosine kinase called **focal adhesion kinase** (**FAK**). In studies of cells cultured in the normal way on rigid substrata, focal adhesions are often prominent sites of tyrosine phosphorylation (**Figure 19–52**A), and FAK is one of the major tyrosine-phosphorylated proteins found at these sites. When integrins cluster at cell–matrix contacts, FAK is recruited by intracellular anchor proteins such as talin (binding to the integrin β subunit) or *paxillin* (which binds to one type of integrin α subunit). The clustered FAK molecules cross-phosphorylate each other on a specific tyrosine, creating a phosphotyrosine docking site for members of the Src family of cytoplasmic tyrosine kinases. In addition to phosphorylating other proteins at the adhesion sites, these kinases then phosphorylate FAK on additional tyrosines, creating docking sites for a variety of additional intracellular signaling proteins. In this way, outside-in signaling from integrins, via FAK and Src-family kinases, is relayed into the cell (as discussed in Chapter 15).

One way to analyze the function of FAK is to examine focal adhesions in cells from mutant mice that lack the protein. FAK-deficient fibroblasts still adhere to

Figure 19–52 Focal adhesions and the role of focal adhesion kinase (FAK). (A) A fibroblast cultured on a fibronectin-coated substratum and stained with fluorescent antibodies: actin filaments are stained *green* and activated proteins that contain phosphotyrosine are *red*, giving *orange* where the two components overlap. The actin filaments terminate at focal adhesions, where the cell attaches to the substratum by means of integrins. Proteins containing phosphotyrosine are also concentrated at these sites, reflecting the local activation of FAK and other protein kinases. Signals generated at such adhesion sites help regulate cell division, growth, and survival. (B, C) The influence of FAK on formation of focal adhesions is shown by a comparison of normal and FAK-deficient fibroblasts, stained with an antibody against vinculin to reveal the focal adhesions. (B) The normal fibroblasts have fewer focal adhesions and have spread after 2 hours in culture. (C) At the same time point, the FAK-deficient fibroblasts have more focal adhesions and have not spread. (A, courtesy of Keith Burridge; B, C, from D. Ilic et al., *Nature* 377:539–544, 1995. With permission from Macmillan Publishers Ltd.)



fibronectin and form focal adhesions. In fact, they form too many focal adhesions; as a result, cell spreading and migration are slowed (Figure 19–52B and C). This unexpected finding suggests that FAK normally helps disassemble focal adhesions and that this loss of adhesions is required for normal cell migration. Many cancer cells have elevated levels of FAK, which may help explain why they are often more motile than their normal counterparts.

Integrins Can Produce Localized Intracellular Effects

Through FAK and other pathways, activated integrins, like other signaling receptors, can induce global cell responses, often including changes in gene expression. But the integrins are especially adept at stimulating localized changes in the cytoplasm close to the cell–matrix contact. We have already mentioned an important example in our discussion of epithelial cell polarity: it is through integrins that the basal lamina plays its part in directing the internal apico-basal organization of epithelial cells.

Localized intracellular effects may be a common feature of signaling by transmembrane cell adhesion proteins in general. In the developing nervous system, for example, the growing tip of an axon is guided mainly by its responses to local adhesive (and repellent) cues in the environment that are recognized by transmembrane cell adhesion proteins, as discussed in Chapter 22. The primary effects of the adhesion proteins are thought to result from the activation of intracellular signaling pathways that act locally in the axon tip, rather than through cell–cell adhesion itself or signals conveyed to the cell body. Through localized activation of the Rho family of small GTPases, for example (as discussed in Chapters 15 and 16), the transmembrane adhesion proteins can control motility and guide forward movement. In this and other ways, practically all the classes of cell-cell and cell-matrix adhesion molecules that we have mentioned, including integrins, are deployed to help guide axon outgrowth in the developing nervous system.

Table 19–5 summarizes the categories of cell adhesion molecules that we have considered in this chapter. In the next section, we turn from the adhesion molecules in cell membranes to look in detail at the extracellular matrix that surrounds cells in connective tissues.

Table 19–5 Cell Adhesion Molecule Families

	SOME FAMILY MEMBERS	Ca^{2+} OR Mg^{2+} DEPENDENCE	HOMOPHILIC OR HETEROPHILIC	CYTOSKELETON ASSOCIATIONS	CELL JUNCTION ASSOCIATIONS
Cell–Cell Adhesion					
Classical cadherins	E, N, P, VE	yes	homophilic	actin filaments (via catenins)	adherens junctions, synapses
Desmosomal cadherins	desmoglein, desmocollin	yes	homophilic	intermediate filaments (via desmoplakin, plakoglobin, and plakophilin)	desmosomes
Ig family members	N-CAM, ICAM	no	both	unknown	neuronal and immunological synapses
Selectins (blood cells and endothelial cells only)	L-, E-, and P-selectins	yes	heterophilic	actin filaments	(no prominent junctional structure)
Integrins on blood cells	αLβ2 (LFA1)	yes	heterophilic	actin filaments	immunological synapses
Cell–Matrix Adhesion					
Integrins	many types	yes	heterophilic	actin filaments (via talin, paxillin, filamin, α-actinin, and vinculin)	focal adhesions
	α6β4	yes	heterophilic	intermediate filaments (via plectin and dystonin)	hemidesmosomes
Transmembrane proteoglycans	syndecans	no	heterophilic	actin filaments	(no prominent junctional structure)

Summary

Integrins are the principal receptors used by animal cells to bind to the extracellular matrix: they function as transmembrane linkers between the extracellular matrix and the cytoskeleton connecting usually to actin, but to intermediate filaments for the specialized integrins at hemidesmosomes. Integrin molecules are heterodimers, and the binding of ligands is associated with dramatic changes of conformation. This creates an allosteric coupling between binding to matrix outside the cell and binding to the cytoskeleton inside it, allowing the integrin to convey signals in both directions across the plasma membrane—from inside to out and from outside to in. Binding of the intracellular anchor protein talin to the tail of an integrin molecule tends to drive the integrin into an extended conformation with increased affinity for its extracellular ligand. Conversely, binding to an extracellular ligand, by promoting the same conformational change, leads to binding of talin and formation of a linkage to the actin cytoskeleton. Complex assemblies of proteins become organized around the intracellular tails of integrins, producing intracellular signals that can influence almost any aspect of cell behavior, from proliferation and survival, as in the phenomenon of anchorage dependence, to cell polarity and guidance of migration.

THE EXTRACELLULAR MATRIX OF ANIMAL CONNECTIVE TISSUES

We have already discussed the basal lamina as an archetypal example of extracellular matrix, common to practically all multicellular animals and an essential feature of epithelial tissues. We now turn to the much more varied and bulky forms of extracellular matrix found in connective tissues (**Figure 19–53**). Here, the extracellular matrix is generally more plentiful than the cells it surrounds, and it determines the tissue's physical properties.

The classes of macromolecules constituting the extracellular matrix in animal tissues are broadly similar, whether we consider the basal lamina or the other forms that matrix can take, but variations in the relative amounts of these different classes of molecules and in the ways in which they are organized give rise to an amazing diversity of materials. The matrix can become calcified to form the rock-hard structures of bone or teeth, or it can form the transparent substance of the cornea, or it can adopt the ropelike organization that gives tendons their enormous tensile strength. It forms the jelly in a jellyfish. Covering the body of a beetle or a lobster, it forms a rigid carapace. Moreover, the extracellular matrix is more than a passive scaffold to provide physical support. It has an active and complex role in regulating the behavior of the cells that touch it, inhabit it, or crawl through its meshes, influencing their survival, development, migration, proliferation, shape, and function.

In this section, we focus our discussion on the extracellular matrix of connective tissues in vertebrates, but bulky forms of extracellular matrix play an

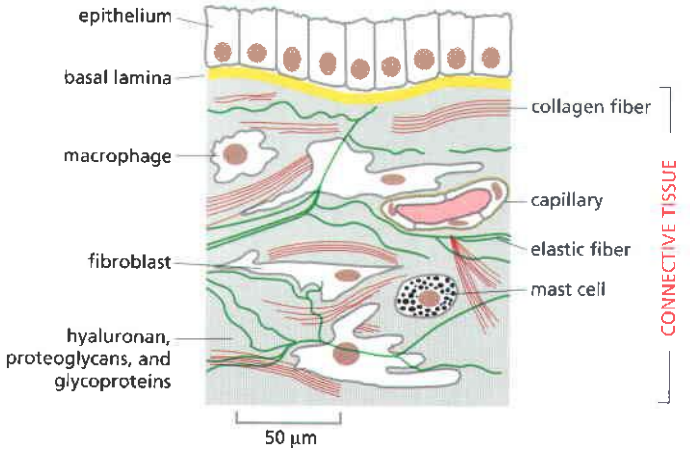


Figure 19–53 The connective tissue underlying an epithelium. This tissue contains a variety of cells and extracellular matrix components. The predominant cell type is the fibroblast, which secretes abundant extracellular matrix.

important part in virtually all multicellular organisms; examples include the cuticles of worms and insects, the shells of mollusks, the cell walls of fungi, and, as we discuss later, the cell walls of plants.

The Extracellular Matrix Is Made and Oriented by the Cells Within It

The macromolecules that constitute the extracellular matrix are mainly produced locally by cells in the matrix. As we discuss later, these cells also help to organize the matrix: the orientation of the cytoskeleton inside the cell can control the orientation of the matrix produced outside. In most connective tissues, the matrix macromolecules are secreted largely by cells called **fibroblasts** (Figure 19–54). In certain specialized types of connective tissues, such as cartilage and bone, however, they are secreted by cells of the fibroblast family that have more specific names: *chondroblasts*, for example, form cartilage, and *osteoblasts* form bone.

The matrix in connective tissue is constructed from the same two main classes of macromolecules as in basal laminae: (1) glycosaminoglycan polysaccharide chains, usually covalently linked to protein in the form of proteoglycans, and (2) fibrous proteins such as collagen. We shall see that the members of both classes come in a great variety of shapes and sizes.

The proteoglycan molecules in connective tissue typically form a highly hydrated, gel-like "ground substance" in which the fibrous proteins are embedded. The polysaccharide gel resists compressive forces on the matrix while permitting the rapid diffusion of nutrients, metabolites, and hormones between the blood and the tissue cells. The collagen fibers strengthen and help organize the matrix, while other fibrous proteins, such as the rubberlike *elastin*, give it resilience. Finally, many matrix proteins help cells migrate, settle, and differentiate in the appropriate locations.

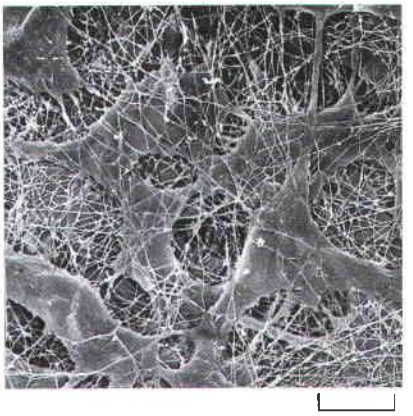


Figure 19–54 Fibroblasts in connective tissue. This scanning electron micrograph shows tissue from the cornea of a rat. The extracellular matrix surrounding the fibroblasts is here composed largely of collagen fibrils. The glycoproteins, hyaluronan, and proteoglycans, which normally form a hydrated gel filling the interstices of the fibrous network, have been removed by enzyme and acid treatment. (From T. Nishida et al., *Invest. Ophthalmol. Vis. Sci.* 29:1887–1890, 1988. With permission from Association for Research in Vision and Opthalmology.)

Glycosaminoglycan (GAG) Chains Occupy Large Amounts of Space and Form Hydrated Gels

Glycosaminoglycans (GAGs) are unbranched polysaccharide chains composed of repeating disaccharide units. They are called GAGs because one of the two sugars in the repeating disaccharide is always an amino sugar (*N*-acetylglucosamine or *N*-acetylgalactosamine), which in most cases is sulfated. The second sugar is usually a uronic acid (glucuronic or iduronic). Because there are sulfate or carboxyl groups on most of their sugars, GAGs are highly negatively charged (Figure 19–55). Indeed, they are the most anionic molecules produced by animal cells. Four main groups of GAGs are distinguished by their sugars, the type of linkage between the sugars, and the number and location of sulfate groups: (1) *hyaluronan*, (2) *chondroitin sulfate* and *dermatan sulfate*, (3) *heparan sulfate*, and (4) *keratan sulfate*.

Polysaccharide chains are too stiff to fold up into the compact globular structures that polypeptide chains typically form. Moreover, they are strongly hydrophilic. Thus, GAGs tend to adopt highly extended conformations that occupy a huge volume relative to their mass (Figure 19–56), and they form gels

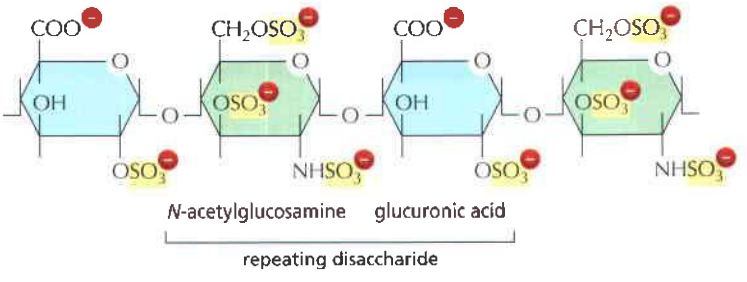


Figure 19–55 The repeating disaccharide sequence of a heparan sulfate glycosaminoglycan (GAG) chain. These chains can consist of as many as 200 disaccharide units, but are typically less than half that size. There is a high density of negative charges along the chain due to the presence of both carboxyl and sulfate groups. The proteoglycans of the basal lamina—perlecan, dystroglycan, and collagen XVIII—all carry heparan sulfate GAGs. The molecule is shown here with its maximal number of sulfate groups. *In vivo*, the proportion of sulfated and nonsulfated groups is variable. Heparin typically has >70% sulfation, while heparan sulfate has <50%.

even at very low concentrations. The weight of GAGs in connective tissue is usually less than 10% of the weight of the fibrous proteins. But, because they form porous hydrated gels, the GAG chains fill most of the extracellular space. Their high density of negative charges attracts a cloud of cations, especially Na^+, that are osmotically active, causing large amounts of water to be sucked into the matrix. This creates a swelling pressure, or turgor, that enables the matrix to withstand compressive forces (in contrast to collagen fibrils, which resist stretching forces). The cartilage matrix that lines the knee joint, for example, can support pressures of hundreds of atmospheres in this way.

Defects in the production of GAGs can affect many different body systems. In one rare human genetic disease, for example, there is a severe deficiency in the synthesis of dermatan sulfate disaccharide. The affected individuals have a short stature, prematurely aged appearance, and generalized defects in their skin, joints, muscles, and bones.

In invertebrates, plants, and fungi, other types of polysaccharides, rather than GAGs, often dominate the extracellular matrix. Thus, in higher plants, as we discuss later, cellulose (polyglucose) chains are packed tightly together in ribbonlike arrays to form the main component of the cell wall. In insects, crustaceans, and other arthropods, chitin (poly-*N*-acetylglucosamine) similarly forms the main component of the exoskeleton. Fungi, too, make their cell walls mainly out of chitin. Together, cellulose and chitin are the most abundant biopolymers on Earth.

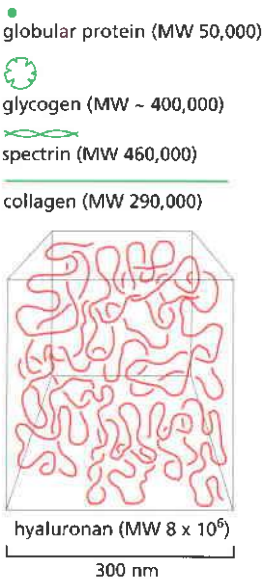


Figure 19–56 The relative dimensions and volumes occupied by various macromolecules. Several proteins, a glycogen granule, and a single hydrated molecule of hyaluronan are shown.

Hyaluronan Acts as a Space Filler and a Facilitator of Cell Migration During Tissue Morphogenesis and Repair

Hyaluronan (also called *hyaluronic acid* or *hyaluronate*) is the simplest of the GAGs (**Figure 19–57**). It consists of a regular repeating sequence of up to 25,000 disaccharide units, is found in variable amounts in all tissues and fluids in adult animals, and is especially abundant in early embryos. Hyaluronan is not typical of the majority of GAGs. In contrast with all of the others, it contains no sulfated sugars, all its disaccharide units are identical, its chain length is enormous (thousands of sugar monomers), and it is not generally linked covalently to any core protein. Moreover, whereas other GAGs are synthesized inside the cell and released by exocytosis, hyaluronan is spun out directly from the cell surface by an enzyme complex embedded in the plasma membrane.

Hyaluronan is thought to have a role in resisting compressive forces in tissues and joints. It is also important as a space filler during embryonic development, where it can be used to force a change in the shape of a structure, as a small quantity expands with water to occupy a large volume (see Figure 19–56). Hyaluronan synthesized locally from the basal side of an epithelium can deform the epithelium by creating a cell-free space beneath it, into which cells subsequently migrate. In the developing heart, for example, hyaluronan synthesis helps in this way to drive formation of the valves and septa that separate the heart's chambers. Similar processes occur in several other organs. When cell

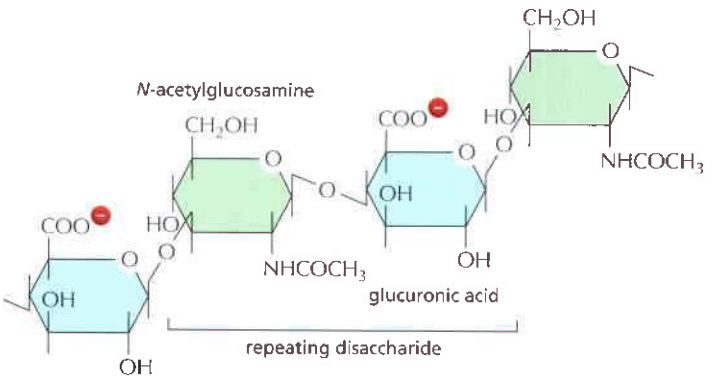


Figure 19–57 The repeating disaccharide sequence in hyaluronan, a relatively simple GAG. This ubiquitous molecule in vertebrates consists of a single long chain of up to 25,000 sugar monomers. Note the absence of sulfate groups.

migration ends, the excess hyaluronan is generally degraded by the enzyme *hyaluronidase*. Hyaluronan is also produced in large quantities during wound healing, and it is an important constituent of joint fluid, in which it serves as a lubricant.

Many of the functions of hyaluronan depend on specific interactions with other molecules, including both proteins and proteoglycans. Some of these molecules that bind to hyaluronan are constituents of the extracellular matrix, while others are integral components of cell surfaces.

Proteoglycans Are Composed of GAG Chains Covalently Linked to a Core Protein

Except for hyaluronan, all GAGs are covalently attached to protein as **proteoglycans**, which are produced by most animal cells. Membrane-bound ribosomes make the polypeptide chain, or *core protein*, of a proteoglycan, which is then threaded into the lumen of the endoplasmic reticulum. The polysaccharide chains are mainly assembled on this core protein in the Golgi apparatus before delivery to the exterior of the cell by exocytosis. First, a special *linkage tetrasaccharide* is attached to a serine side chain on the core protein to serve as a primer for polysaccharide growth; then, one sugar at a time is added by specific glycosyl transferases (**Figure 19–58**). While still in the Golgi apparatus, many of the polymerized sugars are covalently modified by a sequential and coordinated series of reactions. Epimerizations alter the configuration of the substituents around individual carbon atoms in the sugar molecule; sulfations increase the negative charge.

Proteoglycans are usually clearly distinguished from other glycoproteins by the nature, quantity, and arrangement of their sugar side chains. By definition, at least one of the sugar side chains of a proteoglycan must be a GAG. Whereas glycoproteins contain 1–60% carbohydrate by weight, and usually only a few percent, in the form of numerous relatively short, branched oligosaccharide chains, proteoglycans can contain as much as 95% carbohydrate by weight, mostly in the form of long, unbranched GAG chains, each typically about 80 sugars long.

In principle, proteoglycans have the potential for almost limitless heterogeneity. Even a single type of core protein can carry highly variable numbers and types of attached GAG chains. Moreover, the underlying repeating sequence of disaccharides in each GAG can be modified by a complex pattern of sulfate groups. The core proteins, too, are diverse, though many of them share some characteristic domains such as the LINK domain, involved in binding to GAGs.

Proteoglycans can be huge. The proteoglycan *aggrecan*, for example, which is a major component of cartilage, has a mass of about 3×10^6 daltons with over 100 GAG chains. Other proteoglycans are much smaller and have only 1–10 GAG chains; an example is *decorin*, which is secreted by fibroblasts and has a single

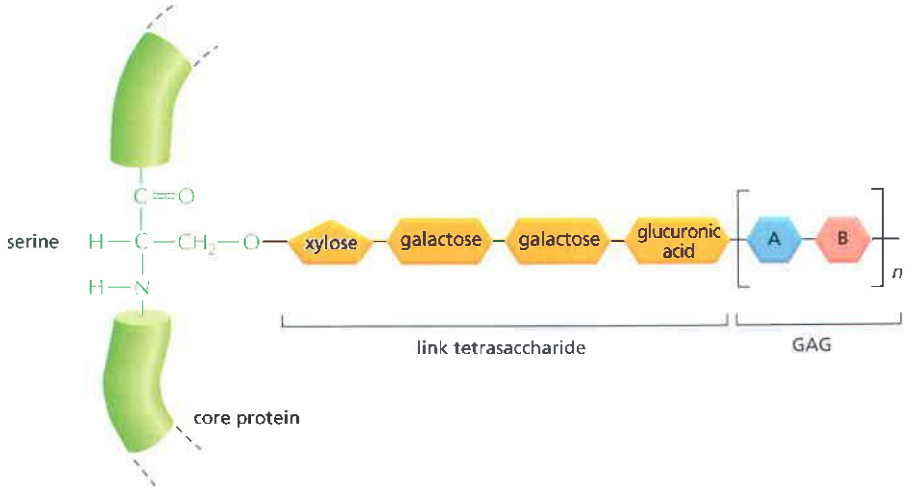
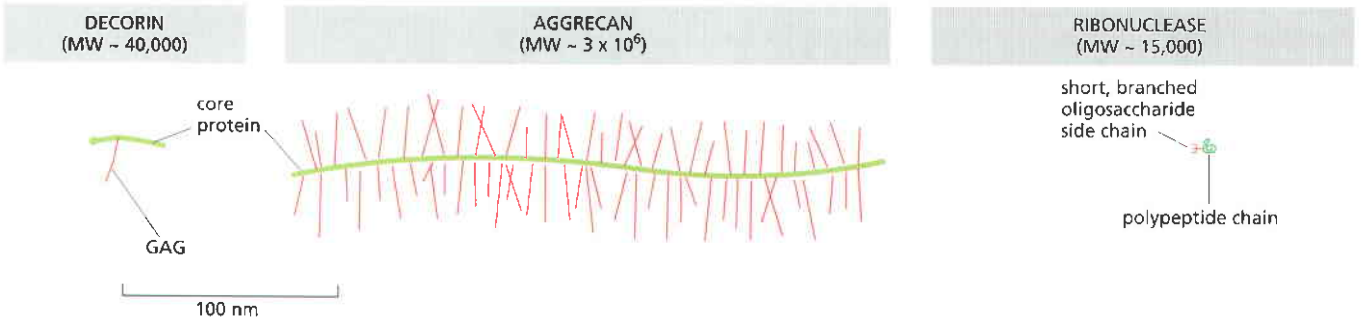


Figure 19–58 The linkage between a GAG chain and its core protein in a proteoglycan molecule. A specific link tetrasaccharide is first assembled on a serine side chain. In most cases, it is unclear how the particular serine is selected, but it seems that a specific local conformation of the polypeptide chain, rather than a specific linear sequence of amino acids, is recognized. The rest of the GAG chain, consisting mainly of a repeating disaccharide unit, is then synthesized, with one sugar being added at a time. In chondroitin sulfate, the disaccharide is composed of D-glucuronic acid and *N*-acetyl-D-galactosamine; in heparan sulfate, it is D-glucosamine (or L-iduronic acid) and *N*-acetyl-D-glucosamine; in keratan sulfate, it is D-galactose and *N*-acetyl-D-glucosamine.



GAG chain (Figure 19–59). Decorin binds to collagen fibrils and regulates fibril assembly and fibril diameter; mice that cannot make decorin have fragile skin that has reduced tensile strength. The GAGs and proteoglycans of these various types can associate to form even larger polymeric complexes in the extracellular matrix. Molecules of aggrecan, for example, assemble with hyaluronan in cartilage matrix to form aggregates that are as big as a bacterium (Figure 19–60). Moreover, besides associating with one another, GAGs and proteoglycans associate with fibrous matrix proteins such as collagen and with protein meshworks such as the basal lamina, creating extremely complex composites (Figure 19–61).

Figure 19–59 Examples of a small (decorin) and a large (aggrecan) proteoglycan found in the extracellular matrix. The figure compares these two proteoglycans with a typical secreted glycoprotein molecule, pancreatic ribonuclease B. All three are drawn to scale. The core proteins of both aggrecan and decorin contain oligosaccharide chains as well as the GAG chains, but these are not shown. Aggrecan typically consists of about 100 chondroitin sulfate chains and about 30 keratan sulfate chains linked to a serine-rich core protein of almost 3000 amino acids. Decorin "decorates" the surface of collagen fibrils, hence its name.

Proteoglycans Can Regulate the Activities of Secreted Proteins

Proeteoglycans are as diverse in function as they are in chemistry and structure. Their GAG chains, for example, can form gels of varying pore size and charge density; one possible function, therefore, is to serve as selective sieves

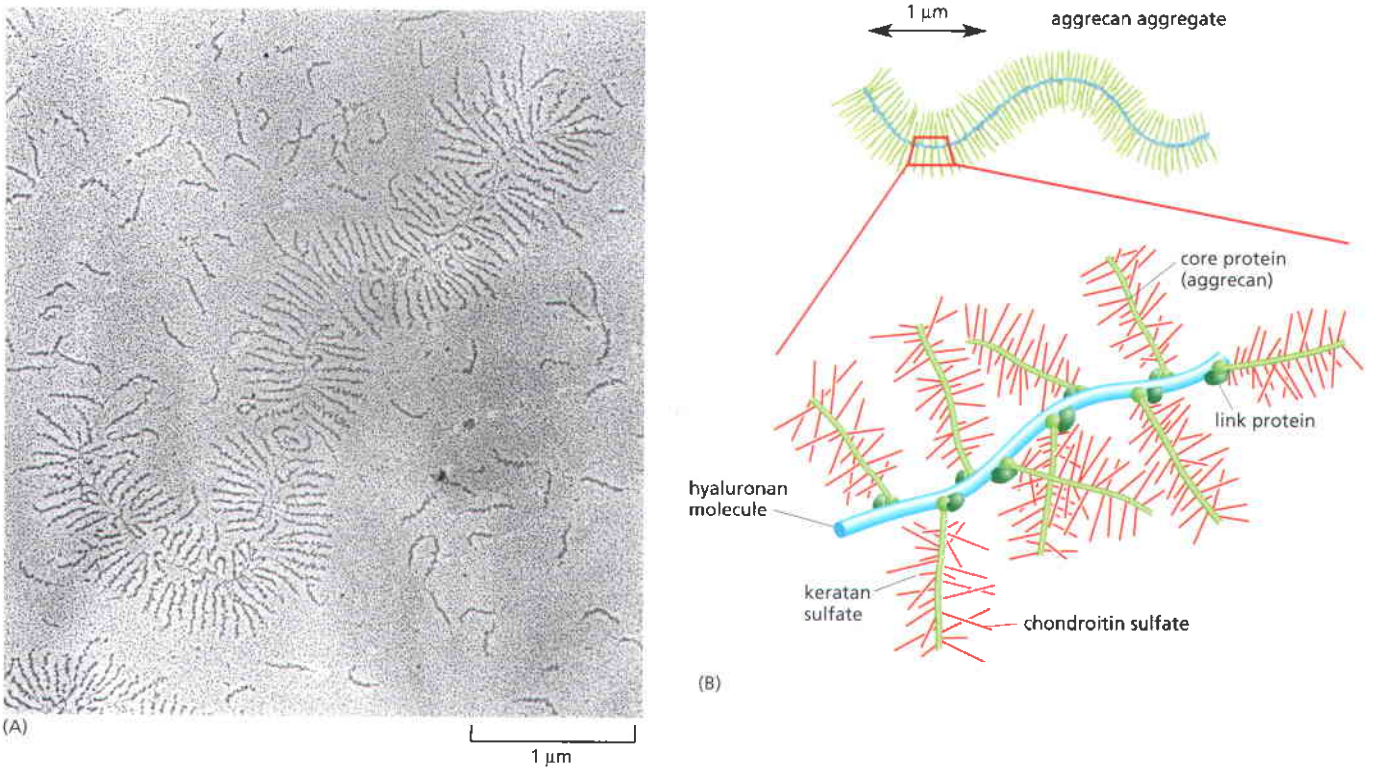


Figure 19–60 An aggrecan aggregate from fetal bovine cartilage. (A) An electron micrograph of an aggrecan aggregate shadowed with platinum. Many free aggrecan molecules are also visible. (B) A drawing of the giant aggrecan aggregate shown in (A). It consists of about 100 aggrecan monomers (each like the one shown in Figure 19–59) noncovalently bound through the N-terminal domain of the core protein to a single hyaluronan chain. A link protein binds both to the core protein of the proteoglycan and to the hyaluronan chain, thereby stabilizing the aggregate. The link proteins are members of a family of hyaluronan-binding proteins, some of which are cell-surface proteins. The molecular weight of such a complex can be 10^8 or more, and it occupies a volume equivalent to that of a bacterium, which is about 2×10^{-12} cm^3. (A, courtesy of Lawrence Rosenberg.)

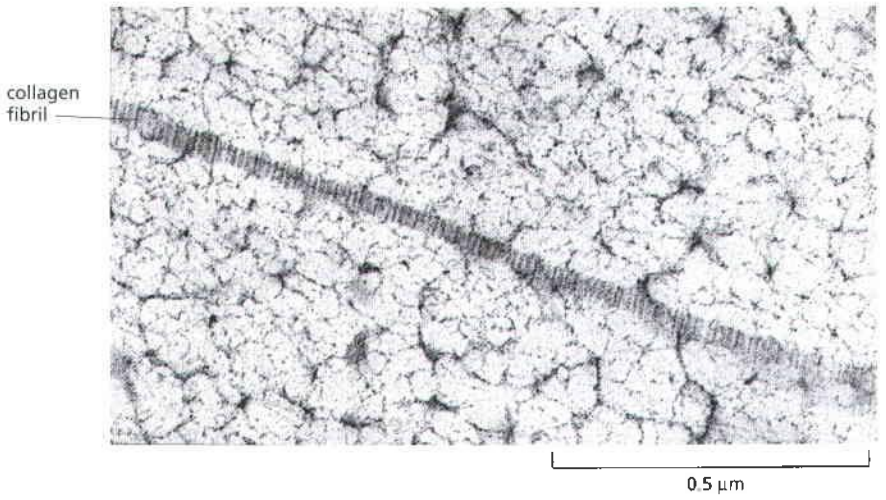


Figure 19–61 Proteoglycans in the extracellular matrix of rat cartilage. The tissue was rapidly frozen at −196°C, and fixed and stained while still frozen (a process called freeze substitution) to prevent the GAG chains from collapsing. In this electron micrograph, the proteoglycan molecules are seen to form a fine filamentous network in which a single striated collagen fibril is embedded. The more darkly stained parts of the proteoglycan molecules are the core proteins; the faintly stained threads are the GAG chains. (Reproduced from E.B. Hunziker and R.K. Schenk, *J. Cell Biol.* 98:277–282, 1984. With permission from The Rockefeller University Press.)

to regulate the traffic of molecules and cells according to their size and charge, as in the thick basal lamina of the kidney glomerulus (see p. 1167).

Proteoglycans have an important role in chemical signaling between cells. They bind various secreted signal molecules, such as certain protein growth factors—controlling their diffusion through the matrix, their range of action, and their lifetime, as well as enhancing or inhibiting their signaling activity. For example, the heparan sulfate chains of proteoglycans bind to *fibroblast growth factors (FGFs)*, which (among other effects) stimulate a variety of cell types to proliferate; this interaction oligomerizes the growth factor molecules, enabling them to cross-link and activate their cell-surface receptors, which are transmembrane tyrosine kinases (see Figure 15–54A). In inflammatory responses, heparan sulfate proteoglycans immobilize secreted chemotactic attractants called *chemokines* (discussed in Chapter 25) on the endothelial surface of a blood vessel at an inflammatory site. This allows the chemokines to remain there for a prolonged period, stimulating white blood cells to leave the bloodstream and migrate into the inflamed tissue.Whereas in most cases the signal molecules bind to the GAG chains of the proteoglycan, this is not always so. Some members of the *transforming growth factor β (TGFβ)* family bind to the core proteins of several matrix proteoglycans, including decorin; binding to decorin inhibits the growth factor activity. Signal molecules (including TGFβ) can also bind to fibrous proteins in the matrix. *Vascular endothelial growth factor (VEGF)*, for example, binds to fibronectin.

Proteoglycans also bind, and regulate the activities of, other types of secreted proteins, including proteolytic enzymes (proteases) and protease inhibitors. Thus they play a part in controlling both the assembly and the degradation of other components of the extracellular matrix, including collagen.

Cell-Surface Proteoglycans Act as Co-Receptors

Not all proteoglycans are secreted components of the extracellular matrix. Some are integral components of plasma membranes and have their core protein either inserted across the lipid bilayer or attached to the lipid bilayer by a glycosylphosphatidylinositol (GPI) anchor. Some of these plasma membrane proteoglycans act as *co-receptors* that collaborate with conventional cell-surface receptor proteins. In addition, some conventional receptors have one or more GAG chains and are therefore proteoglycans themselves.

Among the best-characterized plasma membrane proteoglycans are the *syndecans*, which have a membrane-spanning core protein whose intracellular domain is thought to interact with the actin cytoskeleton and with signaling molecules in the cell cortex. Syndecans are located on the surface of many types of cells, including fibroblasts and epithelial cells. In fibroblasts, syndecans can be found in cell-matrix adhesions, where they modulate integrin function by

Table 19–6 Some Common Proteoglycans

PROTEOGLYCAN	APPROXIMATE MOLECULAR WEIGHT OF CORE PROTEIN	TYPE OF GAG CHAINS	NUMBER OF GAG CHAINS	LOCATION	FUNCTIONS
Aggrecan	210,000	chondroitin sulfate + keratan sulfate (in separate chains)	~130	cartilage	mechanical support; forms large aggregates with hyaluronan
Betaglycan	36,000	chondroitin sulfate/ dermatan sulfate	1	cell surface and matrix	binds TGFβ
Decorin	40,000	chondroitin sulfate/ dermatan sulfate	1	widespread in connective tissues	binds to type I collagen fibrils and TGFβ
Perlecan	600,000	heparan sulfate	2–15	basal laminae	structural and filtering function in basal lamina
Syndecan-1	32,000	chondroitin sulfate + heparan sulfate (in separate chains)	1–3	cell surface	cell adhesion; binds FGF and other growth factors
Dally (in *Drosophila*)	60,000	heparan sulfate	1–3	cell surface	co-receptor for Wingless and Decapentaplegic signaling proteins

interacting with fibronectin on the cell surface and with cytoskeletal and signaling proteins inside the cell. Syndecans also bind FGFs and present them to FGF receptor proteins on the same cell. Similarly, another plasma membrane proteoglycan, called *betaglycan*, binds TGFβ and presents it to TGFβ receptors.

The importance of proteoglycans as co-receptors and as regulators of the distribution and activity of signal molecules is illustrated by the severe developmental defects that can occur when specific proteoglycans are inactivated by mutation. In *Drosophila*, for example, the products of the *Dally* and *Dally-like* genes, coding for members of the *glypican* family, with heparan sulfate side chains, are needed for signaling by no less than four of the main signal proteins that govern the patterning of the embryo (Wingless, Hedgehog, FGF, and Decapentaplegic (Dpp), as discussed in Chapter 22). For some of these signal pathways, mutations in *Dally* or *Dally-like* mimic the effects of mutations in the genes that code for the signal proteins themselves.

The structure, function, and location of some of the proteoglycans discussed in this chapter are summarized in **Table 19–6**.

Collagens Are the Major Proteins of the Extracellular Matrix

The fibrous proteins are no less important than the proteoglycans as components of the extracellular matrix. Foremost among them are the **collagens,** a family of fibrous proteins found in all multicellular animals. Collagens are secreted in large quantities by connective tissue cells, and in smaller quantities by many other cell types. As a major component of skin and bone, they are the most abundant proteins in mammals, where they constitute 25% of the total protein mass.

The primary feature of a typical collagen molecule is its long, stiff, triple-stranded helical structure, in which three collagen polypeptide chains, called α *chains,* are wound around one another in a ropelike superhelix (**Figure 19–62**). Collagens are extremely rich in proline and glycine, both of which are important in the formation of the triple-stranded helix. Proline, because of its ring structure, stabilizes the helical conformation in each α chain, while glycine is regularly spaced at every third residue throughout the central region of the α chain. Being the smallest amino acid (having only a hydrogen atom as a side chain), glycine allows the three helical α chains to pack tightly together to form the final collagen superhelix.

The human genome contains 42 distinct genes coding for different collagen α chains. Different combinations of these genes are expressed in different tissues. Although in principle thousands of types of triple-stranded collagen

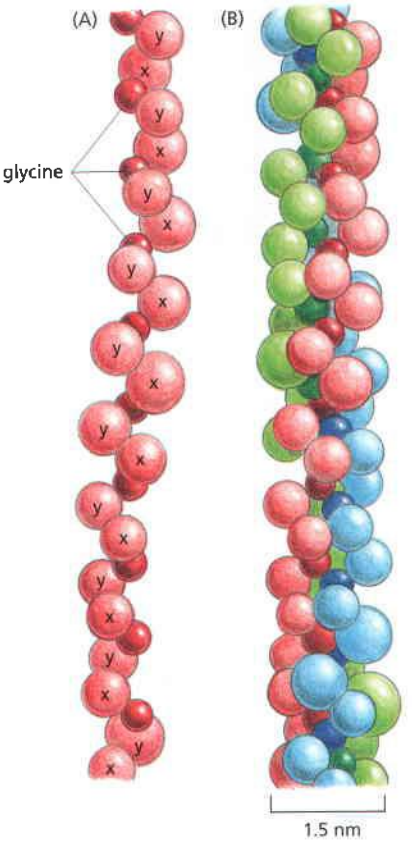


Figure 19–62 The structure of a typical collagen molecule. (A) A model of part of a single collagen α chain in which each amino acid is represented by a sphere. The chain is about 1000 amino acids long. It is arranged as a left-handed helix, with three amino acids per turn and with glycine as every third amino acid. Therefore, an α chain is composed of a series of triplet Gly-X-Y sequences, in which X and Y can be any amino acid (although X is commonly proline and Y is commonly hydroxyproline). (B) A model of part of a collagen molecule in which three α chains, each shown in a different color, are wrapped around one another to form a triple-stranded helical rod. Glycine is the only amino acid small enough to occupy the crowded interior of the triple helix. Only a short length of the molecule is shown; the entire molecule is 300 nm long. (From a model by B.L. Trus.)

molecules could be assembled from various combinations of the 42 α chains, only a limited number of triple-helical combinations are possible, and less than 40 types of collagen molecules have been found. Type I is by far the most common, being the principal collagen of skin and bone. It belongs to the class of **fibrillar collagens**, or fibril-forming collagens, which have long ropelike structures with few or no interruptions. After being secreted into the extracellular space, these collagen molecules assemble into higher-order polymers called *collagen fibrils*, which are thin structures (10–300 nm in diameter) many hundreds of micrometers long in mature tissues and clearly visible in electron micrographs (**Figure 19–63**; see also Figure 19–61). Collagen fibrils often aggregate into larger, cablelike bundles, several micrometers in diameter, that are visible in the light microscope as *collagen fibers*.

Collagen types IX and XII are called *fibril-associated collagens* because they decorate the surface of collagen fibrils. They are thought to link these fibrils to one another and to other components in the extracellular matrix. Type IV, as we have already seen, is a *network-forming collagen*, forming a major part of basal laminae, while type VII molecules form dimers that assemble into specialized structures called *anchoring fibrils*. Anchoring fibrils help attach the basal lamina of multilayered epithelia to the underlying connective tissue and therefore are especially abundant in the skin.

There are also a number of "collagen-like" proteins, including type XVII, which has a transmembrane domain and is found in hemidesmosomes, and type XVIII, which we mentioned earlier as the core protein of a proteoglycan in basal laminae.

Many proteins appear to have evolved by repeated duplications of an original DNA sequence, giving rise to a repetitive pattern of amino acids. The genes that encode the α chains of most of the fibrillar collagens provide a good example: they are very large (up to 44 kilobases in length) and contain about 50 exons. Most of the exons are 54, or multiples of 54, nucleotides long, suggesting that these collagens originated through multiple duplications of a primordial gene

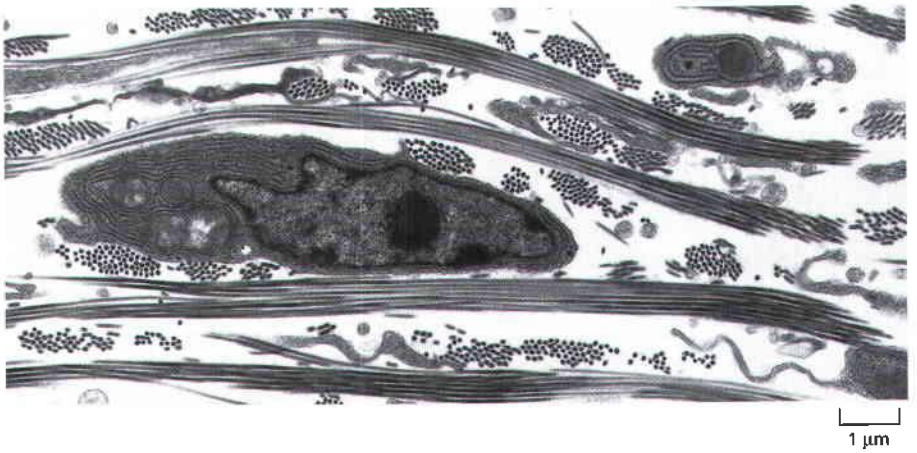


Figure 19–63 A fibroblast surrounded by collagen fibrils in the connective tissue of embryonic chick skin. In this electron micrograph, the fibrils are organized into bundles that run approximately at right angles to one another. Therefore, some bundles are oriented longitudinally, whereas others are seen in cross section. The collagen fibrils are produced by the fibroblasts, which contain abundant endoplasmic reticulum, where secreted proteins such as collagen are synthesized. (From C. Ploetz, E.I. Zycband and D.E. Birk, *J. Struct. Biol.* 106:73–81, 1991. With permission from Elsevier.)

Table 19–7 Some Types of Collagen and Their Properties

	TYPE	POLYMERIZED FORM	TISSUE DISTRIBUTION	MUTANT PHENOTYPE
Fibril-forming (fibrillar)	I	fibril	bone, skin, tendons, ligaments, cornea, internal organs (accounts for 90% of body collagen)	severe bone defects, fractures
	II	fibril	cartilage, invertebral disc, notochord, vitreous humor of the eye	cartilage deficiency, dwarfism
	III	fibril	skin, blood vessels, internal organs	fragile skin, loose joints, blood vessels prone to rupture
	V	fibril (with type I)	as for type I	fragile skin, loose joints, blood vessels prone to rupture
	XI	fibril (with type II)	as for type II	myopia, blindness
Fibril-associated	IX	lateral association with type II fibrils	cartilage	osteoarthritis
Network-forming	IV	sheetlike network	basal lamina	kidney disease (glomerulonephritis), deafness
	VII	anchoring fibrils	beneath stratified squamous epithelia	skin blistering
Transmembrane	XVII	non-fibrillar	hemidesmosomes	skin blistering
Proteoglycan core protein	XVIII	non-fibrillar	basal lamina	myopia, detached retina, hydrocephalus

Note that types I, IV, V, IX, and XI are each composed of two or three types of α chains (distinct, nonoverlapping sets in each case), whereas types II, III, VII, XII, XVII, and XVIII are composed of only one type of α chain each. Only 10 types of collagen are shown, but about 27 types of collagen and 42 types of α chains have been identified in humans.

containing 54 nucleotides and encoding exactly 6 Gly-X-Y repeats (see Figure 19–62).

Table 19–7 provides additional details for some of the collagen types discussed in this chapter.

Collagen Chains Undergo a Series of Post-Translational Modifications

Individual collagen polypeptide chains are synthesized on membrane-bound ribosomes and injected into the lumen of the endoplasmic reticulum (ER) as larger precursors, called *pro-α chains*. These precursors not only have the short amino-terminal signal peptide required to direct the nascent polypeptide to the ER, but also have, at both their N- and C-terminal ends, additional amino acids, called *propeptides*, that are clipped off at a later step of collagen assembly. Moreover, in the lumen of the ER, selected prolines and lysines are hydroxylated to form hydroxyproline and hydroxylysine, respectively, and some of the hydroxy-lysines are glycosylated. Each pro-α chain then combines with two others to form a hydrogen-bonded, triple-stranded, helical molecule known as *procollagen*.

Hydroxylysines and *hydroxyprolines* (**Figure 19–64**) are infrequently found in other animal proteins, although hydroxyproline is abundant in some proteins in the plant cell wall. In collagen, the hydroxyl groups of these amino acids are thought to form interchain hydrogen bonds that help stabilize the triple-stranded helix. Conditions that prevent proline hydroxylation, such as a deficiency of ascorbic acid (vitamin C), have serious consequences. In *scurvy*, the often fatal disease caused by a dietary deficiency of vitamin C that was common in sailors until the nineteenth century, the defective pro-α chains that are synthesized fail to form a stable triple helix and are immediately degraded within the cell, and synthesis of new collagen is inhibited. In healthy tissues, collagen is continually degraded and replaced (with a turnover time of months or years, depending on the tissue). In scurvy, replacement fails, and within a few months, with the gradual loss of the preexisting normal collagen in the matrix, blood vessels become fragile, teeth become loose in their sockets, and wounds cease to heal.

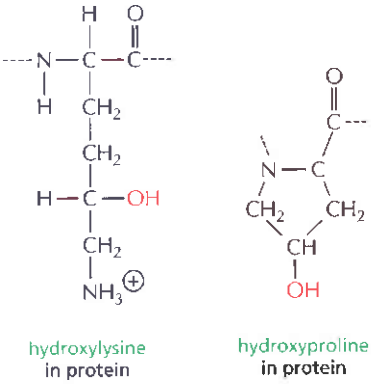


Figure 19–64 Hydroxylysine and hydroxyproline. These modified amino acids are common in collagen. They are formed by enzymes that act after the lysine and proline have been incorporated into procollagen molecules.

Propeptides Are Clipped Off Procollagen After its Secretion, to Allow Assembly of Fibrils

After secretion, the propeptides of the fibrillar procollagen molecules are removed by specific proteolytic enzymes outside the cell. This converts the procollagen molecules to collagen molecules, which assemble in the extracellular space to form much larger **collagen fibrils**. The propeptides have at least two functions. First, they guide the intracellular formation of the triple-stranded collagen molecules. Second, because they are retained until after secretion, they prevent the intracellular formation of large collagen fibrils, which could be catastrophic for the cell.

The process of fibril formation is driven, in part, by the tendency of the collagen molecules, which are more than a thousandfold less soluble than procollagen molecules, to self-assemble. The fibrils begin to form close to the cell surface, often in deep infoldings of the plasma membrane formed by the fusion of secretory vesicles with the cell surface. The underlying cortical cytoskeleton can influence the sites, rates, and orientation of fibril assembly.

When viewed in an electron microscope, collagen fibrils have characteristic cross-striations every 67 nm, reflecting the regularly staggered packing of the individual collagen molecules in the fibril. After the fibrils have formed in the extracellular space, they are greatly strengthened by the formation of covalent cross-links between lysine residues of the constituent collagen molecules (**Figure 19–65**). The types of covalent bonds involved are found only in collagen and elastin. If cross-linking is inhibited, the tensile strength of the fibrils is drastically reduced: collagenous tissues become fragile, and structures such as skin, tendons, and blood vessels tend to tear. The extent and type of cross-linking vary from tissue to tissue. Collagen is especially highly cross-linked in the Achilles tendon, for example, where tensile strength is crucial.

Figure 19–66 summarizes the steps in the synthesis and assembly of collagen fibrils. Given the large number of enzymatic processes involved, it is not surprising that there are many human genetic diseases that affect fibril formation. Mutations affecting type I collagen cause *osteogenesis imperfecta*, characterized by weak bones that fracture easily. Mutations affecting type II collagen cause *chondrodysplasias*, characterized by abnormal cartilage, which leads to bone and joint deformities. And mutations affecting type III collagen cause *Ehlers–Danlos syndrome*, characterized by fragile skin and blood vessels and hypermobile joints.

Secreted Fibril-Associated Collagens Help Organize the Fibrils

In contrast to GAGs, which resist compressive forces, collagen fibrils form structures that resist tensile forces. The fibrils have various diameters and are organized in different ways in different tissues. In mammalian skin, for example, they are woven in a wickerwork pattern so that they resist tensile stress in multiple directions; leather consists of this material, suitably preserved. In tendons, collagen fibrils are organized in parallel bundles aligned along the major axis of

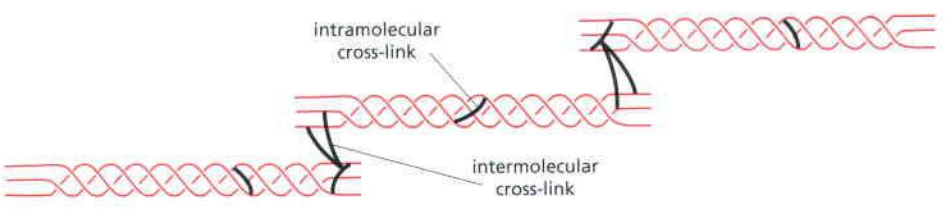


Figure 19–65 Cross-links formed between modified lysine side chains within a collagen fibril. Covalent intramolecular and intermolecular cross-links are formed in several steps. First, the extracellular enzyme lysyl oxidase deaminates certain lysines and hydroxylysines, to yield highly reactive aldehyde groups. The aldehydes then react spontaneously to form covalent bonds with each other or with other lysines or hydroxylysines. Most of the cross-links form between the short nonhelical segments at each end of the collagen molecules.

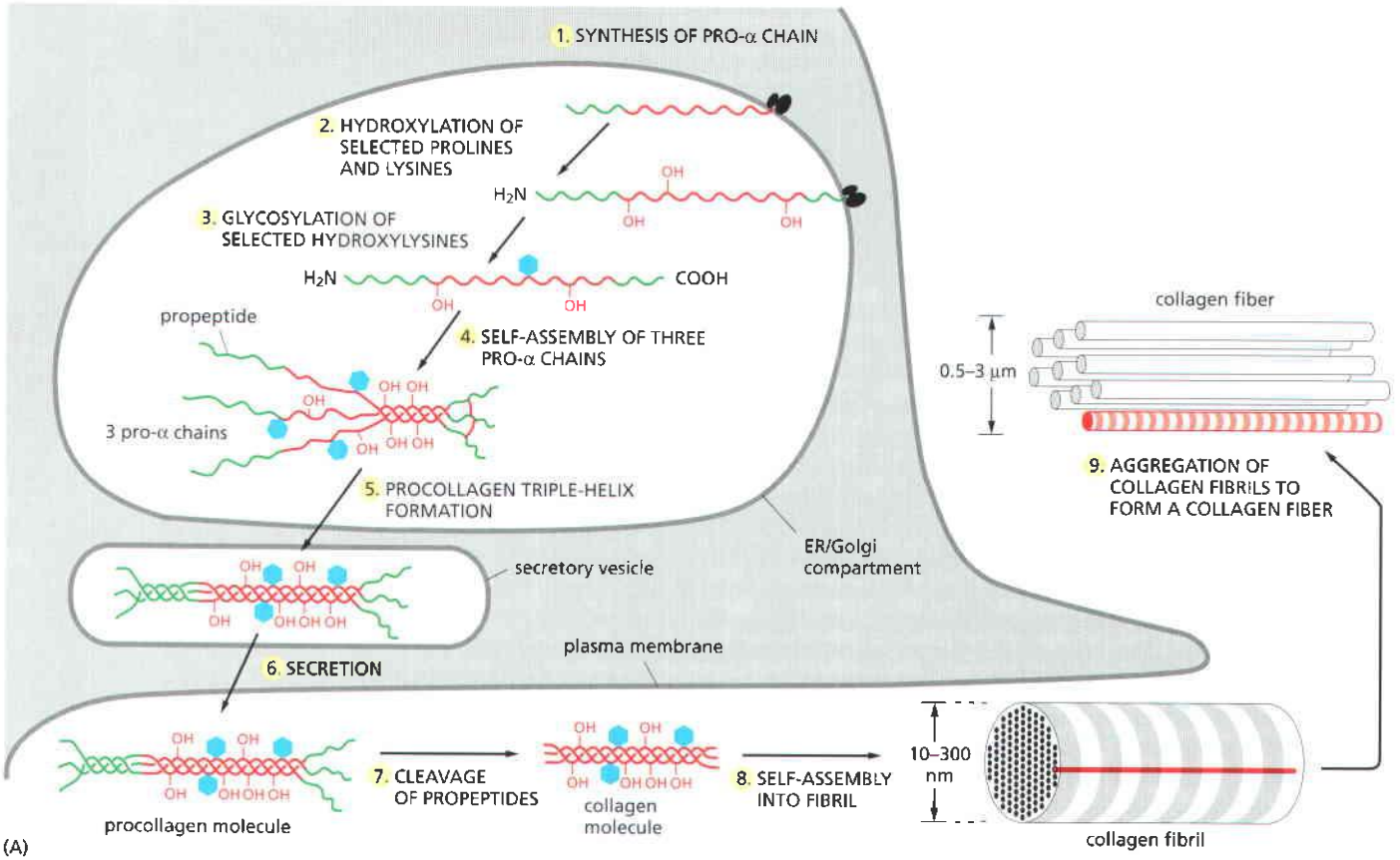


Figure 19–66 The intracellular and extracellular events in the formation of a collagen fibril. (A) Note that procollagen assembles into collagen fibrils in the extracellular space, often within large infoldings in the plasma membrane (not shown). As one example of how collagen fibrils can form ordered arrays in the extracellular space, they are shown further assembling into large collagen fibers, which are visible in the light microscope. The covalent cross-links that stabilize the extracellular assemblies are omitted in this picture. (B) Electron micrograph of a negatively stained collagen fibril reveals its typical striated appearance. (B, courtesy of Robert Horne.)

tension. In mature bone and in the cornea, they are arranged in orderly plywoodlike layers, with the fibrils in each layer lying parallel to one another but nearly at right angles to the fibrils in the layers on either side. The same arrangement occurs in tadpole skin (**Figure 19–67**).

The connective tissue cells themselves determine the size and arrangement of the collagen fibrils. The cells can express one or more genes for the different types of fibrillar procollagen molecules. But even fibrils composed of the same mixture of fibrillar collagen molecules have different arrangements in different tissues. How is this achieved? Part of the answer is that cells can regulate the disposition of the collagen molecules after secretion by guiding collagen fibril formation in close association with the plasma membrane (see Figure 19–66). In addition, cells can influence this organization by secreting, along with their fibrillar collagens, different kinds and amounts of other matrix macromolecules. In particular, they secrete the fibrous protein fibronectin, as we shall discuss later, and this precedes the formation of collagen fibrils and helps guide their organization.

Fibril-associated collagens, such as types IX and XII collagens, are thought to be especially important in this regard. They differ from fibrillar collagens in several ways.

1. Their triple-stranded helical structure is interrupted by one or two short nonhelical domains, which makes the molecules more flexible than fibrillar collagen molecules.

2. They are not cleaved after secretion and therefore retain their propeptides.
3. They do not aggregate with one another to form fibrils in the extracellular space. Instead, they bind in a periodic manner to the surface of fibrils formed by the fibrillar collagens. Type IX molecules bind to type-II-collagen-containing fibrils in cartilage, the cornea, and the vitreous of the eye (**Figure 19–68**), whereas type XII molecules bind to type-I-collagen-containing fibrils in tendons and various other tissues.

Fibril-associated collagens are thought to mediate the interactions of collagen fibrils with one another and with other matrix macromolecules to help determine the organization of the fibrils in the matrix.

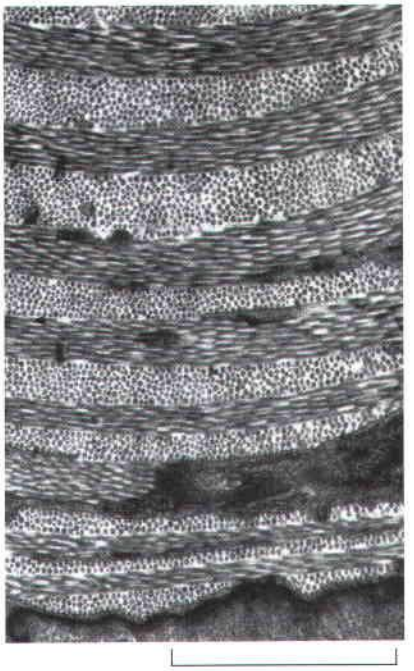


Figure 19–67 Collagen fibrils in the tadpole skin. This electron micrograph shows the plywoodlike arrangement of the fibrils: successive layers of fibrils are laid down nearly at right angles to each other. This organization is also found in mature bone and in the cornea. (Courtesy of Jerome Gross.)

Cells Help Organize the Collagen Fibrils They Secrete by Exerting Tension on the Matrix

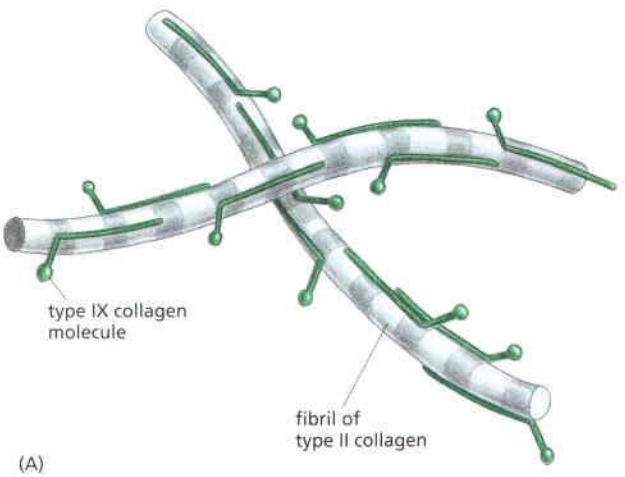
Cells interact with the extracellular matrix mechanically as well as chemically, and studies in culture suggest that this mechanical interaction can have dramatic effects on the architecture of connective tissue. Thus, when fibroblasts are mixed with a meshwork of randomly oriented collagen fibrils that form a gel in a culture dish, the fibroblasts tug on the meshwork, drawing in collagen from their surroundings and thereby causing the gel to contract to a small fraction of its initial volume. By similar activities, a cluster of fibroblasts surrounds itself with a capsule of densely packed and circumferentially oriented collagen fibers.

If two small pieces of embryonic tissue containing fibroblasts are placed far apart on a collagen gel, the intervening collagen becomes organized into a compact band of aligned fibers that connect the two explants (**Figure 19–69**). The fibroblasts subsequently migrate out from the explants along the aligned collagen fibers. Thus, the fibroblasts influence the alignment of the collagen fibers, and the collagen fibers in turn affect the distribution of the fibroblasts.

Fibroblasts may have a similar role in organizing the extracellular matrix inside the body, first of all synthesizing the collagen fibrils and depositing them in the correct orientation, then working on the matrix they have secreted, crawling over it and tugging on it so as to create tendons and ligaments and the tough, dense layers of connective tissue that ensheathe and bind together most organs.

Elastin Gives Tissues Their Elasticity

Many vertebrate tissues, such as skin, blood vessels, and lungs, need to be both strong and elastic in order to function. A network of **elastic fibers** in the extracellular matrix of these tissues gives them the required resilience so that they



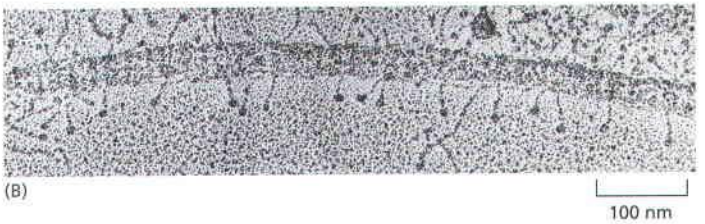




Figure 19–68 Type IX collagen. (A) Type IX collagen molecules binding in a periodic pattern to the surface of a fibril containing type II collagen. (B) Electron micrograph of a rotary-shadowed type-II-collagen-containing fibril in cartilage, sheathed in type IX collagen molecules. (C) An individual type IX collagen molecule. (B and C, from L. Vaughan et al., *J. Cell Biol.* 106:991–997, 1988. With permission from The Rockefeller University Press.)

can recoil after transient stretch (Figure 19–70). Elastic fibers are at least five times more extensible than a rubber band of the same cross-sectional area. Long, inelastic collagen fibrils are interwoven with the elastic fibers to limit the extent of stretching and prevent the tissue from tearing.

The main component of elastic fibers is **elastin**, a highly hydrophobic protein (about 750 amino acids long), which, like collagen, is unusually rich in proline and glycine but, unlike collagen, is not glycosylated and contains some hydroxyproline but no hydroxylysine. Soluble *tropoelastin* (the biosynthetic precursor of elastin) is secreted into the extracellular space and assembled into elastic fibers close to the plasma membrane, generally in cell-surface infoldings. After secretion, the tropoelastin molecules become highly cross-linked to one another, generating an extensive network of elastin fibers and sheets. A mechanism similar to the one that operates in cross-linking collagen molecules forms cross-links between the lysines.

The elastin protein is composed largely of two types of short segments that alternate along the polypeptide chain: hydrophobic segments, which are responsible for the elastic properties of the molecule; and alanine- and lysine-rich α-helical segments, which form cross-links between adjacent molecules. Each segment is encoded by a separate exon. There is still uncertainty concerning the conformation of elastin molecules in elastic fibers and how the structure of these fibers accounts for their rubberlike properties. However, it seems that parts of the elastin polypeptide chain, like the polymer chains in ordinary rubber, adopt a loose "random coil" conformation, and it is the random coil nature of the component molecules cross-linked into the elastic fiber network that allows the network to stretch and recoil like a rubber band (Figure 19–71).

Elastin is the dominant extracellular matrix protein in arteries, comprising 50% of the dry weight of the largest artery—the aorta. Mutations in the elastin gene causing a deficiency of the protein in mice or humans result in narrowing of the aorta or other arteries and excessive proliferation of smooth muscle cells in the arterial wall. Apparently, the normal elasticity of an artery is required to restrain the proliferation of these cells.

Elastic fibers do not consist solely of elastin. The elastin core is covered with a sheath of *microfibrils*, each of which has a diameter of about 10 nm. The microfibrils appear before elastin in developing tissues and seem to provide scaffolding to guide elastin deposition. Arrays of microfibrils are elastic in their own right, and in some places they persist in the absence of elastin: they help to hold the lens in its place in the eye, for example. Microfibrils are composed of a number of distinct glycoproteins, including the large glycoprotein *fibrillin*, which binds to elastin and is essential for the integrity of elastic fibers. Mutations in the fibrillin gene result in *Marfan's syndrome*, a relatively common human genetic disorder. In the most severely affected individuals the aorta is

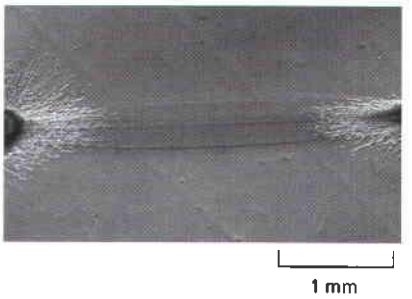


Figure 19–69 The shaping of the extracellular matrix by cells. This micrograph shows a region between two pieces of embryonic chick heart (rich in fibroblasts as well as heart muscle cells) that were cultured on a collagen gel for 4 days. A dense tract of aligned collagen fibers has formed between the explants, presumably as a result of the fibroblasts in the explants tugging on the collagen. (From D. Stopak and A.K. Harris, *Dev. Biol.* 90:383–398, 1982. With permission from Academic Press.)



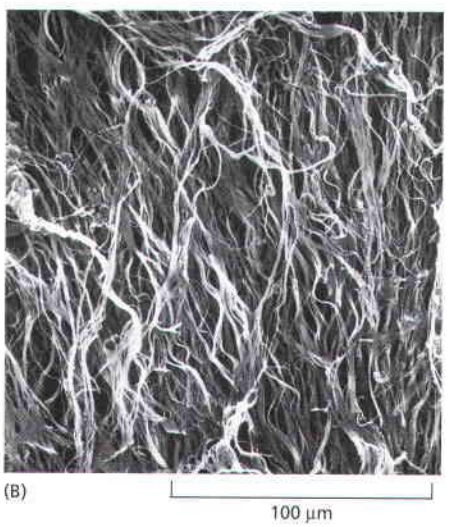


Figure 19–70 Elastic fibers. These scanning electron micrographs show (A) a low-power view of a segment of a dog's aorta and (B) a high-power view of the dense network of longitudinally oriented elastic fibers in the outer layer of the same blood vessel. All the other components have been digested away with enzymes and formic acid. (From K.S. Haas et al., *Anat. Rec.* 230:86–96, 1991. With permission from Wiley-Liss.)

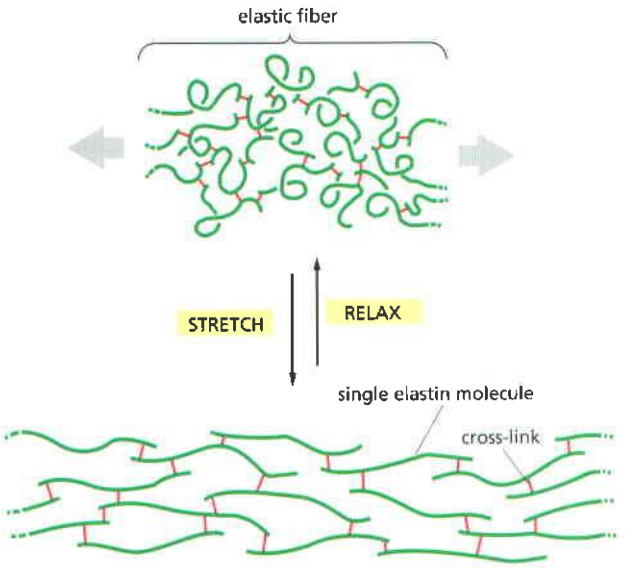


Figure 19–71 Stretching a network of elastin molecules. The molecules are joined together by covalent bonds *(red)* to generate a cross-linked network. In this model, each elastin molecule in the network can extend and contract in a manner resembling a random coil, so that the entire assembly can stretch and recoil like a rubber band.

prone to rupture; other common effects include displacement of the lens and abnormalities of the skeleton and joints. Affected individuals are often unusually tall and lanky: Abraham Lincoln is suspected to have had the condition.

Fibronectin Is an Extracellular Protein That Helps Cells Attach to the Matrix

The extracellular matrix contains a number of noncollagen proteins that typically have multiple domains, each with specific binding sites for other matrix macromolecules and for receptors on the surface of cells. These proteins therefore contribute to both organizing the matrix and helping cells attach to it. Like the proteoglycans, they also guide cell movements in developing tissues, by serving as tracks along which cells can migrate or as repellents that keep cells out of forbidden areas.

The first of this class of matrix proteins to be well characterized was **fibronectin**, a large glycoprotein found in all vertebrates and important for many cell–matrix interactions. Thus, for example, mutant mice that are unable to make fibronectin die early in embryogenesis because their endothelial cells fail to form proper blood vessels. The defect is thought to result from abnormalities in the interactions of these cells with the surrounding extracellular matrix, which normally contains fibronectin.

Fibronectin is a dimer composed of two very large subunits joined by disulfide bonds at one end. Each subunit is folded into a series of functionally distinct domains separated by regions of flexible polypeptide chain (**Figure 19–72**). The domains in turn consist of smaller modules, each of which is serially repeated and usually encoded by a separate exon, suggesting that the fibronectin gene, like the collagen genes, evolved by multiple exon duplications. In the human genome, there is only one fibronectin gene, containing about 50 exons of similar size, but the transcripts can be spliced in different ways to produce many different fibronectin isoforms. Sites in the main type of module, called the **type III fibronectin repeat**, bind to integrins and thereby to cell surfaces. This module is about 90 amino acids long and occurs at least 15 times in each subunit. The type III fibronectin repeat is among the most common of all protein domains in vertebrates.

Tension Exerted by Cells Regulates the Assembly of Fibronectin Fibrils

Fibronectin can exist both in a soluble form, circulating in the blood and other body fluids, and as insoluble *fibronectin fibrils*, in which fibronectin dimers are

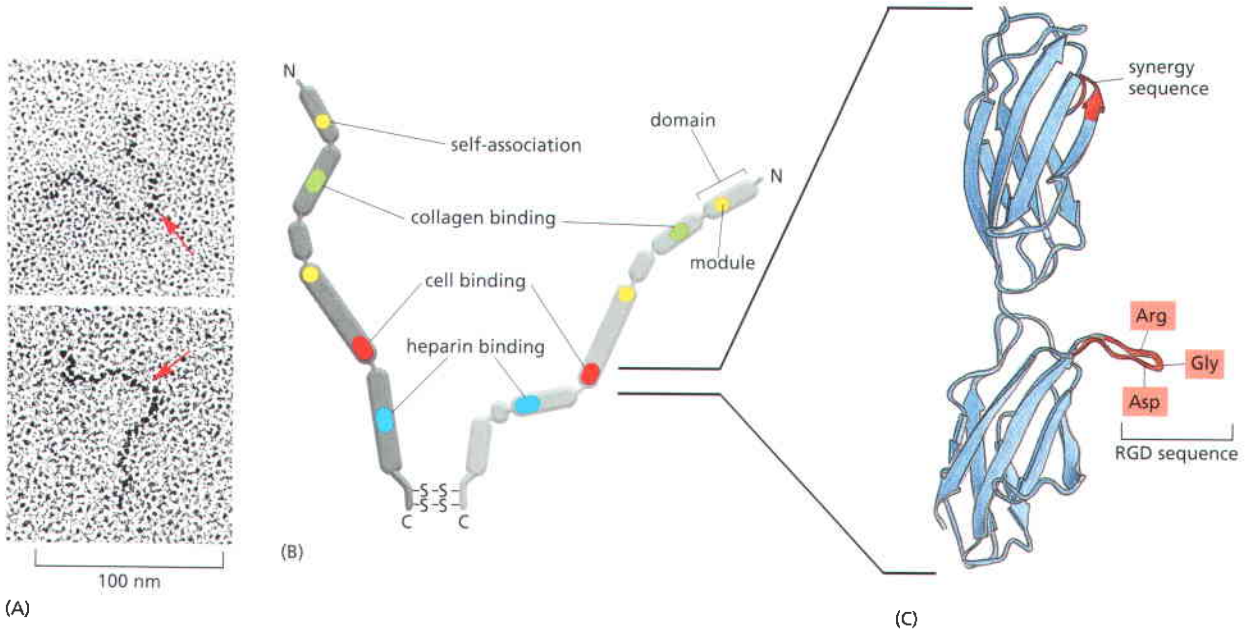


Figure 19–72 The structure of a fibronectin dimer. (A) Electron micrographs of individual fibronectin dimer molecules shadowed with platinum; *red arrows* mark the C-termini. (B) The two polypeptide chains are similar but generally not identical (being made from the same gene but from differently spliced mRNAs). They are joined by two disulfide bonds near the C-termini. Each chain is almost 2500 amino acids long and is folded into five or six domains connected by flexible polypeptide segments. Individual domains contain modules specialized for binding to a particular molecule or to a cell, as indicated for five of the domains. For simplicity, not all of the known binding sites are shown (there are other cell-binding sites, for example). (C) The three-dimensional structure of two type III fibronectin repeat modules as determined by x-ray crystallography. The type III repeat is the main repeating module in fibronectin. Both the Arg-Gly-Asp (RGD) and the "synergy" sequences shown in *red* form part of the major cell-binding site, and are important for binding. (A, from J. Engel et al., *J. Mol. Biol.* 150:97–120, 1981. With permission from Academic Press; C, from Daniel J. Leahy, *Annu. Rev. Cell Dev. Biol.* 13:363–393, 1997. With permission from Annual Reviews.)

cross-linked to one another by additional disulfide bonds and form part of the extracellular matrix. Unlike fibrillar collagen molecules, however, which can self-assemble into fibrils in a test tube, fibronectin molecules assemble into fibrils only on the surface of cells, and only where those cells possess appropriate fibronectin-binding proteins—in particular, integrins. The integrins provide a linkage from the fibronectin outside the cell to the actin cytoskeleton inside it. The linkage transmits tension to the fibronectin molecules—provided that they also have an attachment to some other structure—and stretches them, exposing a cryptic (hidden) binding site in the fibronectin molecules (**Figure 19–73**). This allows them to bind directly to one another and to recruit additional fibronectin molecules to form a fibril (**Figure 19–74**). This dependence on tension and interaction with cell surfaces ensures that fibronectin fibrils assemble where there is a mechanical need for them and not in inappropriate locations such as the bloodstream.

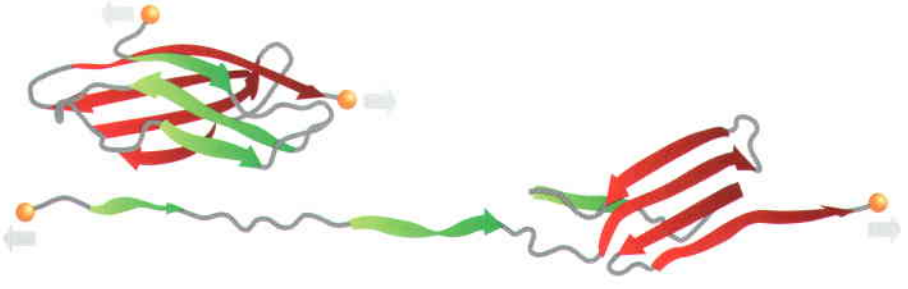


Figure 19–73 Unfolding of a type III fibronectin domain in response to tension. The stretching of fibronectin in this way exposes cryptic binding sites that cause the stretched molecules to assemble into filaments, as shown in Figure 19–74. (From V. Vogel and M. Sheetz, *Nat. Rev. Mol. Cell Biol.* 7:265–275, 2006. With permission from Macmillan Publishers Ltd.)

Many other extracellular matrix proteins have multiple repeats similar to the type III fibronectin repeat, and it is possible that tension exerted on these proteins also uncovers cryptic binding sites and thereby influences their polymerization.

Fibronectin Binds to Integrins Through an RGD Motif

One way to analyze a complex multifunctional protein molecule such as fibronectin is to chop it into pieces and determine the function of its individual domains. When fibronectin is treated with a low concentration of a proteolytic enzyme, the polypeptide chain is cut in the connecting regions between the domains, leaving the domains themselves intact. One can then show that one of its domains binds to collagen, another to heparin, another to specific receptors on the surface of various types of cells, and so on (see Figure 19–72B). Synthetic peptides corresponding to different segments of the cell-binding domain have been used to identify a specific tripeptide sequence (*Arg-Gly-Asp*, or *RGD*), which is found in one of the type III repeats (see Figure 19–72C), as a central feature of the cell-binding site. Even very short peptides containing this **RGD sequence** can compete with fibronectin for the binding site on cells, thereby inhibiting the attachment of the cells to a fibronectin matrix. If these peptides are coupled to a solid surface, they cause cells to adhere to it.

Several extracellular proteins besides fibronectin also have an RGD sequence that mediates cell-surface binding. Some of these are involved in blood clotting, and peptides containing the RGD sequence have been useful in the development of anti-clotting drugs. Some snakes use a similar strategy to cause their victims to bleed: they secrete RGD-containing anti-clotting proteins called *disintegrins* into their venom.

The cell-surface receptors that bind RGD-containing proteins are members of the integrin family. Each integrin, however, specifically recognizes its own small set of matrix molecules, indicating that tight binding requires more than just the RGD sequence. Moreover, RGD sequences are not the only sequence motifs used for binding to integrins: many integrins recognize and bind to other motifs instead.

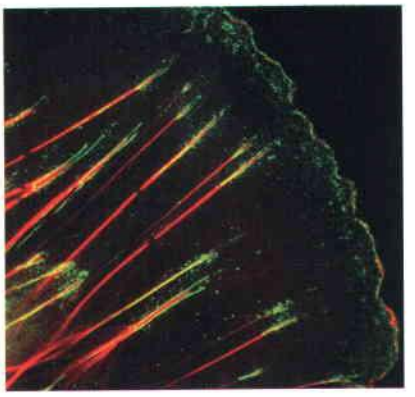


Figure 19–74 Organization of fibronectin into fibrils at the cell surface. The picture shows the front end of a migrating mouse fibroblast. The extracellular adhesion protein fibronectin is stained *green*, and the intracellular filaments of actin, *red*. The fibronectin is initially present as small dot-like aggregates near the leading edge of the cell. It accumulates at focal adhesions (sites of anchorage of actin filaments) and becomes organized into fibrils parallel to the actin filaments. Integrin molecules spanning the cell membrane link the fibronectin outside the cell to the actin filaments inside it. Tension exerted on the fibronectin molecules through this linkage is thought to stretch them, exposing binding sites that promote fibril formation. (Courtesy of Roumen Pankov and Kenneth Yamada.)

Cells Have to Be Able to Degrade Matrix, as Well as Make it

The ability of cells to degrade and destroy extracellular matrix is as important as their ability to make it and bind to it. Rapid matrix degradation is required in processes such as tissue repair, and even in the seemingly static extracellular matrix of adult animals there is a slow, continuous turnover, with matrix macromolecules being degraded and resynthesized. This allows bone, for example, to be remodelled so as to adapt to the stresses on it, as discussed in Chapter 23.

From the point of view of individual cells, the ability to cut through matrix is crucial in two ways: it enables them to divide while embedded in matrix, and it enables them to travel through it. As we have already mentioned, cells in connective tissues generally need to be able to stretch out in order to divide. If a cell lacks the enzymes needed to cut through the surrounding matrix or is embedded in a matrix that resists their action, it remains rounded, unable to extend processes because the matrix is impenetrable; as a result, the cell is strongly inhibited from dividing, as well as being hindered from migrating.

Localized degradation of matrix components is also required wherever cells have to escape from confinement by a basal lamina. It is needed during normal branching growth of epithelial structures such as glands, for example, to allow the population of epithelial cells to expand, and needed also when white blood cells migrate across the basal lamina of a blood vessel into tissues in response to infection or injury. Less benignly, matrix degradation is important both for the spread of cancer cells through the body and for the ability of the cancer cells to proliferate in the tissues that they invade (discussed in Chapter 20).

Matrix Degradation Is Localized to the Vicinity of Cells

In general, matrix components are degraded by extracellular proteolytic enzymes (proteases) that act close to the cells that produce them. Thus, antibodies that recognize the products of proteolytic cleavage stain matrix only around cells. Many of these proteases belong to one of two general classes. Most are **matrix metalloproteases**, which depend on bound Ca^{2+} or Zn^{2+} for activity; the others are **serine proteases**, which have a highly reactive serine in their active site. Together, metalloproteases and serine proteases cooperate to degrade matrix proteins such as collagen, laminin, and fibronectin. Some metalloproteases, such as the *collagenases*, are highly specific, cleaving particular proteins at a small number of sites. In this way, the structural integrity of the matrix is largely retained, while the limited amount of proteolysis that occurs is sufficient for cell migration. Other metalloproteases may be less specific, but, because they are anchored to the plasma membrane, they can act just where they are needed; it is this type of matrix metalloprotease that is crucial for a cell's ability to divide when embedded in matrix.

Clearly, the activity of the proteases that degrade the matrix components has to be tightly controlled, if the fabric of the body is not to collapse in a heap. Three basic control mechanisms operate:

Local activation: Many proteases are secreted as inactive precursors that can be activated locally when needed. An example is *plasminogen*, an inactive protease precursor that is abundant in the blood. Proteases called *plasminogen activators* cleave plasminogen locally to yield the active serine protease *plasmin*, which helps break up blood clots. *Tissue-type plasminogen activator (tPA)* is often given to patients who have just had a heart attack or thrombotic stroke; it helps dissolve the arterial clot that caused the attack, thereby restoring blood-flow to the tissue.

Confinement by cell-surface receptors: Many cells have receptors on their surface that bind proteases, thereby confining the enzyme to the sites where it is needed. *Urokinase-type plasminogen activator (uPA)* is an example. It is found bound to receptors on the growing tips of axons and at the leading edge of some migrating cells, where it may serve to clear a pathway for their migration. Receptor-bound uPA may also help some cancer cells metastasize (**Figure 19–75**).

Secretion of inhibitors: The action of proteases is confined to specific areas by various secreted protease inhibitors, including the *tissue inhibitors of metalloproteases (TIMPs)* and the serine protease inhibitors known as *serpins*. The inhibitors are protease-specific and bind tightly to the activated enzyme, blocking its activity. An attractive idea is that the inhibitors are secreted by cells at the margins of areas of active protein degradation to protect uninvolved matrix; they may also protect cell-surface proteins required for cell adhesion and migration. The overexpression of TIMPs inhibits the migration of some cell types, emphasizing the importance of metalloproteases for the migration.

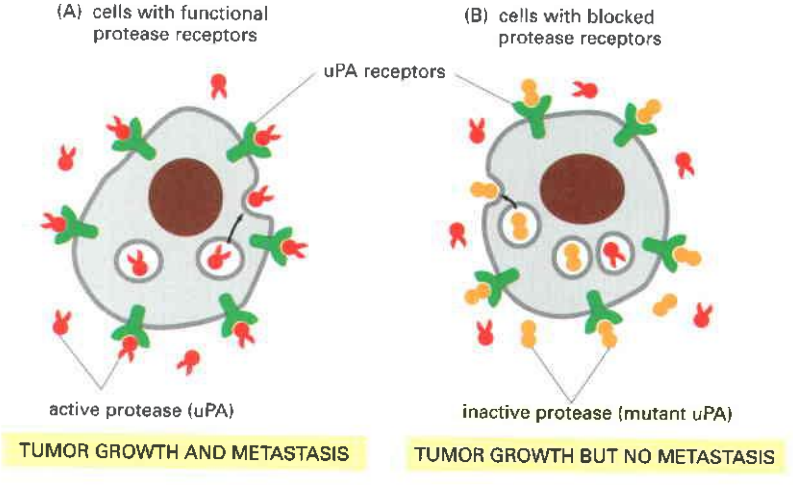


Figure 19–75 The importance of proteases bound to cell-surface receptors. (A) Human prostate cancer cells make and secrete the serine protease uPA, which binds to cell-surface uPA receptor proteins. (B) The same cells have been transfected with DNA that encodes an excess of an inactive form of uPA, which binds to the uPA receptors but has no protease activity. By occupying most of the uPA receptors, the inactive uPA prevents the active protease from binding to the cell surface. Both types of cells secrete active uPA, grow rapidly, and produce tumors when injected into experimental animals. But the cells in (A) metastasize widely, whereas the cells in (B) do not. To metastasize via the blood, tumor cells have to crawl through basal laminae and other extracellular matrices on the way into and out of the bloodstream. This experiment suggests that proteases must be bound to the cell surface to facilitate migration through the matrix.

Summary

Cells in connective tissues are embedded in an intricate extracellular matrix that not only binds the cells together but also influences their survival, development, shape, polarity, and migratory behavior. The matrix contains various protein fibers interwoven in a hydrated gel composed of a network of glycosaminoglycan (GAG) chains.

GAGs are a heterogeneous group of negatively charged polysaccharide chains that (except for hyaluronan) are covalently linked to protein to form proteoglycan molecules. The negative charges attract counterions, which have a powerful osmotic effect, drawing water into the matrix and keeping it swollen so as to occupy a large volume of extracellular space. Proteoglycans are also found on the surface of cells, where they often function as co-receptors to help cells respond to secreted signal proteins.

Fiber-forming proteins give the matrix strength and resilience. They also form structures to which cells can be anchored, often via large multidomain glycoproteins such as laminin and fibronectin that have multiple binding sites for integrins on the cell surface. Elasticity is provided by elastin molecules, which form an extensive cross-linked network of fibers and sheets that can stretch and recoil. The fibrillar collagens (types I, II, III, V, and XI) provide tensile strength. They are ropelike, triple-stranded helical molecules that aggregate into long fibrils in the extracellular space. The fibrils in turn can assemble into various highly ordered arrays. Fibril-associated collagen molecules, such as types IX and XII, decorate the surface of collagen fibrils and influence the interactions of the fibrils with one another and with other matrix components.

Matrix components are degraded by extracellular proteolytic enzymes. Most of these are matrix metalloproteases, which depend on bound Ca^{2+} or Zn^{2+} for activity, while others are serine proteases, which have a reactive serine in their active site. The degradation of matrix components is subject to complex controls, and cells can, for example, cause a localized degradation of matrix components to clear a path through the matrix.

THE PLANT CELL WALL

Each cell in a plant deposits, and is in turn completely enclosed by, an elaborate extracellular matrix called the *plant cell wall*. It was the thick cell walls of cork, visible in a primitive microscope, that in 1663 enabled Robert Hooke to distinguish and name cells for the first time. The walls of neighboring plant cells, cemented together to form the intact plant (**Figure 19–76**), are generally thicker, stronger, and, most important of all, more rigid than the extracellular matrix produced by animal cells. In evolving relatively rigid walls, which can be up to many micrometers thick, early plant cells forfeited the ability to crawl about and adopted a sedentary lifestyle that has persisted in all present-day plants.

The Composition of the Cell Wall Depends on the Cell Type

All cell walls in plants have their origin in dividing cells, as the cell plate forms during cytokinesis to create a new partition wall between the daughter cells (discussed in Chapter 17). The new cells are usually produced in special regions called *meristems* (discussed in Chapter 22), and they are generally small in comparison with their final size. To accommodate subsequent cell growth, the walls of the newborn cells, called **primary cell walls**, are thin and extensible, although tough. Once growth stops, the wall no longer needs to be extensible: sometimes the primary wall is retained without major modification, but, more commonly, a rigid, **secondary cell wall** is produced by depositing new layers of matrix inside the old ones. These new layers generally have a composition that is significantly different from that of the primary wall. The most common additional polymer in secondary walls is **lignin**, a complex network of covalently linked phenolic compounds found in the walls of the xylem vessels and fiber cells of woody tissues.

Although the cell walls of higher plants vary in both composition and organization, they are all constructed, like animal extracellular matrices, using a

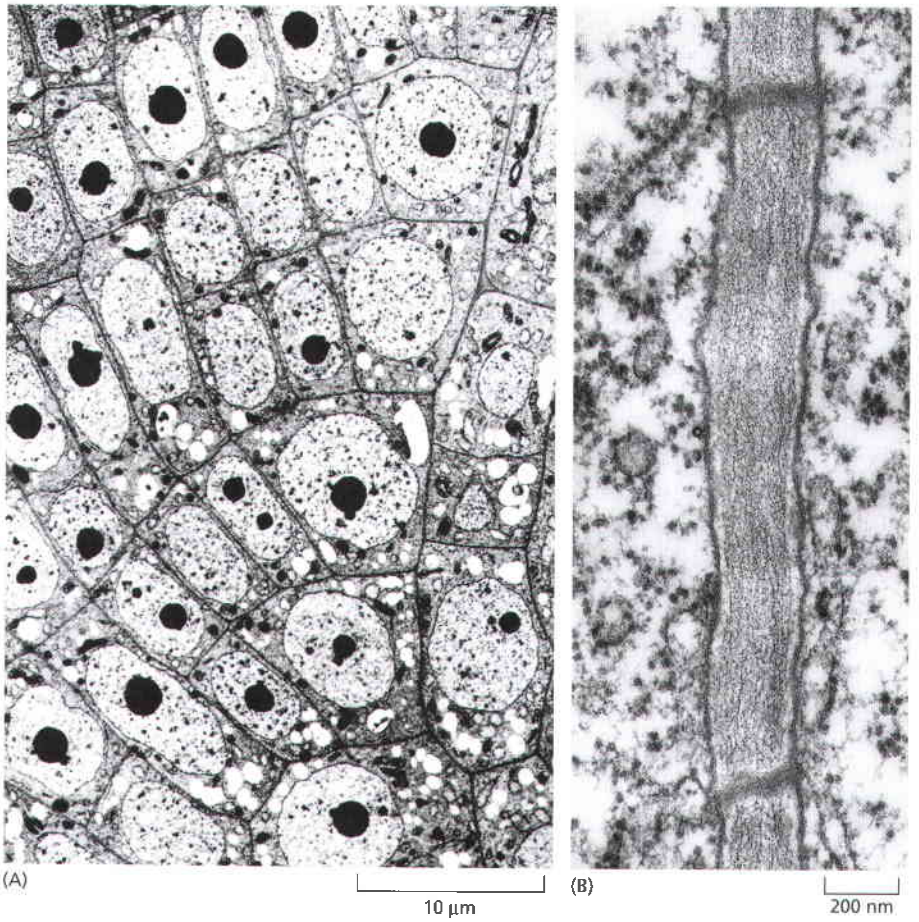


Figure 19–76 Plant cell walls. (A) Electron micrograph of the root tip of a rush, showing the organized pattern of cells that results from an ordered sequence of cell divisions in cells with relatively rigid cell walls. In this growing tissue, the cell walls are still relatively thin, appearing as fine black lines between the cells in the micrograph. (B) Section of a typical cell wall separating two adjacent plant cells. The two dark transverse bands correspond to plasmodesmata that span the wall (see Figure 19–38). (A, courtesy of C. Busby and B. Gunning, *Eur. J. Cell Biol.* 21:214–233, 1980. With permission from Elsevier; B, courtesy of Jeremy Burgess.)

structural principle common to all fiber-composites, including fiberglass and reinforced concrete. One component provides tensile strength, while another, in which the first is embedded, provides resistance to compression. While the principle is the same in plants and animals, the chemistry is different. Unlike the animal extracellular matrix, which is rich in protein and other nitrogen-containing polymers, the plant cell wall is made almost entirely of polymers that contain no nitrogen, including *cellulose* and lignin. For a sedentary organism that depends on CO_2, H_2O, and sunlight, these two abundant biopolymers represent "cheap," carbon-based, structural materials, helping to conserve the scarce fixed nitrogen available in the soil that generally limits plant growth. Thus trees, for example, make a huge investment in the cellulose and lignin that comprise the bulk of their biomass.

In the cell walls of higher plants, the tensile fibers are made from the polysaccharide cellulose, the most abundant organic macromolecule on Earth, tightly linked into a network by *cross-linking glycans*. In primary cell walls, the matrix in which the cross-linked cellulose network is embedded is composed of *pectin*, a highly hydrated network of polysaccharides rich in galacturonic acid. Secondary cell walls contain additional molecules to make them rigid and permanent; lignin, in particular, forms a hard, waterproof filler in the interstices between the other components. All of these molecules are held together by a combination of covalent and noncovalent bonds to form a highly complex structure, whose composition, thickness and architecture depend on the cell type.

The plant cell wall thus has a "skeletal" role in supporting the structure of the plant as a whole, a protective role as an enclosure for each cell individually, and a transport role, helping to form channels for the movement of fluid in the plant. When plant cells become specialized, they generally adopt a specific shape and produce specially adapted types of walls, according to which the different types of cells in a plant can be recognized and classified (**Figure 19–77**;

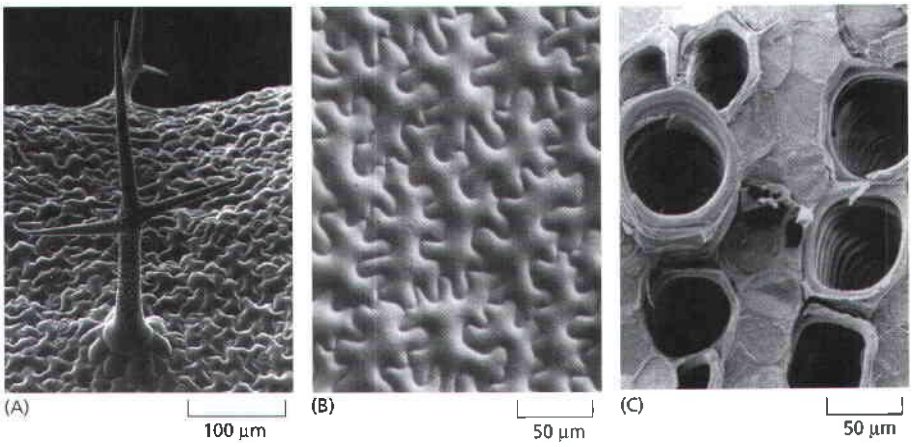


Figure 19–77 Some specialized plant cell types with appropriately modified cell walls. (A) A trichome, or hair, on the upper surface of an *Arabidopsis* leaf. This spiky, protective single cell is shaped by the local deposition of a tough, cellulose-rich wall. (B) Surface view of tomato leaf epidermal cells. The cells fit together snugly like the pieces of a jigsaw puzzle, providing a strong outer covering for the leaf. The outer cell wall is reinforced with a cuticle and with waxes that waterproof the leaf and help defend it against pathogens. (C) This view into young xylem elements shows the thick, lignified, hoop-reinforced secondary cell wall that creates robust tubes for the transport of water throughout the plant. (A, courtesy of Paul Linstead; B and C, courtesy of Kim Findlay.)

see also Panel 22–2, pp. 1404–1405).We focus here, however, on the primary cell wall and the molecular architecture that underlies its remarkable combination of strength, resilience, and plasticity, as seen in the growing parts of a plant.

The Tensile Strength of the Cell Wall Allows Plant Cells to Develop Turgor Pressure

The aqueous extracellular environment of a plant cell consists of the fluid contained in the walls that surround the cell. Although the fluid in the plant cell wall contains more solutes than does the water in the plant's external milieu (for example, soil), it is still hypotonic in comparison with the cell interior. This osmotic imbalance causes the cell to develop a large internal hydrostatic pressure, or **turgor pressure**, which pushes outward on the cell wall, just as an inner tube pushes outward on a tire. The turgor pressure increases just to the point where the cell is in osmotic equilibrium, with no net influx of water despite the salt imbalance (see Panel 11–1, p. 664). The turgor pressure generated in this way may reach 10 or more atmospheres, about five times that in the average car tire. This pressure is vital to plants because it is the main driving force for cell expansion during growth, and it provides much of the mechanical rigidity of living plant tissues. Compare the wilted leaf of a dehydrated plant, for example, with the turgid leaf of a well-watered one. It is the mechanical strength of the cell wall that allows plant cells to sustain this internal pressure.

The Primary Cell Wall Is Built from Cellulose Microfibrils Interwoven with a Network of Pectic Polysaccharides

Cellulose gives the primary cell wall tensile strength. Each cellulose molecule consists of a linear chain of at least 500 glucose residues that are covalently linked to one another to form a ribbonlike structure, which is stabilized by hydrogen bonds within the chain (**Figure 19–78**). In addition, hydrogen bonds between adjacent cellulose molecules cause them to stick together in overlapping parallel arrays, forming bundles of about 40 cellulose chains, all of which have the same polarity. These highly ordered crystalline aggregates, many micrometers long, are called **cellulose microfibrils**, and they have a tensile strength comparable to steel (see Figure 19–78). Sets of microfibrils are arranged in layers, or lamellae, with each microfibril about 20–40 nm from its neighbors and connected to them by long cross-linking glycan molecules that are attached by hydrogen bonds to the surface of the microfibrils. The primary cell wall consists of several such lamellae arranged in a plywoodlike network (**Figure 19–79**).

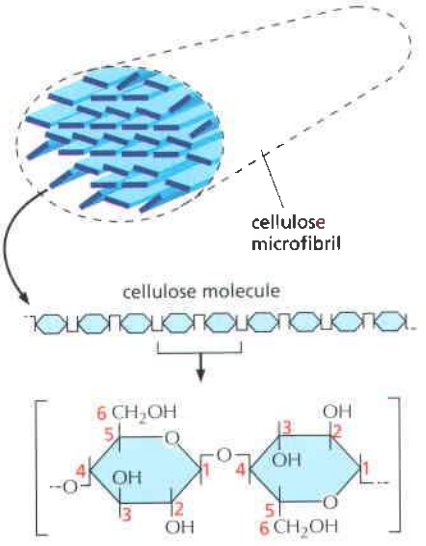


Figure 19–78 Cellulose. Cellulose molecules are long, unbranched chains of β1,4-linked glucose units. Each glucose residue is inverted with respect to its neighbors, and the resulting disaccharide repeat occurs hundreds of times in a single cellulose molecule. About 40 individual cellulose molecules assemble to form a strong, hydrogen-bonded cellulose microfibril.

The **cross-linking glycans** are a heterogeneous group of branched polysaccharides that bind tightly to the surface of each cellulose microfibril and thereby help to cross-link the microfibrils into a complex network. Their function is analogous to that of the fibril-associated collagens discussed earlier (see Figure 19–68). There are many classes of cross-linking glycans, but they all have a long linear backbone composed of one type of sugar (glucose, xylose, or mannose) from which short side chains of other sugars protrude. It is the backbone sugar molecules that form hydrogen bonds with the surface of cellulose microfibrils, cross-linking them in the process. Both the backbone and the side-chain sugars vary according to the plant species and its stage of development.

Coextensive with this network of cellulose microfibrils and cross-linking glycans is another cross-linked polysaccharide network based on **pectins** (see Figure19–79). Pectins are a heterogeneous group of branched polysaccharides that contain many negatively charged galacturonic acid units. Because of their negative charge, pectins are highly hydrated and associated with a cloud of cations, resembling the glycosaminoglycans of animal cells in the large amount of space they occupy (see Figure 19–56). When Ca^{2+} is added to a solution of pectin molecules, it cross-links them to produce a semirigid gel (it is pectin that is added to fruit juice to make jam set). Certain pectins are particularly abundant in the *middle lamella*, the specialized region that cements together the walls of adjacent cells (see Figure 19–79); here, Ca^{2+} cross-links are thought to help hold cell-wall components together. Although covalent bonds also play a part in linking the components, very little is known about their nature. Regulated separation of cells at the middle lamella underlies such processes as the ripening of tomatoes and the abscission (detachment) of leaves in the fall.

In addition to the two polysaccharide-based networks that form the bulk of all plant primary cell walls, proteins are present, contributing up to about 5% of the wall's dry mass. Many of these proteins are enzymes, responsible for wall turnover and remodeling, particularly during growth. Another class of wall proteins, like collagen, contain high levels of hydroxyproline. These proteins are thought to strengthen the wall, and they are produced in greatly increased amounts as a local response to attack by pathogens. From the genome sequence of *Arabidopsis*, it has been estimated that more than 700 genes are required to synthesize, assemble, and remodel the plant cell wall. Some of the main polymers found in the primary and secondary cell wall are listed in **Table 19–8**.

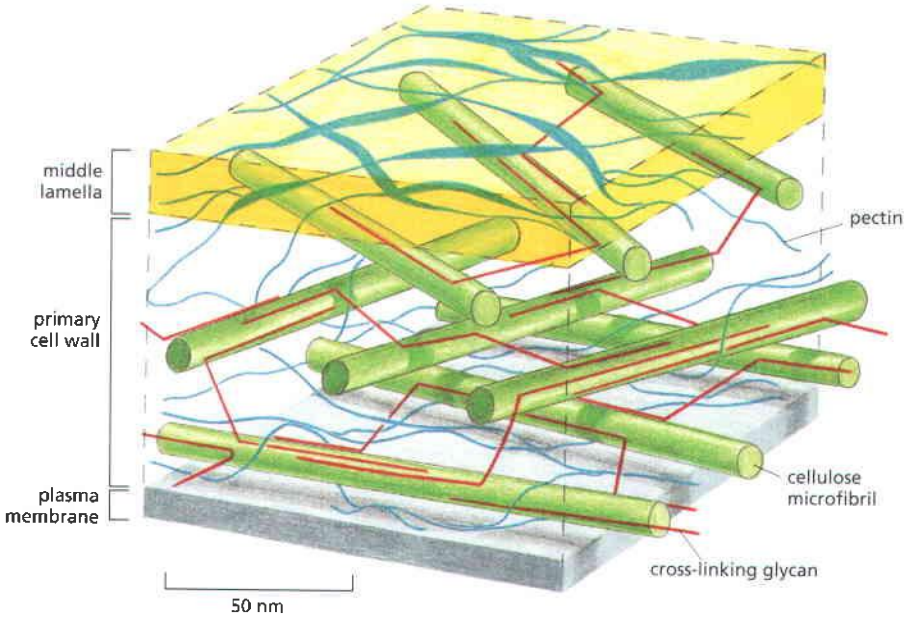


Figure 19–79 Scale model of a portion of a primary plant cell wall showing the two major polysaccharide networks. The orthogonally arranged layers of cellulose microfibrils *(green)* are tied into a network by the cross-linking glycans *(red)* that form hydrogen bonds with the microfibrils. This network is coextensive with a network or matrix of pectic polysaccharides *(blue)*. The network of cellulose and cross-linking glycans provides tensile strength, while the pectin network resists compression. Cellulose, cross-linking glycans, and pectin are typically present in roughly equal amounts in a primary cell wall. The middle lamella is especially rich in pectin, and it cements adjacent cells together.

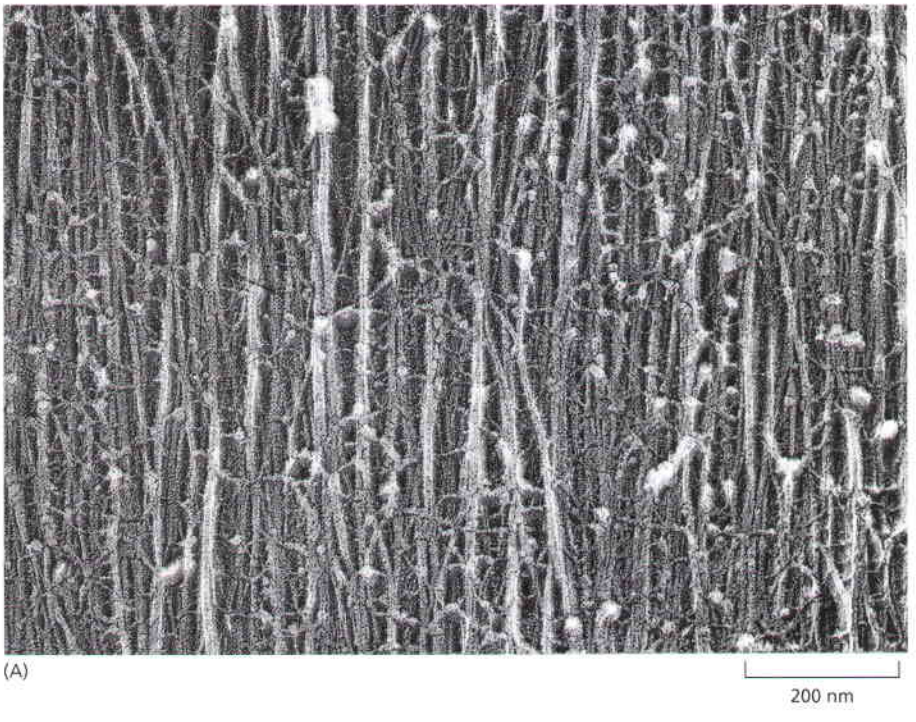
Table 19–8 The Polymers of the Plant Cell Wall

POLYMER	COMPOSITION	FUNCTIONS
Cellulose	linear polymer of glucose	fibrils confer tensile strength on all walls
Cross-linking glycans	xyloglucan, glucuronoarabinoxylan, and mannan	cross-link cellulose fibrils into robust network
Pectin	homogalacturonans and rhamnogalacturonans	forms negatively charged, hydrophilic network that gives compressive strength to primary walls; cell–cell adhesion
Lignin	cross-linked coumaryl, coniferyl, and sinapyl alcohols	forms strong waterproof polymer that reinforces secondary cell walls
Proteins and glycoproteins	enzymes, hydroxyproline-rich proteins	responsible for wall turnover and remodeling; help defend against pathogens

Oriented Cell-Wall Deposition Controls Plant Cell Growth

Once a plant cell has left the meristem where it is generated, it can grow dramatically, commonly by more than a thousand times in volume. The manner of this expansion determines the final shape of each cell, and hence the final form of the plant as a whole. Turgor pressure inside the cell drives the expansion, but it is the behavior of the cell wall that governs its direction and extent. Complex wall-remodeling activities are required, as well as the deposition of new wall materials. Because of their crystalline structure, the individual cellulose microfibrils in the wall are unable to stretch, and this gives them a crucial role in the process. For the cell wall to stretch or deform, the microfibrils must either slide past one another or become more widely separated, or both. The orientation of the microfibrils in the innermost layers of the wall governs the direction in which the cell expands. Cells in plants therefore anticipate their future morphology by controlling the orientation of the cellulose microfibrils that they deposit in the wall (Figure 19–80).

Unlike most other matrix macromolecules, which are made in the endoplasmic reticulum and Golgi apparatus and are secreted, cellulose, like hyaluronan in animals, is spun out from the surface of the cell by a plasma-membrane-bound enzyme complex (cellulose synthase), which uses as its substrate the sugar nucleotide UDP-glucose supplied from the cytosol. Each enzyme complex, or *rosette*, has a six-fold symmetry and contains the protein products of three separate cellulose synthase *(CESA)* genes. Each CESA protein



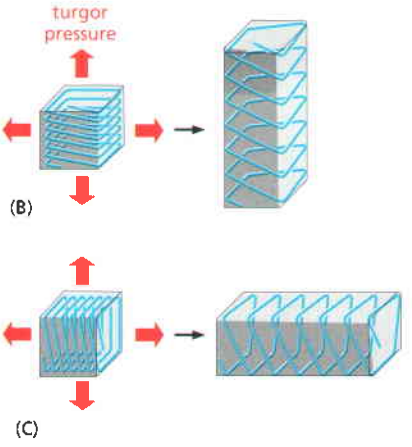


Figure 19–80 Cellulose microfibrils influence the direction of cell elongation. (A) The orientation of cellulose microfibrils in the primary cell wall of an elongating carrot cell is shown in this electron micrograph of a shadowed replica from a rapidly frozen and deep-etched cell wall. The cellulose microfibrils are aligned parallel to one another and perpendicular to the axis of cell elongation. The microfibrils are cross-linked by, and interwoven with, a complex web of matrix molecules (compare with Figure 19–79). (B,C) The cells in (A) and (B) start off with identical shapes (shown here as cubes) but with different net orientations of cellulose microfibrils in their walls. Although turgor pressure is uniform in all directions, cell wall loosening allows each cell to elongate only in a direction perpendicular to the orientation of the innermost layer of microfibrils, which have great tensile strength. Cell expansion occurs in concert with the insertion of new wall material. The final shape of an organ, such as a shoot, is determined in part by the direction in which its component cells can expand. (A, courtesy of Brian Wells and Keith Roberts.)

is essential for the production of a cellulose microfibril. Three CESA genes are required for primary cell wall synthesis and a different three for secondary cell wall synthesis.

As they are being synthesized, the nascent cellulose chains assemble into microfibrils. These are spun out on the extracellular surface of the plasma membrane, forming a layer, or lamella, in which all the microfibrils have more or less the same alignment (see Figure 19–79). Each new lamella is deposited internally to the previous one, so that the wall consists of concentrically arranged lamellae, with the oldest on the outside. The most recently deposited microfibrils in elongating cells commonly lie perpendicular to the axis of cell elongation, although the orientation of the microfibrils in the outer lamellae that were laid down earlier may be different (see Figure 19–80B and C).

Microtubules Orient Cell-Wall Deposition

An important clue to the mechanism that dictates this orientation came from observations of the microtubules in plant cells. These are frequently arranged in the cortical cytoplasm with the same orientation as the cellulose microfibrils that are currently being deposited in the cell wall in that region. These cortical microtubules form a *cortical array* close to the cytosolic face of the plasma membrane, held there by poorly characterized proteins (**Figure 19–81**). The

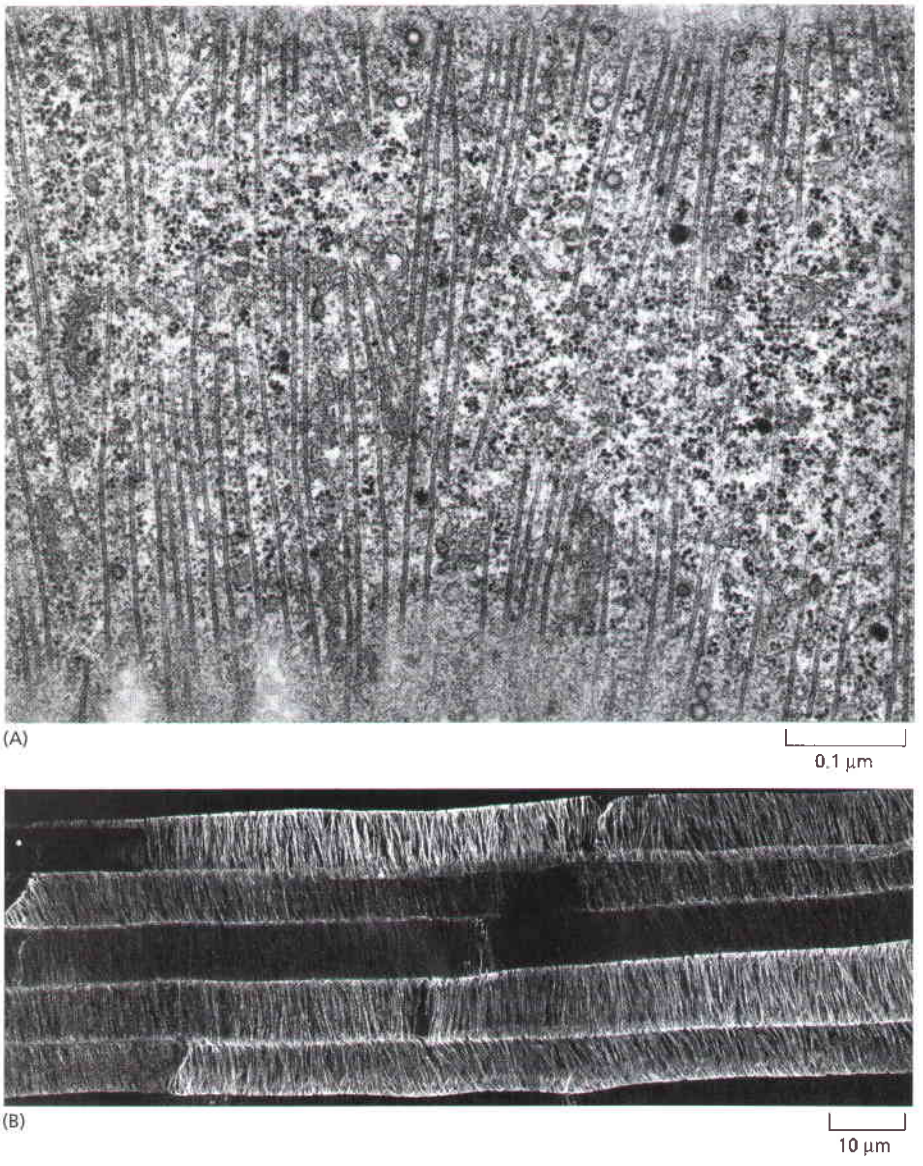


Figure 19–81 The cortical array of microtubules in a plant cell. (A) A grazing section of a root-tip cell from timothy grass, showing a cortical array of microtubules lying just below the plasma membrane. These microtubules are oriented perpendicularly to the long axis of the cell. (B) Epidermal cells in a root tip of *Arabidopsis* that has been stained, by immunofluorescence, to reveal the entire cortical microtubule arrays in these rapidly elongating cells. The clear transverse arrays are at right angles to the axis of root elongation. (A, courtesy of Brian Gunning; B, courtesy of Keiko Sugimoto-Shirasu.)

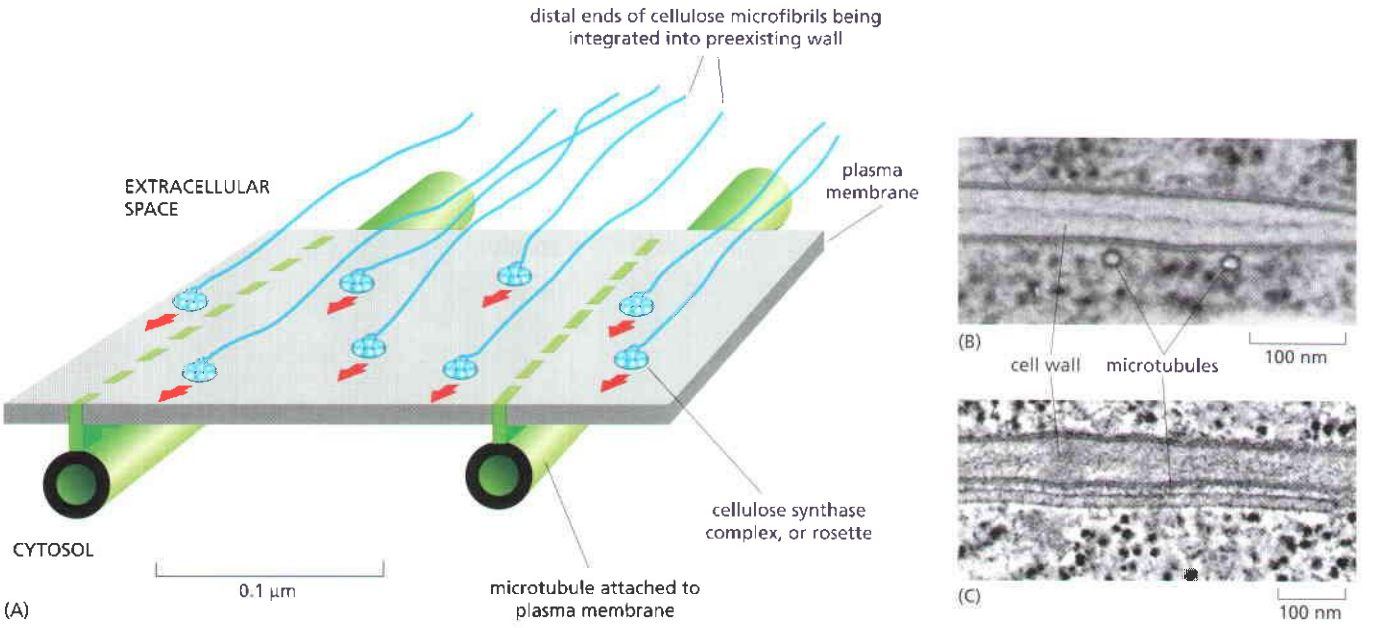


Figure 19–82 One model of how the orientation of newly deposited cellulose microfibrils might be determined by the orientation of cortical microtubules. (A) The large cellulose synthase complexes are integral membrane proteins that continuously synthesize cellulose microfibrils on the outer face of the plasma membrane. The distal ends of the stiff microfibrils become integrated into the texture of the wall, and their elongation at the proximal end pushes the synthase complex along in the plane of the membrane. Because the cortical array of microtubules is attached to the plasma membrane in a way that confines this complex to defined membrane channels, the orientation of these microtubules—when they are present—determines the axis along which the new microfibrils are laid down. (B,C) Two electron micrographs show the tight association of the cortical microtubules with the plasma membrane. One shows the microtubules in cross section while the other shows a longitudinal section. Both emphasize the constant gap of about 20 nm between membrane and microtubule; the connecting molecules responsible remain obscure. (B and C, courtesy of Andrew Staehelin.)

congruent orientation of the cortical array of microtubules (lying just inside the plasma membrane) and cellulose microfibrils (lying just outside) is seen in many types and shapes of plant cells and is present during both primary and secondary cell-wall deposition, suggesting a causal relationship.

This relationship can be tested by treating a plant tissue with a microtubule-depolymerizing drug so as to disassemble the entire system of cortical microtubules. The consequences for subsequent cellulose deposition, however, are not as straightforward as might be expected. The drug treatment does not disrupt the production of new cellulose microfibrils, and in some cases cells can continue to deposit new microfibrils in the preexisting orientation. Any developmental switch in the orientation of the microfibril pattern that would normally occur between successive lamellae, however, is invariably blocked. It seems that a preexisting orientation of microfibrils can be propagated even in the absence of microtubules, but any change in the deposition of cellulose microfibrils requires that intact microtubules be present to determine the new orientation.

These observations are consistent with the following model. The cellulose-synthesizing rosettes embedded in the plasma membrane spin out long cellulose molecules. As the synthesis of cellulose molecules and their self-assembly into microfibrils proceeds, the distal end of each microfibril presumably forms indirect cross-links to the previous layer of wall material, orienting the new microfibril in parallel with the old ones as it becomes integrated into the texture of the wall. Since the microfibril is stiff, the rosette at its growing, proximal end has to move as it deposits the new material, traveling in the plane of the membrane in the direction defined by the way in which the far end of the microfibril is anchored in the existing wall. In this way, each layer of microfibrils would tend to be spun out from the membrane in the same orientation as the layer laid down previously, with the rosettes following the direction of the preexisting oriented microfibrils outside the cell. Oriented microtubules inside the cell, however, can force a change in the direction in which the rosettes move: they can create boundaries in the plasma membrane that act like the banks of a canal to constrain rosette movement (**Figure 19–82**). In this view, cellulose synthesis can occur independently of microtubules; but it is constrained spatially when cortical microtubules are present to define membrane microdomains within which the enzyme complex can move.

In this way, plant cells can change their direction of expansion by a sudden change in the orientation of their cortical array of microtubules. Because plant cells cannot move (being constrained by their walls), the entire morphology of a multicellular plant presumably depends on a coordinated, highly patterned deployment of cortical microtubule orientations during plant development. It is not known how these orientations are controlled, although it has been shown

that the microtubules can reorient rapidly in response to extracellular stimuli, including low-molecular-weight plant growth regulators such as ethylene and gibberellic acid (see Figure 22–119).

Microtubules are not, however, the only cytoskeletal elements that influence wall deposition. Local foci of cortical actin filaments can also direct the deposition of new wall material at specific sites on the cell surface, contributing to the elaborate final shaping of many differentiated plant cells, such as the leaf trichomes and epidermal cells shown in Figure 19–77. In cells that cannot organize their cortical actin network, as in *Arp2/3* mutants, the cells lose their characteristic shapes, as illustrated in Figure 16–35.

Summary

Plant cells are surrounded by a tough extracellular matrix, or cell wall, which is responsible for many of the unique features of a plant's life style. The wall is composed of a network of cellulose microfibrils and cross-linking glycans, embedded in a highly cross-linked matrix of pectic polysaccharides. In secondary cell walls, lignin may be deposited making them waterproof, hard, and woody. A cortical array of microtubules can control the orientation of newly deposited cellulose microfibrils, which in turn determine the direction of cell expansion and therefore the final shape of the cell and, ultimately, of the plant as a whole.

PROBLEMS

Which statements are true? Explain why or why not.

19–1 Given the numerous processes inside cells that are regulated by changes in Ca^{2+} concentration, it seems likely that Ca^{2+}-dependent cell–cell adhesions are also regulated by changes in Ca^{2+} concentration.

19–2 Tight junctions perform two distinct functions: they seal the space between cells to restrict paracellular flow and they fence off membrane domains to prevent mixing of apical and basolateral proteins.

19–3 Integrins can convert mechanical signals into molecular signals.

19–4 The elasticity of elastin derives from its high content of α helices, which act as molecular springs.

Discuss the following problems.

19–5 Comment on the following (1922) quote from Warren Lewis, who was one of the pioneers of cell biology. "Were the various types of cells to lose their stickiness for one another and for the supporting extracellular matrix, our bodies would at once disintegrate and flow off into the ground in a mixed stream of cells."

19–6 Cell adhesion molecules were originally identified using antibodies raised against cell-surface components to block cell aggregation. In the adhesion-blocking assays, the researchers found it necessary to use antibody fragments, each with a single binding site (so-called Fab fragments), rather than intact IgG antibodies, which are Y-shaped molecules with two identical binding sites. The Fab fragments were generated by digesting the IgG antibodies with papain, a protease, to separate the two binding sites (**Figure Q19–1**). Why do you suppose it was necessary to use Fab fragments to block cell aggregation?

19–7 The food poisoning bacterium *Clostridium perfringens* makes a toxin that binds to members of the claudin family of proteins, which are the main constituents of tight junctions. When bound to a claudin, the C-terminus of the toxin allows the N-terminus to insert into the adjacent cell membrane, forming holes that kill the cell. The portion of the toxin that binds to the claudins has proven to be a valuable reagent for investigating the properties of tight junctions. MDCK cells are a common choice for studies of tight junctions because they can form an intact epithelial sheet with high transepithelial resistance. MDCK cells express two claudins: claudin-1, which is not bound by the toxin, and claudin-4, which is.

When an intact MDCK epithelial sheet is incubated with the C-terminal toxin fragment, claudin-4 disappears, becoming undetectable within 24 hours. In the absence of claudin-4, the cells remain healthy and the epithelial sheet appears intact. The mean number of strands in the tight junctions that link the cells also decreases over 24 hours from about four to about two, and they are less highly branched. A functional assay for the integrity of the tight junctions shows that transepithelial resistance decreases dramatically in the presence of the toxin, but the resistance can be restored by washing it out (**Figure Q19–2A**). Curiously, the toxin produces these effects only when it is added to the basolateral side of the sheet; it has no effect when added to the apical surface (Figure Q19–2B).

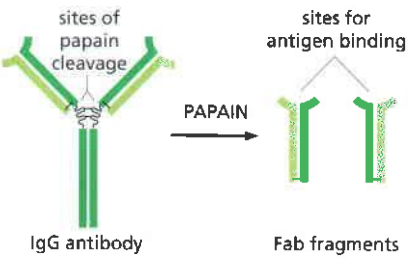


Figure Q19–1 Production of Fab fragments from IgG antibodies by digestion with papain (Problem 19–6).

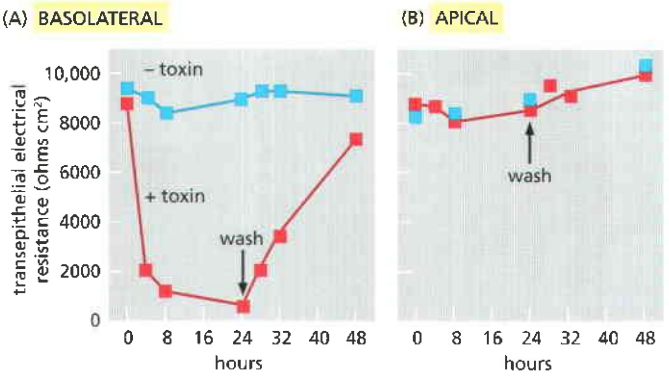


Figure Q19–2 Effects of *Clostridium* toxin on the barrier function of MDCK cells (Problem 19–7). (A) Addition of toxin from the basolateral side of the epithelial sheet. (B) Addition of toxin from the apical side of the epithelial sheet. The higher the resistance (ohms cm^2), the less the paracellular current for a given voltage.

A. How can it be that two tight junction strands remain, even though all of the claudin-4 has disappeared?
B. How do you suppose the toxin fragment causes the tight junction strands to disintegrate?
C. Why do you suppose the toxin works when it is added to the basolateral side of the epithelial sheet, but not when added to the apical side?

19–8 It is not an easy matter to assign particular functions to specific components of the basal lamina, since the overall structure is a complicated composite material with both mechanical and signaling properties. Nidogen, for example, cross-links two central components of the basal lamina by binding to the laminin-γ1 chain and to type IV collagen. Given such a key role, it was surprising that mice with a homozygous knockout of the gene for nidogen-1 were entirely healthy, with no abnormal phenotype. Similarly, mice homozygous for a knockout of the gene for nidogen-2 also appeared completely normal. By contrast, mice that were homozygous for a defined mutation in the gene for laminin-γ1, which eliminated just the binding site for nidogen, died at birth with severe defects in lung and kidney formation. The mutant portion of the laminin-γ1 chain is thought to have no other function than to bind nidogen, and does not affect laminin structure or its ability to assemble into basal lamina. How would you explain these genetic observations, which are summarized in **Table Q19–1**? What would you predict would be the phenotype of a mouse that was homozygous for knockouts of both nidogen genes?

19–9 Discuss the following statement: "The basal lamina of muscle fibers serves as a molecular bulletin board, in which adjoining cells can post messages that direct the differentiation and function of the underlying cells."

Table Q19–1 Phenotypes of mice with genetic defects in components of the basal lamina (Problem 19–8).

PROTEIN	GENETIC DEFECT	PHENOTYPE
nidogen-1	gene knockout (–/–)	none
nidogen-2	gene knockout (–/–)	none
laminin-γ1	nidogen binding-site deletion (+/–)	none
laminin-γ1	nidogen binding-site deletion (–/–)	dead at birth

+/– stands for heterozygous, –/– stands for homozygous.

19–10 The affinity of integrins for matrix components can be modulated by changes to their cytoplasmic domains: a process known as inside-out signaling. You have identified a key region in the cytoplasmic domains of αIIbβ3 integrin that seems to be required for inside-out signaling (**Figure Q19–3**). Substitution of alanine for either D723 in the β chain or R995 in the α chain leads to a high level of spontaneous activation, under conditions where the wild-type chains are inactive. Your advisor suggests that you convert the aspartate in the β chain to an arginine (D723R) and the arginine in the α chain to an aspartate (R995D). You compare all three α chains (R995, R995A, and R995D) against all three β chains (D723, D723A, and D723R). You find that all pairs have a high level of spontaneous activation, except D723 vs R995 (the wild type) and D723R vs R995D, which have low levels. Based on these results, how do you think the αIIbβ3 integrin is held in its inactive state?

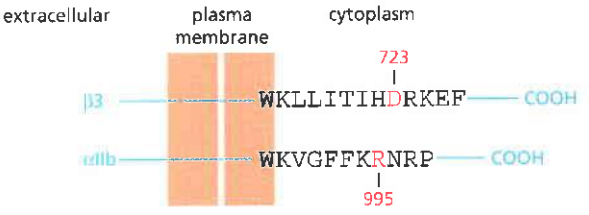


Figure Q19–3 Schematic representation of αIIbβ3 integrin (Problem 19–10). The D723 and R995 residues are indicated. (From P.E. Hughes et al. *J. Biol. Chem.* 271:6571–6574, 1996. With permission from the American Society for Biochemistry and Molecular Biology.)

19–11 Carboxymethyl Sephadex is a negatively charged crosslinked dextran that is commonly used for purifying proteins. It comes in the form of dry beads that swell tremendously when added to water. You have packed a chromatography column with the swollen gel. When you start equilibrating the column with a buffer that contains 50 mM NaCl at neutral pH, you are alarmed to see a massive shrinkage in gel volume. Why does the dry Sephadex swell so dramatically when it is placed in water? Why does the swollen gel shrink so much when a salt solution is added?

19–12 At body temperature, L-aspartate in proteins racemizes to D-aspartate at an appreciable rate. Most proteins in the body have a very low level of D-aspartate, if it can be detected at all. Elastin, however, has a fairly high level of D-aspartate. Moreover, the amount of D-aspartate increases in direct proportion to the age of the person from whom the sample was taken. Why do you suppose that most proteins have little if any D-aspartate, while elastin has high, age-dependent levels?

19–13 Your boss is coming to dinner! All you have for a salad is some wilted, day-old lettuce. You vaguely recall that there is a trick to rejuvenating wilted lettuce, but you can't remember what it is. Should you soak the lettuce in salt water, soak it in tap water, or soak it in sugar water, or maybe just shine a bright light on it and hope that photosynthesis will perk it up?

19–14 The hydraulic conductivity of a single water channel is 4.4×10^{-22} m^3 per second per MPa (megapascal) of pressure. What does this correspond to in terms of water molecules per second at atmospheric pressure? [Atmospheric pressure is 0.1 MPa (1 bar) and the concentration of water is 55.5 M.]

REFERENCES

General

Beckerle M (ed) (2002) Cell Adhesion. Oxford: Oxford University Press.

Kreis T & Vale R (ed) (1999) Guidebook to the Extracellular Matrix, Anchor, and Adhesion Proteins, 2nd ed. Oxford: Oxford University Press.

Miner JH (ed) (2005) Extracellular Matrix in Development and Disease (Advances in Developmental Biology, Vol 15). Elsevier Science.

Cadherins and Cell–Cell Adhesion

Garrod DR, Merritt AJ & Nie Z (2002) Desmosomal cadherins. *Curr Opin Cell Biol* 14:537–545.

Gumbiner BM (2005) Regulation of cadherin-mediated adhesion in morphogenesis. *Nature Rev Mol Cell Biol* 6:622–634.

King N, Hittinger CT & Carroll SB (2003) Evolution of key cell signaling and adhesion protein families predates animal origins. *Science* 301:361–363.

Kowalczyk AP & Reynolds AB (2004) Protecting your tail: regulation of cadherin degradation by p120-catenin. *Curr Opin Cell Biol* 16:522–527.

Nelson WJ & Nusse R (2004) Convergence of Wnt, beta-catenin, and cadherin pathways. *Science* 303:1483–1487.

Pokutta S & Weis WI (2007) Structure and mechanism of cadherins and catenins in cell-cell contacts. *Annu Rev Cell Dev Biol*, in press.

Rosen SD (2004) Ligands for L-selectin: homing, inflammation, and beyond. *Annu Rev Immunol* 22:129–156.

Scheiffele P, Fan J, Choih J, Fetter R & Serafini T (2000) Neuroligin expressed in nonneuronal cells triggers presynaptic development in contacting axons. *Cell* 101:657–669.

Shapiro L, Love J & Colman DR (2007) Adhesion molecules in the nervous system: structural insights into function and diversity. *Annu Rev Neurosci* 30:451–474.

Sheng M & Hoogenraad CC (2007) The postsynaptic architecture of excitatory synapses: a more quantitative view. *Annu Rev Biochem* 76:823–847.

Takeichi M (2007) The cadherin superfamily in neuronal connections and interactions. *Nature Rev Neurosci* 8:11–20.

Vestweber D (2002) Regulation of endothelial cell contacts during leukocyte extravasation. *Curr Opin Cell Biol* 14:587–593.

Yamagata M, Sanes JR & Weiner JA (2003) Synaptic adhesion molecules. *Curr Opin Cell Biol* 15:621–632.

Yang J, Mani SA, Donaher JL et al (2004) Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* 117:927–939.

Tight Junctions and the Organization of Epithelia

Furuse M, Hata M, Furuse K et al (2002) Claudin-based tight junctions are crucial for the mammalian epidermal barrier: a lesson from claudin-1-deficient mice. *J Cell Biol* 156:1099–1111.

Ikenouchi J, Furuse M, Furuse K et al (2005) Tricellulin constitutes a novel barrier at tricellular contacts of epithelial cells. *J Cell Biol* 171:939–945.

Henrique D & Schweisguth F (2003) Cell polarity: the ups and downs of the Par6/aPKC complex. *Curr Opin Genet Dev* 13:341–350.

Nelson WJ (2003) Adaptation of core mechanisms to generate cell polarity. *Nature* 422:766–774.

O'Brien LE, Jou TS, Pollack AL et al (2001) Rac1 orientates epithelial apical polarity through effects on basolateral laminin assembly. *Nature Cell Biol* 3:831–838.

Seifert JR & Mlodzik M (2007) Frizzled/PCP signalling: a conserved mechanism regulating cell polarity and directed motility. *Nature Rev Genet* 8:126–138.

Shin K, Fogg VC & Margolis B (2006) Tight junctions and cell polarity. *Annu Rev Cell Dev Biol* 22:207–236.

Passageways from Cell to Cell: Gap Junctions and Plasmodesmata

Cilia ML & Jackson D (2004) Plasmodesmata form and function. *Curr Opin Cell Biol* 16:500–506.

Gaietta G, Deerinck TJ, Adams SR et al (2002) Multicolor and electron microscopic imaging of connexin trafficking. *Science* 296:503–507.

Goodenough DA & Paul DL (2003) Beyond the gap: functions of unpaired connexon channels. *Nature Rev Mol Cell Biol* 4:285–294.

Segretain D & Falk MM (2004) Regulation of connexin biosynthesis, assembly, gap junction formation, and removal. *Biochim Biophys Acta* 1662:3–21.

Wei CJ, Xu X & Lo CW (2004) Connexins and cell signaling in development and disease. *Annu Rev Cell Dev Biol* 20:811–838.

The Basal Lamina

Miner JH & Yurchenco PD (2004) Laminin functions in tissue morphogenesis. *Annu Rev Cell Dev Biol* 20:255–284.

Sanes JR (2003) The basement membrane/basal lamina of skeletal muscle. *J Biol Chem* 278:12601–12604.

Sasaki T, Fassler R & Hohenester E (2004) Laminin: the crux of basement membrane assembly. *J Cell Biol* 164:959–963.

Yurchenco PD, Amenta PS & Patton BL (2004) Basement membrane assembly, stability and activities observed through a developmental lens. *Matrix Biol* 22:521–538.

Integrins and Cell–Matrix Adhesion

Galbraith CG, Yamada KM & Sheetz MP (2002) The relationship between force and focal complex development. *J Cell Biol* 159:695–705.

Hynes RO (2002) Integrins: bidirectional, allosteric signaling machines. *Cell* 110:673–687.

Luo BH & Springer TA (2006) Integrin structures and conformational signaling. *Curr Opin Cell Biol* 18:579–586.

Nayal A, Webb DJ & Horwitz AF (2004) Talin: an emerging focal point of adhesion dynamics. *Curr Opin Cell Biol* 16:94–98.

Vogel V & Sheetz M (2006) Local force and geometry sensing regulate cell functions. *Nature Rev Mol Cell Biol* 7:265–275.

The Extracellular Matrix of Animal Connective Tissues

Aszodi A, Legate KR, Nakchbandi I & Fassler R (2006) What mouse mutants teach us about extracellular matrix function. *Annu Rev Cell Dev Biol* 22:591–621.

Bulow HE & Hobert O (2006) The molecular diversity of glycosaminoglycans shapes animal development. *Annu Rev Cell Dev Biol* 22:375–407.

Couchman JR (2003) Syndecans: proteoglycan regulators of cell-surface microdomains? *Nature Rev Mol Cell Biol* 4:926–937.

Hotary KB, Allen ED, Brooks PC et al (2003) Membrane type I matrix metalloproteinase usurps tumor growth control imposed by the three-dimensional extracellular matrix. *Cell* 114:33–45.

Kielty CM, Sherratt MJ & Shuttleworth CA (2002) Elastic fibres. *J Cell Sci* 115:2817–2828.

Larsen M, Artym VV, Green JA & Yamada KM (2006) The matrix reorganized: extracellular matrix remodeling and integrin signaling. *Curr Opin Cell Biol* 18:463–471.

Lin X (2004) Functions of heparan sulfate proteoglycans in cell signaling during development. *Development* 131:6009–6021.

Mott JD & Werb Z (2004) Regulation of matrix biology by matrix metalloproteases. *Curr Opin Cell Biol* 16:558–564.

Prockop DJ & Kivirikko KI (1995) Collagens: molecular biology, diseases, and potentials for therapy. *Annu Rev Biochem* 64:403–434.

Toole BP (2001) Hyaluronan in morphogenesis. *Semin Cell Dev Biol* 12:79–87.

The Plant Cell Wall

Carpita NC & McCann M (2000) The Cell Wall. In Biochemistry and Molecular Biology of Plants (Buchanan BB,Gruissem W & Jones RL eds), pp 52–108. Rockville, MD: ASPB.

Carpita NC, Campbell M & Tierney M (eds) (2001) Plant cell walls. *Plant Mol Biol* 47:1–340. (special issue)

Dhugga KS (2001) Building the wall: genes and enzyme complexes for polysaccharide synthases. *Curr Opin Plant Biol* 4:488–493.

McCann MC & Roberts K (1991) Architecture of the primary cell wall. In The Cytoskeletal Basis of Plant Growth and Form (Lloyd CW ed), pp 109–129. London: Academic Press.

Smith LG & Oppenheimer DG (2005) Spatial control of cell expansion by the plant cytoskeleton. *Annu Rev Cell Dev Biol* 21:271–295.

Somerville C (2006) Cellulose synthesis in higher plants. *Annu Rev Cell Dev Biol* 22:53–78.

20

Cancer

About one in five of us will die of cancer, but that is not why we devote a chapter to this disease. Cancer cells break the most basic rules of cell behavior by which multicellular organisms are built and maintained, and they exploit every kind of opportunity to do so. In studying these transgressions, we discover what the normal rules are and how they are enforced. Thus, in the context of cell biology, cancer has a unique importance, and the emphasis given to cancer research has profoundly benefited a much wider area of biomedical science than that of cancer alone.

The effort to combat cancer has led to many fundamental discoveries in cell biology. Many proteins have been discovered because abnormalities in their function can lead to uncontrolled growth, increased division, decreased death, or other aberrant characteristics of cancer cells. Among them are proteins involved in DNA repair (Chapter 5), cell signaling (Chapter 15), the cell cycle and cell growth (Chapter 17), programmed cell death (apoptosis, Chapter 18), and tissue architecture (Chapter 19).

In this chapter, we first examine the nature of cancer by describing the natural history of the disease from a cellular standpoint. In a second section, we discuss the molecular changes that make a cell cancerous. Finally, we end the chapter by discussing how our enhanced understanding of the molecular basis of cancer is leading to improved methods for its prevention and treatment.

In This Chapter

CANCER AS A MICROEVOLUTIONARY PROCESS 1205

THE PREVENTABLE CAUSES OF CANCER 1224

FINDING THE CANCER-CRITICAL GENES 1230

THE MOLECULAR BASIS OF CANCER-CELL BEHAVIOR 1241

CANCER TREATMENT: PRESENT AND FUTURE 1256

CANCER AS A MICROEVOLUTIONARY PROCESS

The body of an animal operates as a society or ecosystem. The individual members are cells that reproduce by cell division and organize into collaborative assemblies called tissues. This society is very peculiar, however, because self-sacrifice—as opposed to survival of the fittest—is the rule. Ultimately, all of the somatic cell lineages in animals are committed to die: they leave no progeny and instead dedicate their existence to support of the germ cells, which alone have a chance of continued survival (discussed in Chapter 21). There is no particular mystery in this, for the body is a clone derived from a fertilized egg, and the genome of the somatic cells is the same as that of the germ cell lineage that gives rise to sperm or eggs. By their self-sacrifice for the sake of the germ cells, the somatic cells help to propagate copies of their own genes.

Thus, unlike free-living cells such as bacteria, which compete to survive, the cells of a multicellular organism are committed to collaboration. To coordinate their behavior, the cells send, receive, and interpret an elaborate set of extracellular signals that serve as *social controls*, directing each of them how to act (discussed in Chapter 15). As a result, each cell behaves in a socially responsible manner—resting, growing, dividing, differentiating, or dying—as needed for the good of the organism. Molecular disturbances that upset this harmony mean trouble for a multicellular society. In a human body with more than 10^{14} cells, billions of cells experience mutations every day, potentially disrupting the social controls. Most dangerously, a mutation may give one cell a selective advantage, allowing it to grow and divide more vigorously and survive more readily than its neighbors and to become a founder of a growing mutant clone. A mutation that

promotes such selfish behavior by individual members of the cooperative can jeopardize the future of the whole enterprise. Over time, repeated rounds of mutation, competition, and natural selection operating within the population of somatic cells can cause matters to go from bad to worse. These are the basic ingredients of cancer: it is a disease in which an individual mutant clone of cells begins by prospering at the expense of its neighbors, but in the end the descendants of this clone can destroy the whole cellular society.

In this section, we discuss the development of cancer as a microevolutionary process. In a human, this process occurs on a time scale of years or decades, in a subpopulation of cells in the body. But the final result depends on the same principles of mutation and natural selection that have driven the evolution of living organisms for billions of years.

Cancer Cells Reproduce Without Restraint and Colonize Other Tissues

Cancer cells are defined by two heritable properties: (1) they reproduce in defiance of the normal restraints on cell growth and division, and (2) they invade and colonize territories normally reserved for other cells. It is the combination of these properties that makes cancers particularly dangerous. An abnormal cell that grows (increases in mass) and proliferates (divides) out of control will give rise to a tumor, or *neoplasm*—literally, a new growth. As long as the neoplastic cells do not become invasive, however, the tumor is said to be **benign**, and removing or destroying the mass locally usually achieves a complete cure. A tumor is considered a cancer only if it is **malignant**, that is, only if its cells have acquired the ability to invade surrounding tissue. Invasiveness is an essential characteristic of cancer cells. It allows them to break loose, enter blood or lymphatic vessels, and form secondary tumors, called **metastases**, at other sites in the body (**Figure 20–1**). The more widely a cancer spreads, the harder it becomes to eradicate, and it is generally metastases that kill the cancer patient.

Cancers are classified according to the tissue and cell type from which they arise. **Carcinomas** are cancers arising from epithelial cells, and they are by far the most common cancers in humans. **Sarcomas** arise from connective tissue or muscle cells. Cancers that do not fit in either of these two broad categories include the various **leukemias** and **lymphomas**, derived from white blood cells and their precursors (hemopoietic cells), as well as cancers derived from cells of the nervous system. **Figure 20–2** shows the types of cancers that are common in the United States, together with their incidence and death rate. Each broad category has many subdivisions according to the specific cell type, location in the body, and the microscopic appearance of the tumor.

In parallel with the set of names for malignant tumors, there is a related set of names for benign tumors: an *adenoma*, for example, is a benign epithelial tumor with a glandular organization; the corresponding type of malignant tumor is an *adenocarcinoma* (**Figure 20–3**). Similarly a *chondroma* and a *chondrosarcoma* are, respectively, benign and malignant tumors of cartilage.

Most cancers have characteristics that reflect their origin. Thus, for example, the cells of a *basal-cell carcinoma*, derived from a keratinocyte stem cell in the skin, generally continue to synthesize cytokeratin intermediate filaments, whereas the cells of a *melanoma*, derived from a pigment cell in the skin, will



Figure 20–1 Metastasis. Malignant tumors typically give rise to metastases, making the cancer hard to eradicate. Shown in this fusion image is a whole-body scan of a patient with metastatic non-Hodgkins lymphoma (NHL). The background image of the body's tissues was obtained by CT (computed X-ray tomography) scanning. Overlaid on this image, a PET (positron emission tomography) scan that detects the uptake of radioactively labeled fluorodeoxyglucose (FDG) in various tissues reveals the tumor tissue (*yellow*). High FDG uptake indicates cells with unusually active glucose uptake and metabolism, a characteristic of tumors. The yellow spots in the abdominal region reveal multiple NHL metastases. (Courtesy of S. Gambhir.)

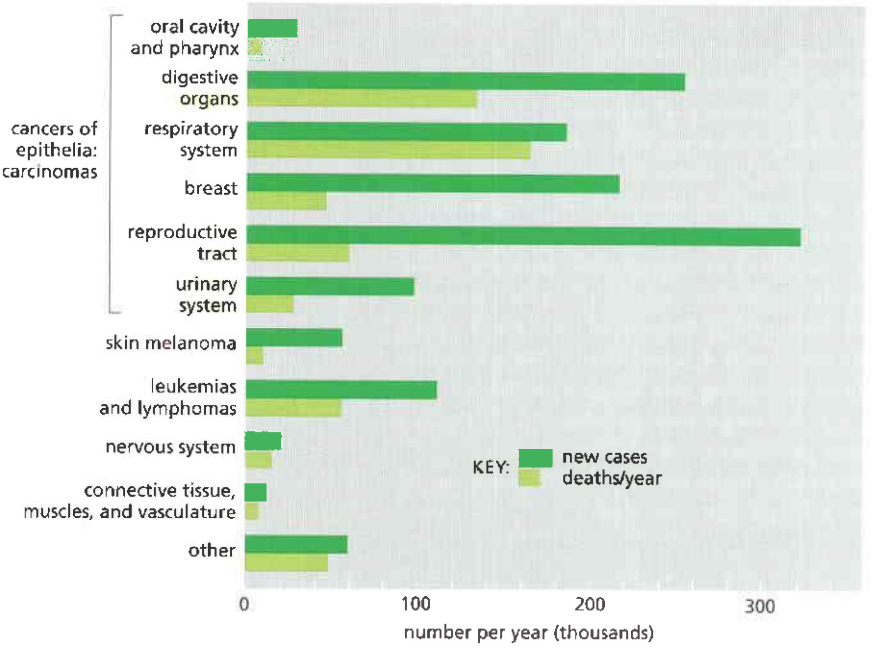


Figure 20–2 Cancer incidence and mortality in the United States. The total number of new cases diagnosed in 2004 in the United States was 1,368,030, and total cancer deaths were 563,700. Note that deaths reflect cases diagnosed at many different times and that somewhat less than half of the people who develop cancer die of it. In the world as a whole, the five most common cancers are those of the lung, stomach, breast, colon/rectum, and uterine cervix (included in the figure under the heading of reproductive tract), and the total number of new cancer cases recorded per year is just over 6 million. Skin cancers other than melanomas are not included in these figures, since almost all are cured easily and many are unrecorded. (Data from American Cancer Society, Cancer Facts and Figures, 2004.)

often (but not always) continue to make pigment granules. Cancers originating from different cell types are, in general, very different diseases. Basal-cell carcinomas of the skin, for example, are only locally invasive and rarely metastasize, whereas melanomas can become much more malignant and often form metastases (reflecting the migratory behavior of normal pigment-cell precursors during development, discussed in Chapter 22). Basal-cell carcinomas are readily cured by surgery or local irradiation, whereas malignant melanomas, once they have metastasized widely, are usually fatal.

About 80% of human cancers are carcinomas, perhaps because most of the cell proliferation in adults occurs in epithelia or because epithelial tissues are most frequently exposed to the various forms of physical and chemical damage that favor the development of cancer.

Most Cancers Derive from a Single Abnormal Cell

Even when a cancer has metastasized, we can usually trace its origins to a single **primary tumor**, arising in a specific organ. The primary tumor is thought to derive by cell division from a single cell that initially experienced some heritable change. Subsequently, additional changes accumulate in some of the descendants of this cell, allowing them to outgrow, out-divide, and often outlive their neighbors. By the time it is first detected, a typical human cancer will have been developing for many years and will already contain about a billion cancer cells or more (**Figure 20–4**). Tumors will usually also contain a variety of other cell types—fibroblasts, for example, in the supporting connective tissue associated

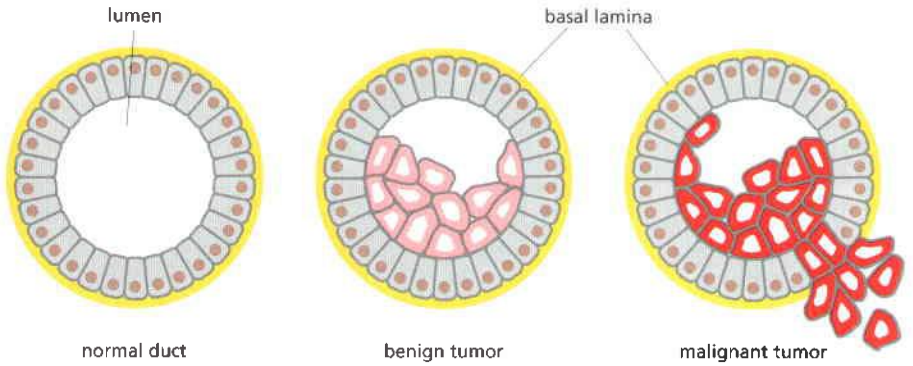


Figure 20–3 Benign versus malignant tumors. A benign glandular tumor (an adenoma) remains inside the basal lamina that marks the boundary of the normal structure (a duct, in this example), whereas a malignant glandular tumor (an adenocarcinoma) destroys duct integrity as shown. There are many forms that such tumors may take.

with a carcinoma, as well as inflammatory cells. How can we be sure that the cancer cells are a clone descended from a single abnormal cell?

Molecular analyses of chromosomes in tumor cells clearly demonstrate the clonal origin of those cells. In almost all patients with *chronic myelogenous leukemia (CML)*, for example, we can distinguish leukemic white blood cells from normal cells by a specific chromosomal abnormality: the so-called *Philadelphia chromosome*, created by a translocation between the long arms of chromosomes 9 and 22 (**Figure 20–5**). When the DNA at the site of translocation is cloned and sequenced, it is found that the site of breakage and rejoining of the translocated fragments is identical in all the leukemic cells in any given patient, but that this site differs slightly (by a few hundred or thousand base pairs) from one patient to another. This is the expected result if, and only if, the cancer in each patient arises from a unique accident occurring in a single cell. We will see later how this translocation promotes the development of CML by creating a novel hybrid gene encoding a protein that promotes cell proliferation.

Many other lines of evidence, from a variety of cancers, point to the same conclusion: most cancers originate from a single aberrant cell (**Figure 20–6**).

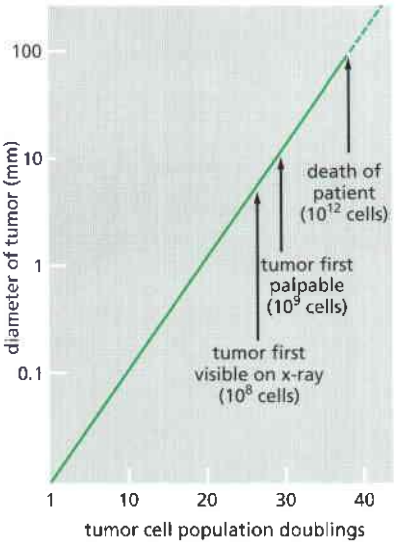


Figure 20–4 The growth of a typical human tumor, such as a tumor of the breast. The diameter of the tumor is plotted on a logarithmic scale. Years may elapse before the tumor becomes noticeable. The doubling time of a typical breast tumor, for example, is about 100 days.

Cancer Cells Contain Somatic Mutations

If a single abnormal cell is to give rise to a tumor, it must pass on its abnormality to its progeny: the aberration has to be heritable. One problem in understanding a cancer is to discover whether a particular heritable aberration is due to a genetic change—that is, an alteration in the cell's DNA sequence—or to an *epigenetic* change—that is, a persistent change in the pattern of gene expression without a change in the DNA sequence. Epigenetic changes occur during normal development, and they are responsible for the stability of the differentiated state and for such phenomena as X-chromosome inactivation and genomic imprinting (discussed in Chapter 7). As we shall discuss later, epigenetic changes are now known to play an important part in the development of cancers.

But it is also clear that the development of a clone of cancer cells also involves genetic changes. The cells of many cancers can be shown to have one or more shared detectable abnormalities in their DNA sequence that distinguish them from the normal cells surrounding the tumor, as in the example of CML just described, and advances in DNA analysis have identified such genetic faults in an increasing proportion of cancer types. This agrees with the finding that many of the agents that provoke the development of cancer also cause genetic changes. Thus, **carcinogenesis** (the generation of cancer) appears to be linked to *mutagenesis* (the production of a change in the DNA sequence). This correlation is particularly clear for two classes of agents: (1) **chemical carcinogens** (which typically cause simple local changes in the nucleotide sequence), and (2) *radiation* such as x-rays (which typically cause chromosome breaks and translocations) or UV light (which causes specific DNA base alterations). We shall discuss these carcinogenic agents in detail later.

The conclusion that the development of cancer depends on somatic mutations is further supported by the finding that people who have inherited a genetic defect in one of several DNA repair mechanisms, causing their cells to accumulate mutations at an elevated rate, show a strong predisposition to develop cancer. People with the disease *xeroderma pigmentosum*, for example, have defects in the system that repairs DNA damage induced by UV light, and they have a greatly increased incidence of skin cancers. Likewise, mice lacking specific DNA repair genes are abnormally prone to cancer.

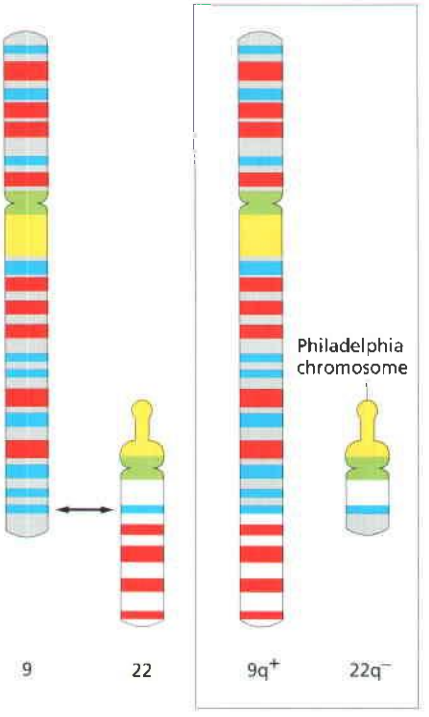
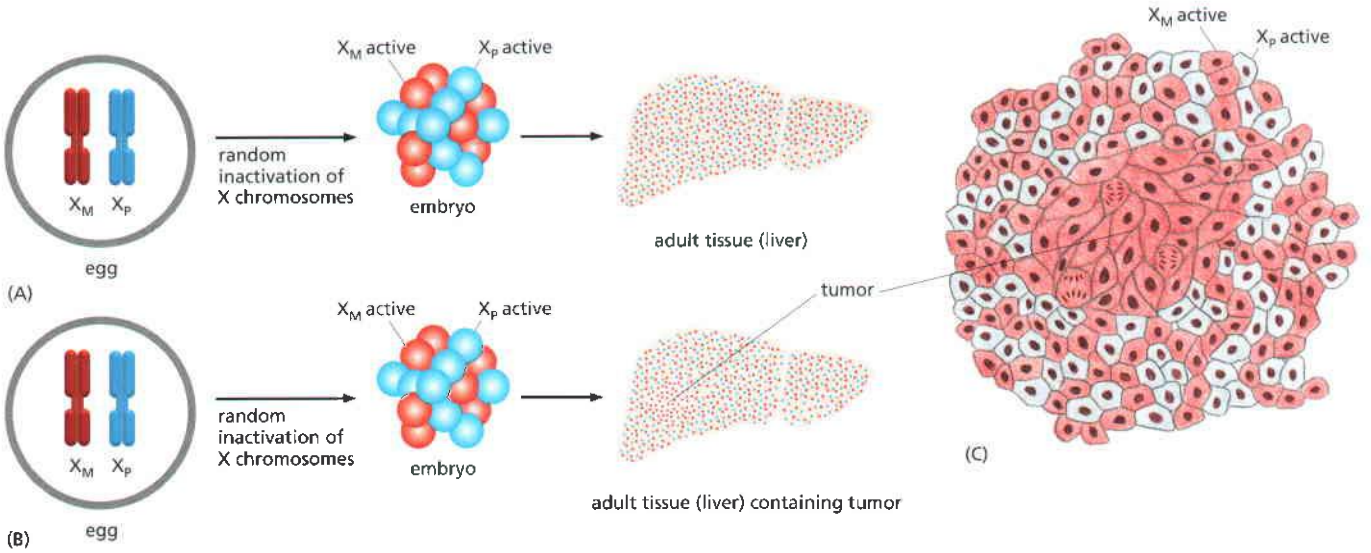


Figure 20–5 The translocation between chromosomes 9 and 22 responsible for chronic myelogenous leukemia. The normal structures of chromosomes 9 and 22 are shown at the left. When a translocation occurs between them at the indicated site, the result is the abnormal pair at the right. The smaller of the two resulting abnormal chromosomes ($22q^-$) is called the Philadelphia chromosome, after the city where the abnormality was first recorded.



A Single Mutation Is Not Enough to Cause Cancer

An estimated 10^{16} cell divisions occur in a normal human body in the course of a typical lifetime; in a mouse, with its smaller number of cells and its shorter life span, the number is about 10^{12}. Even in an environment that is free of mutagens, mutations would occur spontaneously at an estimated rate of about 10^{-6} mutations per gene per cell division—a value set by fundamental limitations on the accuracy of DNA replication and repair (see p. 263). Thus, in a typical lifetime, every single gene is likely to have undergone mutation on about 10^{10} separate occasions in a human, or on about 10^6 occasions in a mouse. Among the resulting mutant cells, we might expect a large number that have sustained deleterious mutations in genes that regulate cell growth and division, causing the cells to disobey the normal restrictions on cell proliferation. From this point of view, the problem of cancer seems to be not why it occurs, but why it occurs so infrequently.

Clearly, if a single mutation were enough to convert a typical healthy cell into a cancer cell that proliferates without restraint, we would not be viable organisms. Many lines of evidence indicate that the development of a cancer typically requires that a substantial number of independent, rare genetic accidents occur in the lineage of one cell. One comes from epidemiological studies of the incidence of cancer as a function of age (**Figure 20–7**). If a single mutation were responsible, occurring with a fixed probability per year, the chance of developing cancer in any given year of life should be independent of age. In fact, for most types of cancer, the incidence rises steeply with age—as would be expected if cancer is caused by a progressive accumulation of random mutations in a single lineage of cells.

Now that many of the specific mutations responsible for the development of cancer have been identified, we can test directly for their presence in the genome of a particular type of cancer cell. Such tests indicate that a human malignant cell generally harbors many mutations. Animal models confirm that a single genetic alteration is insufficient to cause cancer: in genetically engineered mice carrying one cancer-causing mutation, the disease occurs only

Figure 20–6 Evidence from X-inactivation mosaics that demonstrates the monoclonal origin of cancers. (A) As a result of a random process that occurs in the early embryo, practically every normal tissue in a woman's body is a mixture of cells with different X chromosomes heritably inactivated (indicated here by the mixture of *red* cells and *blue* cells in the normal tissue). (B) When the cells of a cancer are tested for their expression of an X-linked marker gene (a specific form of the enzyme G6PD), however, they are usually all found to have the same X chromosome inactivated, as shown in (C). This implies that they are all derived from a single cancerous founder cell.

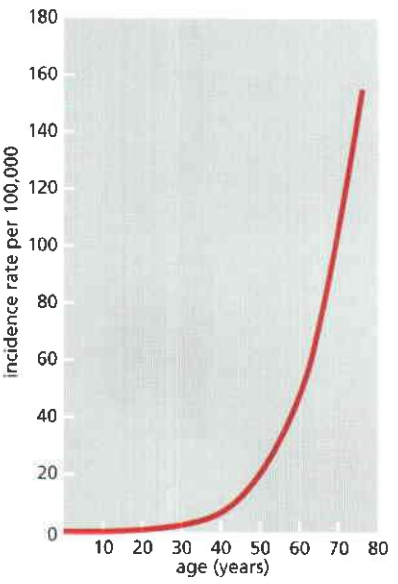


Figure 20–7 Cancer incidence as a function of age. The number of newly diagnosed cases of colon cancer in women in England and Wales in 1 year is plotted as a function of age at diagnosis and expressed relative to the total number of individuals in each age group. The incidence of cancer rises steeply as a function of age. If only a single mutation were required to trigger the cancer and this mutation had an equal chance of occurring at any time, the incidence would be independent of age. (Data from C. Muir et al., Cancer Incidence in Five Continents, Vol. V. Lyon: International Agency for Research on Cancer, 1987.)

after a lag period of months and arises only in a small scattered subset of cells in the body, implying that other genetic events are necessary for cancer to develop. Furthermore, engineered mice with mutations in more than one cancer-critical gene develop cancer faster than mice with single mutations, demonstrating that multiple genetic events in a single cell can cooperate to promote cancerous cell growth and the development of a tumor.

Cancers Develop Gradually from Increasingly Aberrant Cells

For those cancers known to have a specific external cause, the disease does not usually become apparent until long after exposure to the causal agent. The incidence of lung cancer, for example, does not begin to rise steeply until after 20 years of heavy smoking. Similarly, the incidence of leukemias in Hiroshima and Nagasaki did not show a marked rise until about 5 years after the explosion of the atomic bombs, and industrial workers exposed for a limited period to chemical carcinogens do not usually develop the cancers characteristic of their occupation until 10, 20, or even more years after the exposure (**Figure 20–8**). During this long incubation period, the prospective cancer cells undergo a succession of genetic and epigenetic changes. The same seems to apply to cancers where the initial genetic lesion has no such obvious external cause.

The concept that the development of a cancer requires a gradual accumulation of mutations in a number of different genes—different for different cancers, but usually at least five—helps to explain the well-known phenomenon of **tumor progression**, whereby an initial mild disorder of cell behavior evolves gradually into a full-blown cancer. Chronic myelogenous leukemia provides a clear example. It begins as a disorder characterized by a nonlethal overproduction of white blood cells and continues in this form for several years before changing into a much more rapidly progressing illness that usually ends in death within a few months. In the early chronic phase, the leukemic cells are distinguished mainly by the chromosomal translocation (the Philadelphia chromosome) mentioned previously, although there may well be other less visible genetic or epigenetic changes. In the subsequent acute phase, cells that show not only the translocation but also several other chromosomal abnormalities overrun the hemopoietic (blood-forming) system. It appears as though cells of the initial mutant clone have undergone further mutations that make them proliferate even more extensively, so that they come to outnumber both the normal blood cells and the ancestor cells that have only the primary chromosomal translocation.

Carcinomas and other solid tumors are thought to evolve in a similar way. Although many such cancers in humans are not diagnosed until a relatively late stage, in some cases it is possible to observe the earlier steps. Cancers of the *uterine cervix* (the neck of the womb) are one such example, because of the routine screening for this cancer by cervical scraping.

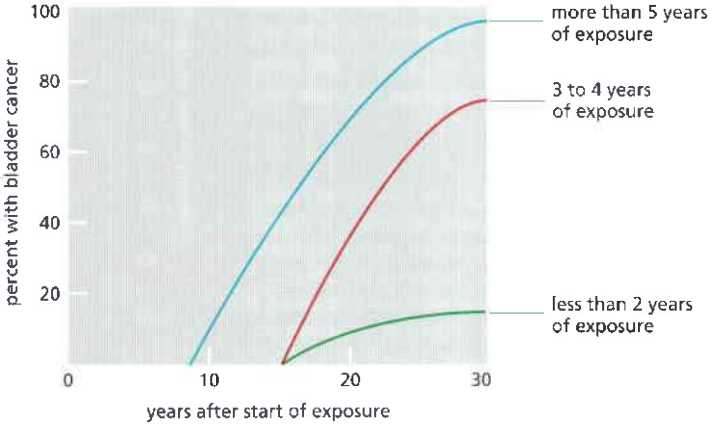


Figure 20–8 Delayed onset of cancer following exposure to a carcinogen. The graph shows the length of the delay before onset of bladder cancer in a set of 78 workers in the chemical industry who had been exposed to the carcinogen 2-naphthylamine, grouped according to the duration of their exposure. (Modified from J. Cairns, Cancer: Science and Society. San Francisco: W.H. Freeman, 1978. After M.H.C. Williams, in Cancer, Vol. III [R.W. Raven, ed.]. London: Butterworth Heinemann, 1958.)

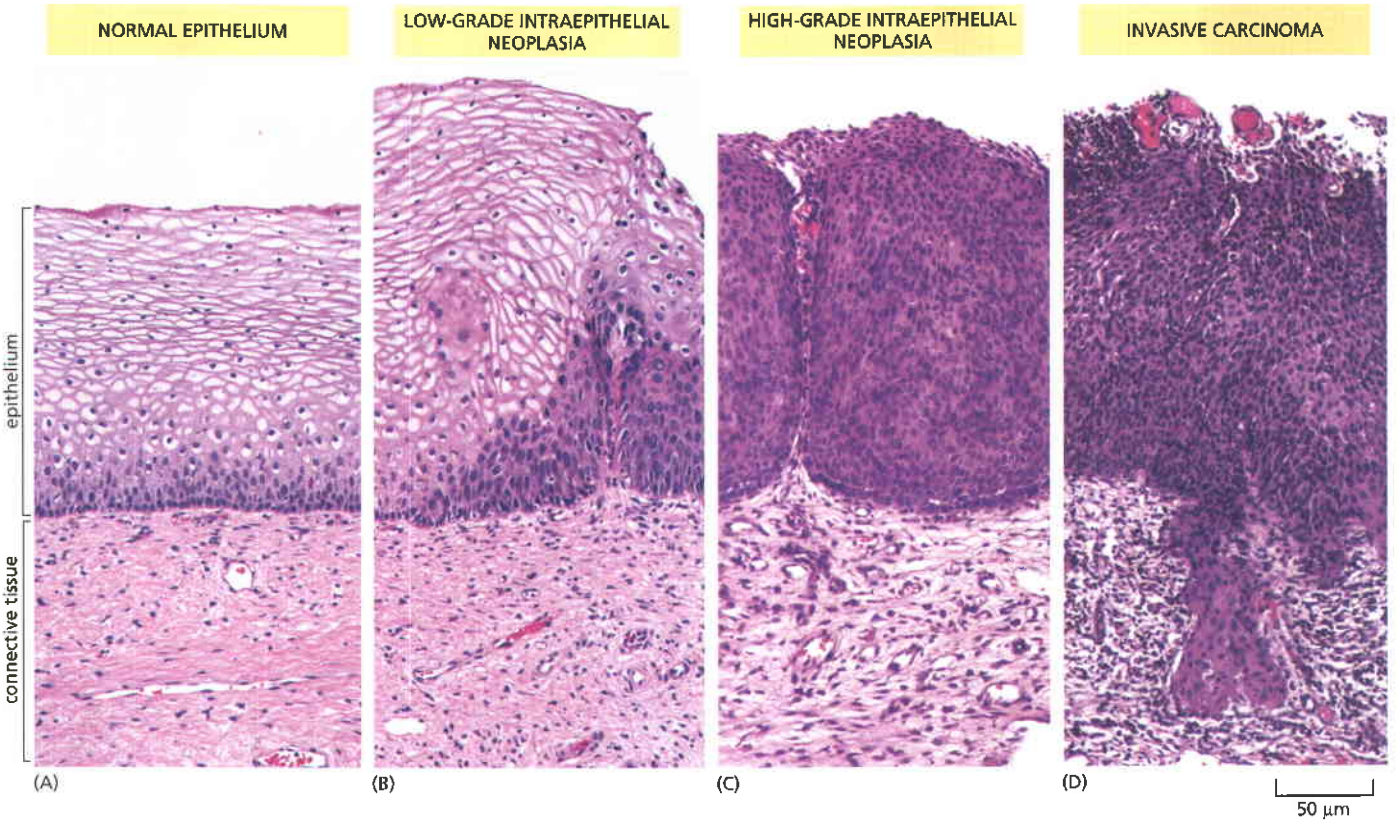


Figure 20–9 The stages of progression in the development of cancer of the epithelium of the uterine cervix. Pathologists use standardized terminology to classify the types of disorders they see, so as to guide the choice of treatment. (A) In the normal stratified squamous epithelium, dividing cells are confined to the basal layer. (B) In low-grade intraepithelial neoplasia, dividing cells can be found throughout the lower third of the epithelium; the superficial cells are still flattened and show signs of differentiation, but this is incomplete. (C) In high-grade intraepithelial neoplasia, cells in all the epithelial layers are proliferating and exhibit defective differentiation. (D) True malignancy begins when the cells move through or destroy the basal lamina that underlies the basal layer of epithelium and invade the underlying connective tissue. (Photographs courtesy of Andrew J. Connolly.)

Cervical Cancers Are Prevented by Early Detection

The epithelium covering the cervix is initially organized as a stratified (multilayered) squamous epithelium (**Figure 20–9A**), which is similar in structure to the epidermis of the skin (see Figure 23–3). In such stratified epithelia, cell proliferation normally occurs only in the basal layer, generating cells that stop dividing and then move out toward the surface, differentiating as they move to form flattened, keratin-rich cells that are eventually sloughed off when they reach the surface. When specimens of cervical epithelium from different women are examined, however, it is not unusual to find patches in which this organization is disturbed in a way that suggests the beginnings of a cancerous transformation. Pathologists describe these changes as *intraepithelial neoplasia*, and classify them as low-grade (mild) or high-grade (moderate to severe).

In the low-grade lesions, the undifferentiated dividing cells are no longer confined to the basal layer but occupy other layers in the lower third of the epithelium; although differentiation proceeds in the upper epithelial layers, it is somewhat disordered (Figure 20–9B). Left alone, most of these mild lesions will spontaneously regress, but about 10% progress to become high-grade lesions, in which most or all of the epithelial layers are occupied by undifferentiated dividing cells, which are usually highly variable in size and shape. Abnormal mitotic structures are frequently seen, and the *karyotype* (the display of the full set of chromosomes) is usually abnormal. But the aberrant cells are still confined to the epithelial side of the basal lamina (Figure 20–9C). At this stage, it is still easy to cure the condition by destroying or surgically removing the abnormal tissue. Fortunately, the presence of such lesions can be detected by

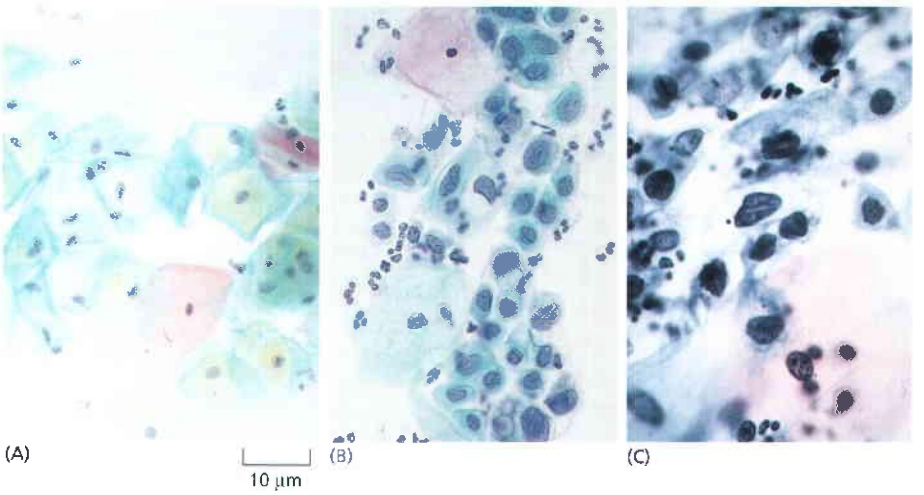


Figure 20–10 Photographs of cells collected by scraping the surface of the uterine cervix (the Papanicolaou or "Pap smear" technique). (A) Normal: the cells are large and well differentiated, with highly condensed nuclei. (B) Precancerous lesion: differentiation and proliferation are abnormal but the lesion is not yet invasive; the cells are in various stages of differentiation, some quite immature. (C) Invasive carcinoma: the cells all appear undifferentiated, with scanty cytoplasm and a relatively large nucleus. For all three panels, debris in the background includes some white blood cells. (Courtesy of Winifred Gray.)

scraping off a sample of cells from the surface of the cervix and viewing it under the microscope (the "Pap smear" technique—**Figure 20–10**), thereby saving many lives.

Without treatment, the abnormal patch of tissue may simply persist and progress no further or may even regress spontaneously. In at least 30–40% of cases, however, progression will occur, giving rise, over a period of several years, to a frank invasive carcinoma (see Figure 20–9D): the cancer cells cross or destroy the basal lamina, invade the underlying tissue, and may metastasize to local lymph nodes via lymphatic vessels. Surgical cure becomes progressively more difficult as the invasive growth spreads.

Tumor Progression Involves Successive Rounds of Random Inherited Change Followed by Natural Selection

From all the evidence, it seems that cancers arise by a process in which an initial population of slightly abnormal cells, descendants of a single abnormal ancestor, evolve from bad to worse through successive cycles of random inherited change followed by natural selection. At each stage, one cell acquires an additional mutation or epigenetic change that gives it a selective advantage over its neighbors, making it better able to thrive in its environment—an environment that, inside a tumor, may be harsh, with low levels of oxygen, scarce nutrients, and the natural barriers to growth presented by the surrounding normal tissues. The offspring of the best-adapted cell continue to divide, eventually taking over the tumor and becoming the dominant clone in the developing lesion (**Figure 20–11**). Tumors grow in fits and starts, as additional advantageous inherited changes arise and the cells bearing them flourish. Tumor progression involves a large element of chance and usually takes many years, which is why the majority of us will die of causes other than cancer.

Why are so many changes needed? Clearly, large animals have had to evolve a very complex set of regulatory mechanisms to keep their cells in check. Without multiple controls, inevitable errors in the maintenance of DNA sequences would

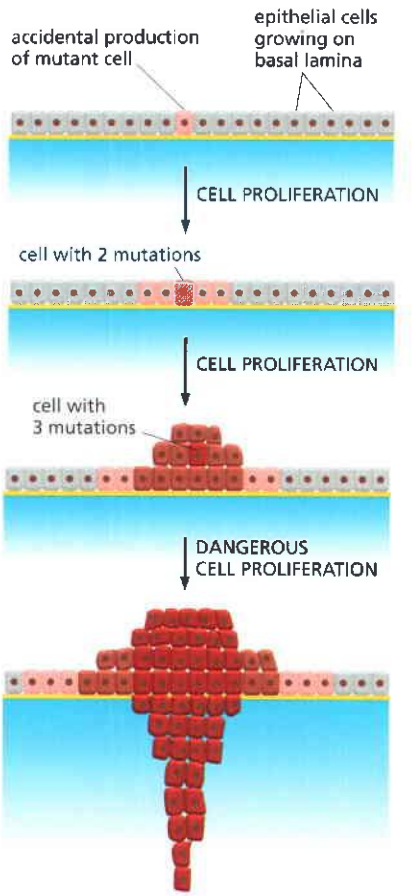


Figure 20–11 Clonal evolution. In this schematic diagram, a tumor develops through repeated rounds of mutation and proliferation, giving rise eventually to a clone of fully malignant cancer cells. At each step, a single cell undergoes a mutation that either enhances cell proliferation or decreases cell death, so that its progeny become the dominant clone in the tumor. Proliferation of each clone hastens the occurrence of the next step of tumor progression by increasing the size of the cell population at risk of undergoing an additional mutation. The final step depicted here is invasion through the basement membrane, an initial step in metastasis. In reality, there are more than the three steps shown here, and a combination of genetic and epigenetic changes are involved.

produce numerous tumors early in life and quickly destroy any large multicellular organism. Thus, we should not be surprised that cells employ multiple regulatory mechanisms to help them maintain tight and precise control over their behavior, and that many different regulatory systems have to be disrupted before a cell can throw off its normal restraints and behave as an asocial cancer cell. The restraints are of many types, and tumor cells meet new barriers to further expansion at each stage of their evolution. Oxygen and nutrients, for example, do not become limiting until a tumor is one or two millimeters in diameter, at which point the cells in the tumor interior may not have adequate access to these necessary resources. Cells must acquire additional mutations, epigenetic changes (or both) to overcome each new barrier, whether physical or physiological.

In general, the rate of evolution in any population of organisms on Earth would be expected to depend mainly on four parameters: (1) the *mutation rate*, that is, the probability per gene per unit time that any given member of the population will undergo genetic change; (2) the *number of reproducing individuals* in the population; (3) the *rate of reproduction*, that is, the average number of generations of progeny produced per unit time; and (4) the *selective advantage* enjoyed by successful mutant individuals, that is, the ratio of the number of surviving fertile progeny they produce per unit time to the number of surviving fertile progeny produced by nonmutant individuals. The same types of factors are also crucial for the evolution of cancer cells in a multicellular organism, except that both genetic and epigenetic changes help drive the evolutionary process.

The Epigenetic Changes That Accumulate in Cancer Cells Involve Inherited Chromatin Structures and DNA Methylation

As just stated, the progression toward cancer differs from normal biological evolution in one important respect: epigenetic changes also occur that give the cells a selective advantage.

For many years, pathologists have used an abnormal appearance of the cell nucleus to identify and classify cancer cells in tumor biopsies. For example, cancer cells sometimes contain an unusually large amount of heterochromatin—a condensed form of interphase chromatin that silences genes (see p. 238). We now understand some of the molecular mechanisms involved in forming this chromatin, and it is possible to associate new heterochromatin formation with the epigenetic silencing of specific genes that would otherwise block tumor progression.

Heterochromatin formation and maintenance involve specific covalent modifications of histones; these in turn attract complexes of chromatin-binding proteins that are stably maintained following DNA replication (discussed in Chapter 4). In this way, genes can be turned off in a cell-to-cell inherited manner without any change in the DNA sequence. This epigenetic form of gene regulation plays a major role in driving the orderly pattern of cell specializations that occur during embryonic development (discussed in Chapter 22). Errors that occur in this process are potentially dangerous because they can be transmitted to the progeny of the cell in which the original change occurred.

We now know that many of the mutations that make a cell cancerous alter the proteins that determine chromatin structures. These include not only enzymes that modify the histones in nucleosomes, but also proteins in the "code reader–writer" complexes that interpret the histone code (see Figure 4–43). These findings provide strong evidence for the importance of chromatin-based epigenetic changes in tumor progression.

During normal developmental processes, an inherited pattern of DNA methylation reinforces many of the gene silencing events caused by the packaging of genes into heterochromatin (see p. 467). Analyses reveal that a large amount of DNA methylation also occurs on selected genes during tumor progression. In summary, although recognized relatively recently, it now appears that the abnormal epigenetic silencing of genes is no less important than mutations in DNA sequences for the development of most cancers (**Figure 20–12**).

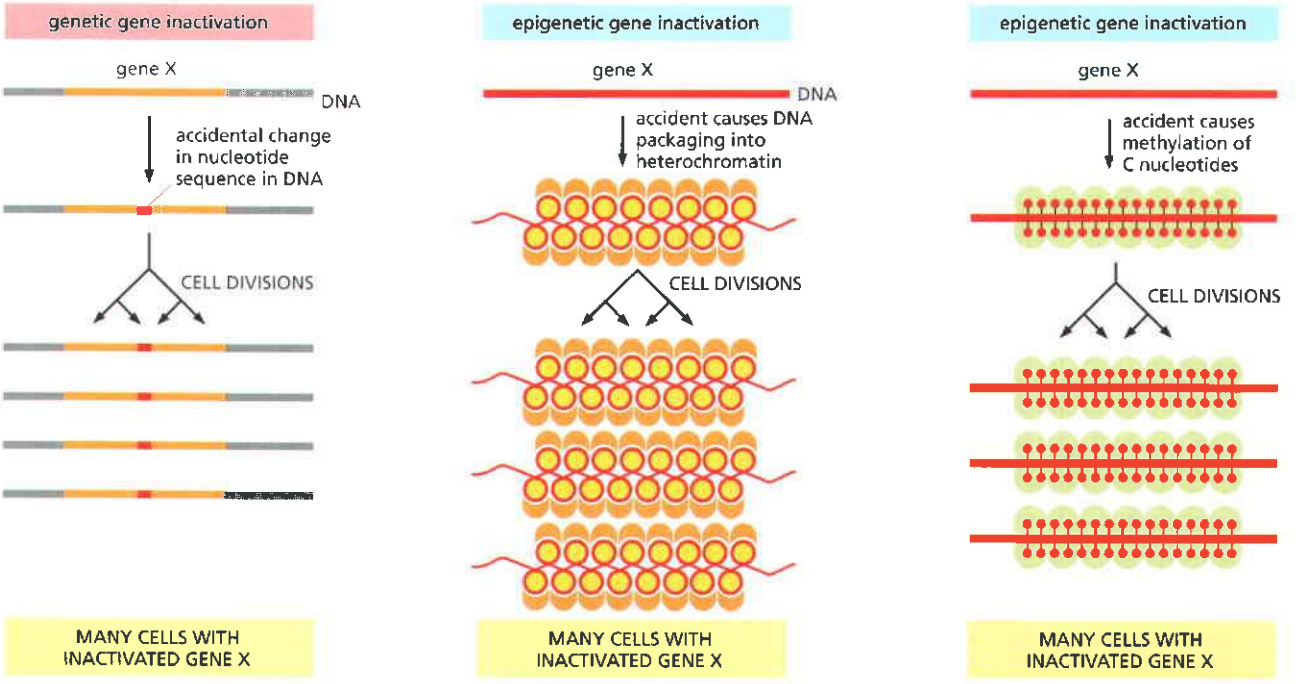


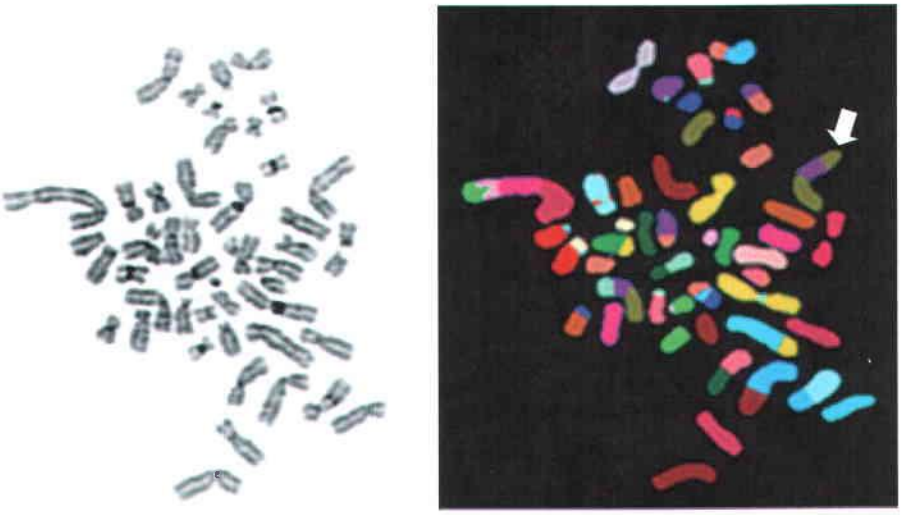
Figure 20–12 Comparison of the genetic and epigenetic changes observed in tumors. A mutation results from an irreversible change in DNA sequence. In contrast to such genetic changes, epigenetic changes are based on alterations that, while heritable from cell to cell, can be reversed either by site-specific changes in histone modification (heterochromatin pathway) or by site-specific DNA demethylation (methylation pathway). Because the epigenetic marks on genes are generally reversed during the formation of eggs and sperm, they are not inherited between generations and therefore have not been extensively studied by geneticists.

As discussed in Chapter 7, it is thought that DNA methylation (an inherited pattern of methylation of C nucleotides in CpG sequences) is a mechanism that is used to more permanently silence genes that have already been turned off, and DNA methylation seems to be a silencing mechanism that normally follows the formation of heterochromatin. But the misregulation that arises during tumor progression may be able to produce gene silencing independently of heterochromatinization, as schematically illustrated here.

Human Cancer Cells Are Genetically Unstable

Most human cancer cells accumulate genetic changes at an abnormally rapid rate and are said to be **genetically unstable**. This instability can take various forms. Some cancer cells are unable to repair certain types of DNA damage or to correct replication errors of various kinds. These cells tend to accumulate more point mutations or other DNA sequence changes than do normal cells. Other cancer cells fail to maintain either the number or the integrity of their chromosomes, and they consequently accumulate gross abnormalities in their karyotype that are visible at mitosis (**Figure 20–13**). The genetic instability is further amplified when some of the DNA changes alter epigenetic control mechanisms in ways that produce extra heterochromatin and DNA methylation. And epigenetic changes by themselves, arising accidentally and independently of prior genetic changes, may also lead—in principle at least—to destabilization of normal patterns of cellular inheritance by facilitating other changes either at the genetic or the epigenetic level. From an evolutionary perspective, none of this should be a surprise: anything that increases the probability of random changes in gene function that are heritable from one cell generation to the next is likely to speed the evolution of a clone of cells toward malignancy.

Although the epigenetic changes in cancer cells have not been so thoroughly analyzed, a great deal is now known about the genetic changes. Different tumors—even those from the same tissues—can show different kinds of genetic instability, caused by heritable alterations in any one of a large number of so-called *DNA maintenance genes* involved in propagating DNA or chromosomes. It is rare for a person to inherit a mutation in one of these genes, but those people that do so have a raised incidence of cancer, confirming that a loss of genetic stability can cause cancer. Generally, such destabilizing mutations



(A) (B)

Figure 20–13 Chromosomes from a breast tumor displaying abnormalities in structure and number. Chromosomes were prepared from a breast tumor cell in metaphase, spread on a glass slide, and stained with (A) a general DNA stain or (B) a combination of fluorescent stains that color each normal human chromosome differently. The staining (displayed in false color) shows multiple translocations, including a doubly translocated chromosome *(white arrow)* made up of two pieces of chromosome 8 *(brown)* and a piece of chromosome 17 *(purple)*. The karyotype also contains 48 chromosomes, instead of the normal 46. (Courtesy of Joanne Davidson and Paul Edwards.)

are not inherited but arise *de novo* in the clone of cells in which a tumor develops, helping the cancer cell to accumulate mutations much more rapidly than its neighbors. Recent analyses show that the cells in a variety of human cancers experience single nucleotide substitutions at a rate that is 10–20 times higher than the rate observed in normal cells. As a result, the sequencing of more than 10,000 genes in a set of breast and colorectal cancers reveals that cancer cells have accumulated enough mutations to cause an amino acid change in the proteins of roughly 100 genes. Most are random changes that affect different genes in different individual tumors. But a subset of genes are found to be repeatedly mutated in a particular type of cancer, suggesting that alterations in as many as 20 genes are needed to drive tumor progression.

Genetic instability does not in itself give a cell a selective advantage. It seems that there is some optimum level of genetic instability for the development of cancer, making a cell mutable enough to evolve rapidly, but not so mutable that it accumulates too many harmful changes and dies. To be favored by selection, a genetically unstable cell must acquire properties that confer some competitive advantage.

Cancerous Growth Often Depends on Defective Control of Cell Death, Cell Differentiation, or Both

Just as an increased mutation rate per cell can raise the probability of cancer, so can any circumstance that increases the number of proliferating cells available for mutating. People who are clinically obese, for example, have a strongly increased risk of many types of cancer, compared with people of normal weight; and this is presumably due, in part at least, to an increase in both the number of cells in the body and the rate at which these cells divide when overnourished or overstimulated by growth factors. The same principle applies to both cancer initiation and cancer progression: the bigger the clone of altered cells resulting from an early inherited change, the greater the chance that at least one of these cells will undergo an additional mutation or epigenetic change that will allow the cancer to progress. Thus, at every stage in the development of cancer, any condition that helps the altered cells to increase in number favors the progression of the tumor.

An early mutation or epigenetic change can have this effect by increasing the rate at which a clone of cells proliferates, as we discuss in detail later. Such changes, however, are not the only—or necessarily the most important—mechanism for increasing cell number. In normal adult tissues, especially those at risk of cancer, cells may proliferate continually; but their numbers remain steady because cell production is balanced by cell loss, as part of the body's homeostatic control mechanisms. Programmed cell death by *apoptosis* usually plays an essential part in this balance, as we discuss in Chapters 18 and 23. If too many

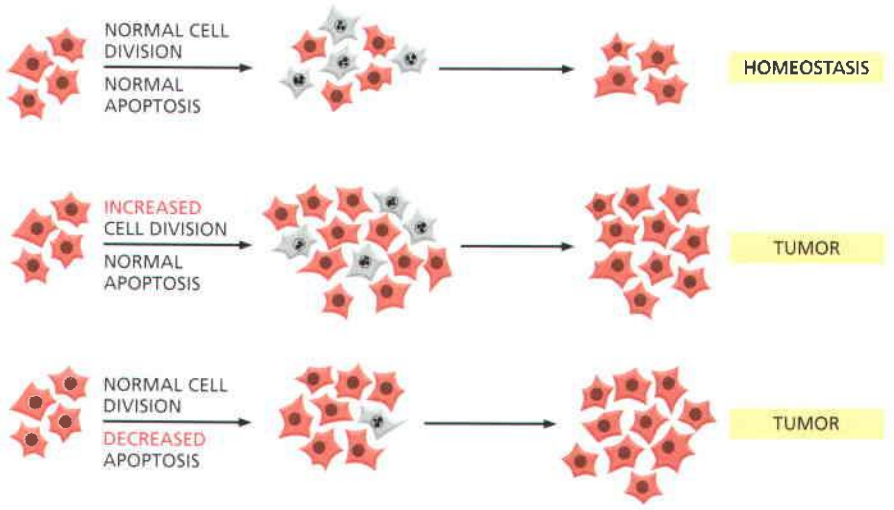


Figure 20–14 Both increased cell division and decreased apoptosis can contribute to tumorigenesis. In normal tissues, apoptosis balances cell division to maintain homeostasis. <GCCC> During the development of cancer, either an increase in cell division or an inhibition of apoptosis can lead to the increased cell numbers important for tumorigenesis. Cells fated to undergo apoptosis are shown in *gray*.

cells are generated, the rate of apoptosis increases to dispose of the surplus. One of the most important properties of many types of cancer cells is that they fail to undergo apoptosis when a normal cell would do so. This greatly contributes to the growth of a tumor (**Figure 20–14**).

Inherited changes can also increase the size of a clone of mutant cells by altering their ability to differentiate, as illustrated by the situation in the uterine cervix, discussed earlier. When a stem cell in the basal layer divides, each daughter cell has a choice—it can either remain a stem cell or commit to a pathway leading to differentiation; the committed cells initially proliferate and then stop dividing and terminally differentiate (the committed, proliferating cells are called *transit amplifying cells*). If the differentiation program is blocked at some stage, proliferating cells accumulate, contributing to the progression from low-grade intraepithelial neoplasia of the cervix to high-grade intraepithelial neoplasia and malignant cancer (see Figure 20–9). Similar considerations apply to the development of cancer in other tissues that rely on stem cells, such as the skin, the lining of the gut, and the hemopoietic system. Several forms of leukemia, for example, seem to arise from a disruption of the normal program of differentiation, such that a committed progenitor of a particular type of blood cell eventually becomes able to divide indefinitely, instead of undergoing terminal differentiation in the normal way and dying after a limited number of cell divisions (as discussed in Chapter 23).

Thus, the accumulation of mutations and epigenetic changes that lead to defects in the normal controls of cell division, apoptosis, and differentiation can all contribute to the development and progression of cancers.

Cancer Cells Are Usually Altered in Their Responses to DNA Damage and Other Forms of Stress

As just discussed, many normal cells permanently stop dividing when they differentiate into specialized cells. Differentiation, however, is not the only reason that proliferating cells stop dividing; they can also do so in response to stress or to damage to their DNA. As described in Chapter 17, normal cells contain a set of *checkpoint control* mechanisms. These arrest the cell cycle temporarily when something goes awry, providing time to correct the problem. Chromosome breakage and other types of damage to DNA generate intracellular signals that activate these checkpoint mechanisms, causing a normal cell to halt its cycle, and thereby giving it a chance to repair the damage before it attempts to progress through the cycle and divide. And if the damage is irreparable, a normal cell, rather than forge ahead and generate daughters with a damaged genome, will either permanently withdraw from the cell cycle or commit suicide by apoptosis.

Cancer cells usually acquire mutations and epigenetic changes that inactivate such responses to DNA damage. From the selfish standpoint of the cancer

cells, the loss of these restraints confers an advantage, allowing them to continue to multiply even with damaged DNA. At the same time, this accelerates the rate at which mutations accumulate in the tumor cell population and can thereby hasten the progression to greater malignancy. The cancer cells are also often defective in other checkpoint control mechanisms that help regulate the cell cycle. Thus, they will often stagger on through the cycle when things go wrong, and in this way will do themselves still further genetic damage, contributing to the chromosome abnormalities that they often display. Even though such genetically damaged cells usually take longer to complete each cell cycle than normal cycling cells, they accumulate over time because they display a greater propensity to proliferate. It is perhaps remarkable that these cells can manage to outcompete their normal neighbors when they are so messed up.

Human Cancer Cells Escape a Built-in Limit to Cell Proliferation

Many normal human cells have a built-in limit to the number of times they can divide when stimulated to proliferate in culture: they permanently stop dividing after a certain number of population doublings (25–50 for human fibroblasts, for example), a phenomenon termed **replicative cell senescence** (discussed in Chapter 17). This cell-division-counting mechanism generally depends on the progressive shortening of the telomeres at the ends of chromosomes, which eventually changes their structure. As discussed in Chapter 5, the replication of telomere DNA during S phase depends on the enzyme **telomerase**, which maintains the special telomeric DNA sequence and promotes the formation of protein cap structures that protect the chromosome ends. Because many proliferating human cells (but not stem cells) are deficient in telomerase, their telomeres shorten with every division, and their protective caps deteriorate. Eventually, the altered chromosome ends trigger a permanent cell-cycle arrest.

The cell-division-counting mechanism just described sets a physiological limit to cell proliferation for most of the types of cells in the body. Human cancer cells must somehow avoid or overcome this barrier to form large tumors. Some rodent cells, by contrast, maintain telomerase activity and normal telomeres as they proliferate and therefore lack this barrier. It has been proposed that humans need replicative cell senescence to help prevent cancer because our comparatively long lifespan would otherwise provide enormous opportunities for tumor progression.

Human cancer cells avoid replicative cell senescence in two ways. First, they acquire genetic and epigenetic changes that disable the checkpoint control so that the cells continue to cycle even when telomeres become uncapped. Mutations that inactivate the p53 pathway have this effect and are very common in cancer cells, as we shall discuss below. As another strategy to escape replicative senescence, cancer cells often maintain telomerase activity as they proliferate, so that their telomeres do not shorten or become uncapped. In some cases, the cancer cells may maintain telomerase activity because the cancer originated in stem cells that have this activity. In other cases, although the cancer originated in cells without appreciable telomerase activity, the cancer cells have acquired the activity as a result of genetic or epigenetic changes that were selected for as their telomeres shortened. Yet other cancer cells evolve alternate mechanisms for elongating chromosome ends. Regardless of the strategy used, the result is that the cancer cells continue to proliferate under conditions when normal cells would stop.

A Small Population of Cancer Stem Cells Maintains Many Tumors

As we discuss in Chapter 23, in a normal adult cell lineage with substantial cell turnover, such as the hemopoietic cell lineage, the cells are typically organized as a hierarchy: rare, slowly dividing stem cells produce cells (the transit amplifying cells) that proliferate extensively and eventually differentiate into the specialized cells characteristic of the lineage. The stem cells maintain the lineage by

producing some daughter cells that can remain stem cells (a process called *self-renewal*), while producing others that commit to a pathway leading to differentiation. There is increasing evidence that many cancers are organized in the same hierarchical way, with rare **cancer stem cells** at the top of the hierarchy responsible for maintaining the population of cells in a tumor. These stem cells are capable of indefinite self-renewal, but they also give rise to rapidly dividing transit amplifying cells that have a limited capacity for self-renewal. It has only recently become widely recognized that these cells with a limited growth potential make up the vast majority of the cells in many cancers.

It has been known for over 40 years that there is only a small chance—usually much less than 1%—that a randomly chosen individual cell from a cancer will be able to generate a new tumor when tested for its ability to do so—for example, by implanting it into an immunodeficient mouse. New technologies for sorting cells have shown that, in many cases, this is because the cancer cells are heterogeneous, and only a small subset of them have the special property needed for tumor propagation. One way to enrich the small population of presumptive cancer stem cells in some cancers uses a cell-permeable fluorescent dye and a fluorescence-activated cell sorter (see Figure 8–2). When a cell suspension prepared from a normal tissue is exposed to the dye and then sorted in this way, the stem cells of the tissue are often found to be highly enriched in a small *side population* of weakly fluorescent cells. The stem cells are only weakly fluorescent because they have an ABC transporter (discussed in Chapter 11) in their plasma membrane that actively pumps the dye out of the cell.

When some cancers are analyzed in the same way, a small side population with an ABC transporter is found to contain most, if not all, of the cells capable of forming new cancers after transplantation into immunodeficient mice. Remarkably, this is also the case for some cancer cell lines that have been propagated for more than thirty years in culture: whereas the side population of each of these lines readily forms tumors after transplantation, the remaining highly fluorescent population, which includes the vast majority of the cells in the line, does not (**Figure 20–15**). Moreover, when the side population isolated from such a cancer cell line is returned to culture, a small side population rapidly reappears and is maintained indefinitely, whereas a side population does not develop in cultures of the remaining cells.

In other studies, cell fractionations using monoclonal antibodies that recognize normal stem cells in the tissue of origin of the cancer have given similar results.

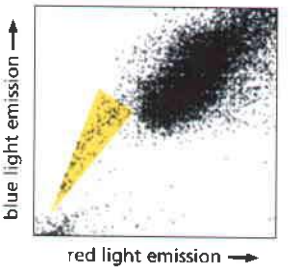


Figure 20–15 Cell sorting reveals cancer stem cells in a small side population of cells. In this example, the cells in an established cell line obtained from a rat glioma tumor were treated with a dye that can pass through the plasma membrane (Hoechst 33342) and then analyzed by cell sorting. About 0.4% of the cells pump out the dye and are therefore only lightly stained, as indicated in *yellow*. Experiments comparing these side population (SP) cells with the remaining cells showed that only the side population cells could proliferate extensively and form tumors. Moreover, whereas the SP cells produced both SP and non-SP cells, the non-SP cells could not produce SP cells. (Adapted from T. Kondo, T. Setoguchi and T. Taga, *Proc. Natl Acad. Sci. U.S.A.* 101:781–786, 2004. With permission from National Academy of Sciences.)

How Do Cancer Stem Cells Arise?

The finding that many cancers are maintained by a small population of cancer stem cells has extremely important implications, both for understanding cancer and for treating it, and it raises a number of intriguing questions. How do these stem cells arise? What is the nature of the cancer stem cell, and what distinguishes it in molecular terms from the majority of cells in the cancer? From the results just described with cultured cancer cell lines, many of these differences must result from epigenetic rather than from genetic changes.

It is clear for some tumors that the cancer originates in a normal tissue stem cell, which gradually accumulates the mutations and epigenetic changes responsible for the asocial behavior of the cancer. For example, this sequence of events is almost certainly the way chronic myelogenous leukemias arise. Thus, when the multipotential hemopoietic stem cells (see p. 1456) are purified from the bone marrow of these patients, many contain the Philadelphia chromosome, with the translocation characteristic of this cancer. Likewise, many human cancers probably arise from normal stem cells, because they develop in epithelia that undergo substantial cell turnover. In these tissues, only stem cells remain in the body and proliferate for long enough to accumulate the number of mutations required to develop into a cancer.

The second way that a cancer stem cell can arise is through a change in a proliferating cell that is more differentiated, such as a transit amplifying cell. The

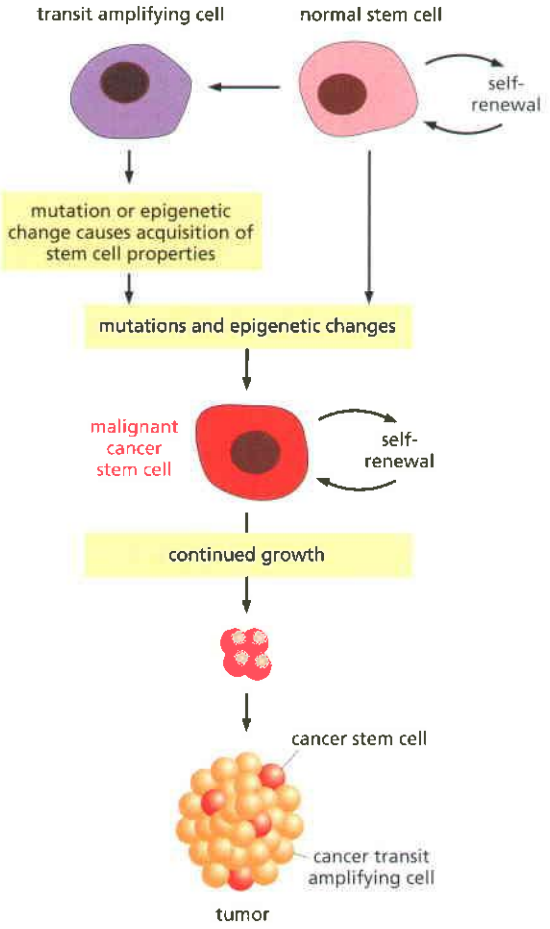


Figure 20–16 Cancers may arise from cancer stem cells. Cancer stem cells are defined as those cancer cells that can self-renew to produce additional malignant stem cells, and at the same time generate non-tumorigenic cells such as transit amplifying cells. Cancer stem cells can arise from normal stem cells that have sustained mutations to make them cancerous; or from more differentiated cells that have undergone mutations or epigenetic changes that give them stem cell properties.

changes must confer the two crucial stem-cell properties: the ability to be retained in the body, and the capacity for prolonged self-renewal. Then, and only then, the altered cell will have the time to accumulate the other mutations and epigenetic changes needed to become a full-blown cancer cell. Some B lymphocyte leukemias, for example, seem to arise in this way. All of the cells in such a leukemia contain the same genetic rearrangements in their antibody genes (discussed in Chapter 25), suggesting that the cancer originated in a committed B lymphocyte precursor cell, rather than in an uncommitted stem cell. Thus, it seems likely that, while most human cancers arise from a stem cell, others arise from a cell that has acquired the capacity for sustained self-renewal through mutation, epigenetic change, or both (**Figure 20–16**).

Most current cancer therapies such as radiation and cytotoxic drugs are likely to preferentially kill the most rapidly proliferating cancer cells in a tumor. Because the cancer stem cells usually divide much more slowly, they may be less sensitive to these treatments. Moreover, in the 1970s, it was proposed that evolution would have selected for mechanisms that protect stem cells from accumulating mutations. In particular, since many mutations arise during DNA replication, could each division of a stem cell utilize a special mechanism that allows the old DNA template strand in each chromosome to be segregated into the one daughter cell that will remain a stem cell? Recent experiments suggest that stem cells, and only stem cells, can indeed coordinate the segregation of their chromosomes in this way (see Figure 23–16). If this is also the case for cancer stem cells, it could make them unusually resistant to many treatments that kill other cells.

If the stem cells are not eradicated, they will very likely regenerate the cancer. This may be a major reason why cancers commonly recur after a dramatic initial response to therapy. To cure cancers, it is clearly crucial to find better ways to target the cancer stem cells and to kill them.

To Metastasize, Malignant Cancer Cells Must Survive and Proliferate in a Foreign Environment

Metastasis is the most deadly—and least understood—aspect of cancer, being responsible for 90% of cancer-associated deaths. By spreading throughout the body, a cancer becomes almost impossible to eradicate by either surgery or localized irradiation. Metastasis is itself a multistep process: the cancer cells have to invade local tissues and vessels, move through the circulation, leave the vessels, and then establish new cellular colonies at distant sites. Each of these events is, in itself, complex, and most of the molecular mechanisms involved are not yet clear.

For a cancer cell to metastasize, it must break free of constraints that keep normal cells in their proper places and prevent them from invading neighboring tissues. Invasiveness is thus one of the defining properties of malignant tumors, which show a disorganized pattern of growth and ragged borders, with extensions into the surrounding tissue (see, for example, Figure 20–9). Although the underlying molecular changes are not well understood, invasiveness almost certainly requires a disruption of the adhesive mechanisms that normally keep cells tethered to their proper neighbors and to the extracellular matrix. As we discuss later, for carcinomas, this change resembles the *epithelial-to-mesenchymal transition (EMT)*, which occurs in some epithelial tissues during normal development (see Figure 19–12).

The next step in metastasis—the establishment of colonies in distant organs—is a complex, slow, and inefficient operation; few cells achieve it. Before it can metastasize successfully, a cell must penetrate a blood vessel or a lymphatic vessel by crossing the basal lamina and the endothelial lining of the vessel so as to enter the blood or lymph, exit from a vessel elsewhere in the body, and grow in the new site, forming at first a small clump of cells known as a micrometastasis. To complete the metastatic process, some of these micrometastases must produce cells that survive and proliferate extensively in the new environment, a difficult process known as colonization (**Figure 20–17**).

The vasculature that grows into tumors is leaky, and it seems that many of the cells in a tumor manage to escape into lymphatic or blood vessels. However, experiments show that only a tiny proportion of these, fewer than one in thousands, perhaps one in millions, survive to form anything more than micrometastases. Many cancers are discovered before they have managed to found metastatic colonies and can be cured if the primary tumor is removed.

While it is possible in principle that each cancer cell entering the bloodstream might have the same tiny chance of surviving, settling, and growing in an alien site, this seems unlikely. More likely, the low rate of colonization reflects the rarity of cells that have the special properties needed. That is, it may be that only a very small population of the cells in a typical cancer have both the stem-cell character that will enable them to divide without limit, and the ability to settle and survive in an alien environment. The rarity of cancer stem cells would explain why so few cells acheive colonization, even after forming a micrometastasis.

The need for special survival capabilities is likely to be an additional restraint. Normal cells depend on extracellular survival signals that are relatively abundant in their normal environment; when deprived of these signals, they activate their cell death machinery and undergo apoptosis (discussed in Chapter 18). Cancer cells capable of metastasis are often resistant to apoptosis compared with normal cells and therefore can more readily survive after escaping from their proper home. They are also less dependent on signals from other cells to grow and divide.

Tumors Induce Angiogenesis

In addition to all the requirements just described, to grow large, a tumor must recruit an adequate blood supply to ensure that it gets sufficient oxygen and nutrients. Thus, *angiogenesis*, the formation of new blood vessels, is required for tumor growth beyond a certain size. Like normal tissues, tumors attract a blood

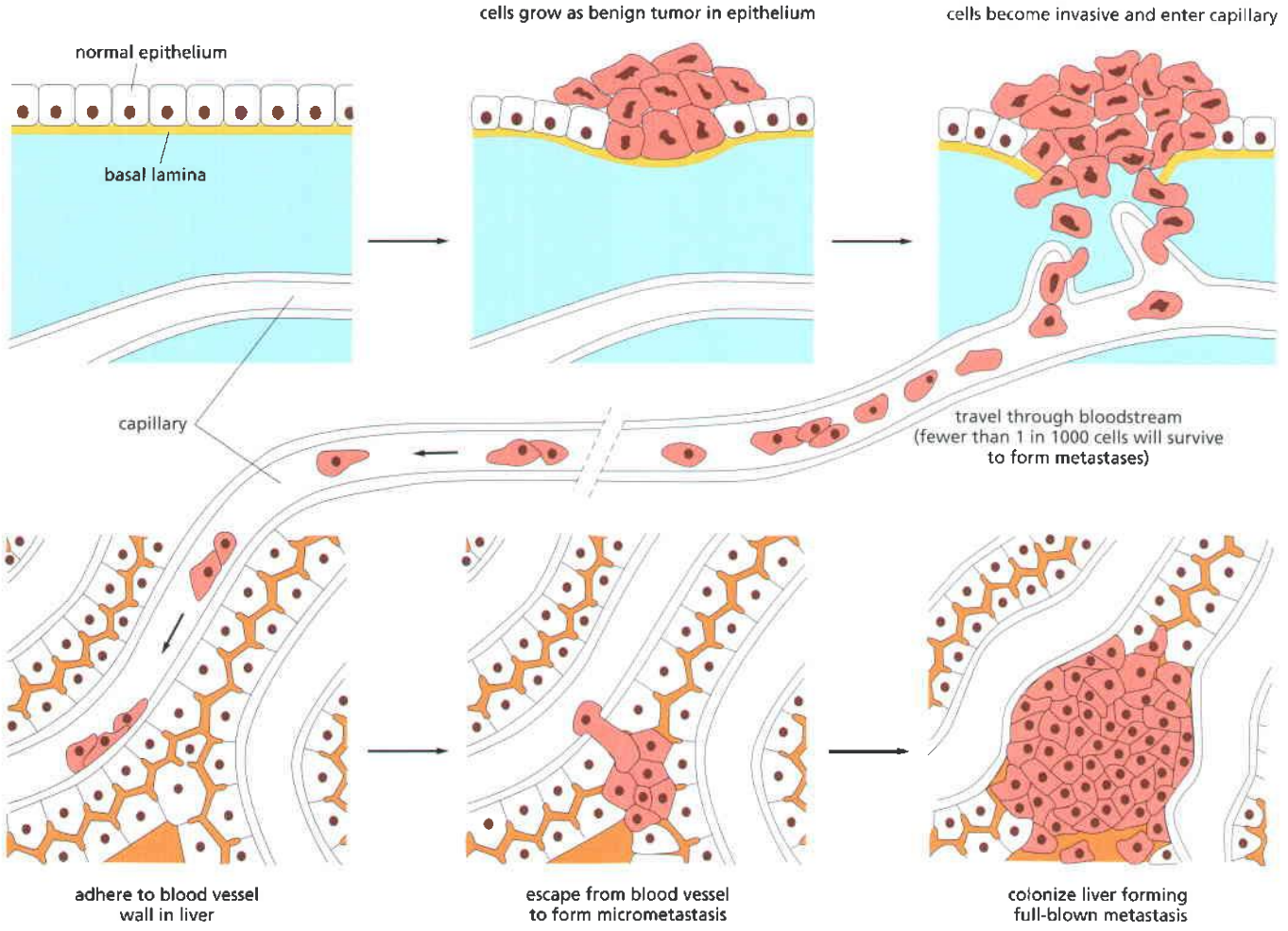


Figure 20–17 Steps in the process of metastasis. This example illustrates the spread of a tumor from an organ such as the bladder to the liver. Tumor cells may enter the bloodstream directly by crossing the wall of a blood vessel, as diagrammed here, or, more commonly perhaps, by crossing the wall of a lymphatic vessel that ultimately discharges its contents (lymph) into the bloodstream. Tumor cells that have entered a lymphatic vessel often become trapped in lymph nodes along the way, giving rise to lymph-node metastases. Studies in animals show that typically far fewer than one in every thousand malignant tumor cells that enter the bloodstream will colonize a new tissue so as to produce a detectable tumor at a new site.

supply by secreting angiogenic signals. These signals are produced in response to hypoxia, which begins to affect the cells as the tumor enlarges beyond a millimeter or two in diameter. The hypoxia activates an *angiogenic switch* to increase the blood supply by provoking an increase in the level of *Hypoxia Inducible Factor-1α (HIF-1α)*, a gene regulatory protein described in Chapter 23; this protein, in turn, activates the transcription of genes that encode pro-angiogenic factors, such as *vascular endothelial growth factor (VEGF)*. These are secreted proteins that attract endothelial cells and stimulate the growth of new blood vessels (see Figure 23–34). The vessels not only help supply the tumor with nutrients and oxygen, but also provide an escape route for its cells to metastasize.

However, the induced vessels are tortuous, heterogeneous in diameter, and leaky, and they have many branches with dead ends. These abnormalities, which probably result from an abnormal balance of signaling molecules, lead to irregular blood flow within the tumor, helping to create additional regions of hypoxia (**Figure 20–18**). Hypoxia, in turn, selects for mutant cancer cells that are better able to survive in these harsh and stressful conditions, which usually means that they are more malignant. Ultimately, however, tumor growth relies on an adequate blood supply, and the abnormal vessels that are attracted into tumors are one obvious target for drug therapy, as we discuss later.

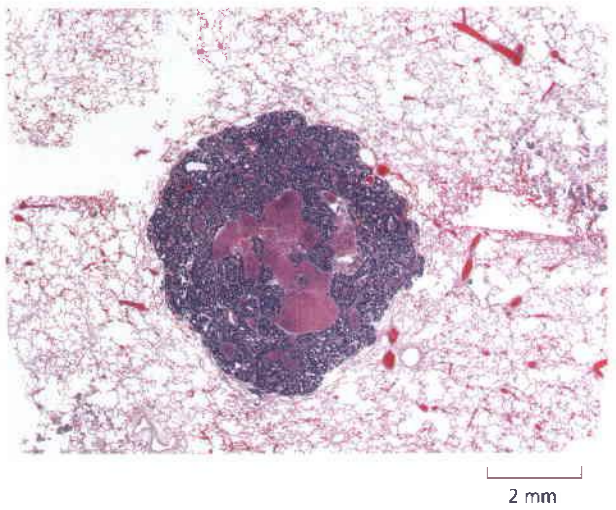


Figure 20–18 Colon adenocarcinoma metastasis in the lung. This tissue slice shows well-differentiated colorectal cancer cells, still forming cohesive glands in the lung. This metastasis has central pink areas of necrosis where the cancer cells have outgrown their blood supply. (Courtesy of Andrew J. Connolly.)

The Tumor Microenvironment Influences Cancer Development

As mentioned earlier, carcinomas are complex entities, containing many cell types in addition to the cancer cells. While the cancer cells are grossly abnormal, being the bearers of mutations in these tumors, the other cells in the tumor—especially those of the supporting connective tissue, or **stroma**—are far from innocent bystanders. This is because the development of a tumor relies on cross-talk between the tumor cells and the tumor stroma, just as the normal development of epithelial organs relies on cross-talk between epithelial cells and mesenchymal cells (discussed in Chapter 22).

The stroma provides the framework for the tumor and is composed of normal connective tissue containing fibroblasts, myofibroblasts, inflammatory white blood cells, the endothelial cells of blood and lymphatic vessels, and their attendant pericytes and smooth muscle cells (**Figure 20–19**). As a carcinoma progresses, the cancer cells induce changes in the stroma by secreting both signal proteins that alter the behavior of the stromal cells and proteolytic enzymes that modify the extracellular matrix. The stromal cells in turn act back on the tumor cells in various ways. They secrete signal proteins that stimulate cancer cell growth and division, and they secrete proteases that further remodel the extracellular matrix. In these ways, the tumor and its stroma evolve together, and the tumor becomes dependent on its particular stromal cells. Experiments in mice indicate that the growth of some transplanted carcinomas depends on the tumor-associated fibroblasts, and normal fibroblasts will not do. As a second example, the growth of skin tumors is significantly retarded in mice lacking a specific matrix metalloproteinase that is normally secreted by stromal mast cells; the protease is responsible for releasing angiogenic factors from the extracellular matrix that stimulate the angiogenesis required for the skin tumor to grow. These and other results suggest that cancer treatments could be designed to inhibit activated stromal cells, in addition to targeting the tumor cells themselves.

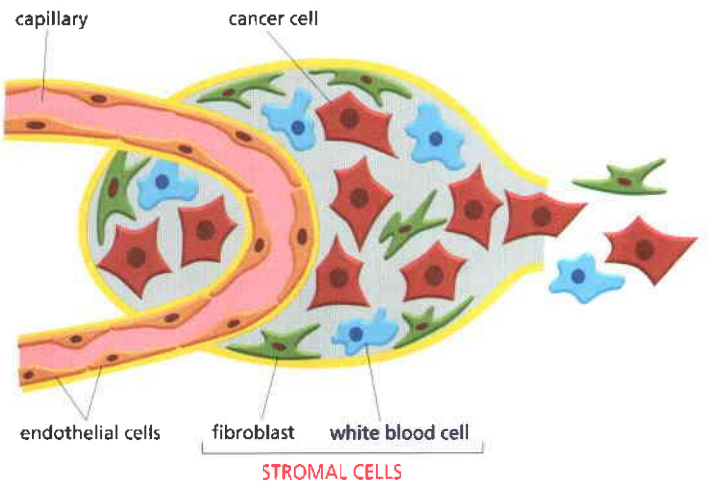


Figure 20–19 The tumor microenvironment plays a role in tumorigenesis. Tumors consist of many cell types including cancer cells, vascular epithelial cells, fibroblasts, and inflammatory white blood cells. Cross-talk between these and other cell types plays an important part in tumor development.

Many Properties Typically Contribute to Cancerous Growth

Clearly, to produce a cancer, a cell must acquire a range of aberrant properties—a collection of subversive new skills—as it evolves. Different cancers require different combinations of these properties. Nevertheless, we can draw up a short list of the key behaviors of cancer cells in general:

1. They are more self-sufficient than normal cells for their growth and proliferation. For example, unlike most normal cells, they can often survive and proliferate in cell culture even when not adherent to a substratum and floating free in suspension.
2. They are relatively insensitive to anti-proliferative extracellular signals.
3. They are less prone than normal cells to undergo apoptosis.
4. They are defective in the intracellular control mechanisms that normally stop cell division permanently in response to stress (such as hypoxia) or DNA damage.
5. They induce help from the normal stromal cells in their local environment.
6. They induce angiogenesis.
7. They escape from their home tissues (that is, they are invasive) and survive and proliferate in foreign sites (that is, they metastasize).
8. They are genetically unstable.
9. They either produce telomerase, or acquire another way of stabilizing their telomeres.

In the later sections of the chapter, we examine the mutations and molecular mechanisms that underlie some of these properties, as well as the environmental factors that promote the development of cancer.

Summary

Cancer cells, by definition, grow and proliferate in defiance of normal controls (that is, they are neoplastic) and are able to invade surrounding tissues and colonize distant organs (that is, they are malignant). By giving rise to secondary tumors, or metastases, they become difficult to eradicate by surgery or local irradiation. Cancers are thought to originate from a single cell that has experienced an initial mutation, but the progeny of this cell must undergo many further changes, requiring numerous additional mutations and epigenetic events, to become cancerous. The cell of origin of a cancer may be either a tissue stem cell, which is already endowed with the ability to self-renew indefinitely, or a more differentiated cell that acquires the capacity for indefinite self-renewal. Tumor progression usually takes many years and reflects the operation of a Darwinian-like process of evolution, in which somatic cells undergo mutation and epigenetic changes accompanied by natural selection.

Cancer cells acquire a variety of special properties as they evolve, multiply, and spread. These include alterations in cell signaling pathways, enabling the cells to ignore the signals from their environment that normally keep cell proliferation under tight control. As part of the evolutionary process of tumor progression, cancer cells acquire defects in differentiation and in the control mechanisms that permanently stop cell division or induce apoptosis in response to cell stress or DNA damage. All of these changes increase the ability of cancer cells to survive, grow, and divide in their original tissue and then to metastasize—which requires survival and proliferation in foreign environments. However, the evolution of a tumor does not simply rely on changes in the cancer cells themselves; it also depends on other cells present in the tumor microenvironment, collectively called stromal cells, including new blood vessels that allow the tumor to grow large and metastasize via the bloodstream.

Since many mutations and epigenetic changes are needed to confer this collection of asocial behaviors, it is not surprising that nearly all cancer cells are genetically unstable. The genetic instability can arise from defects in the ability to repair DNA damage or to correct replication errors of various kinds, which lead to changes in DNA sequence. Also common are defects in chromosome segregation during mitosis, which lead to chromosome instability and changes in karyotype. This genetic instability is selected for in the clones of aberrant cells that are able to produce tumors, because it greatly accelerates the accumulation of the further genetic and epigenetic changes that are required for tumor progression.

THE PREVENTABLE CAUSES OF CANCER

The development of a cancer generally requires many steps, each governed by multiple factors—some dependent on the genetic constitution of the individual, others dependent on his or her environment and way of life. A certain irreducible background incidence of cancer is to be expected regardless of circumstances: mutations can never be absolutely avoided because they are an inescapable consequence of fundamental limitations on the accuracy of DNA replication and repair, as discussed in Chapter 5. If a human could live long enough, it is inevitable that at least one of his or her cells would eventually accumulate a set of mutations sufficient for cancer to develop.

Nevertheless, there is evidence that environmental factors, including the food we eat, accelerate the onset of most cancers. This is demonstrated most clearly by a comparison of cancer incidence in different countries: for almost every cancer that is common in one country, there is another country where the incidence is much lower. Because migrant populations tend to adopt the pattern of cancer incidence typical of the host country, the differences appear to be due mostly to environmental, not genetic, factors. From such findings, it is estimated that 80–90% of cancers should be avoidable, or at least postponable (**Figure 20–20**).

Unfortunately, different cancers have different environmental risk factors, and a population that escapes one such danger is usually exposed to another. This is not, however, inevitable. There are some human subgroups whose way of life substantially reduces the total cancer death rate among individuals of a given age. Under the current conditions in the United States and Europe,



(A)

environmental and lifestyle factors	cancer	% total cases
• occupational exposure	various types	1–2
• tobacco related	lung, kidney, bladder	24
• diet: low in vegetables, high salt, high nitrate	stomach, esophagus	5
• diet: high fat, low fiber, fried and broiled foods	bowel, pancreas, prostate, breast	37
• tobacco and alcohol	mouth, throat	2

(B)

Figure 20–20 Cancer incidence is related to environmental influences. (A) This map of the world shows the rates of cancer increasing (*red arrows*) or decreasing (*blue arrows*) when specific populations move from one location to another. These observations suggest the importance of environmental factors, including diet, in dictating cancer risk. (B) Some estimated effects of environment and lifestyle on cancer in the United States. The table shows the percentage of total cancer cases attributable to each specified factor. (B, adapted from Cancer Facts and Figures, American Cancer Society, 1990. With permission from American Cancer Society.)

approximately one in five people will die of cancer. But the incidence of cancer among strict Mormons in Utah (but, importantly, not among non-practicing members of the same family), who avoid alcohol, coffee, cigarettes, drugs, and casual sex, is only about half the incidence among Americans in general. Cancer incidence is also low in certain relatively affluent populations in Africa.

Although such observations on human populations indicate that cancer can often be avoided, it has been difficult in most cases—with the exception of tobacco—to identify the specific environmental risk factors or to establish how they act. We will first look at what has been learned about the external agents that are known to cause cancer. We will then consider some of the triumphs, and difficulties, in finding ways to prevent human cancer. We will discuss the treatment of cancer in the last section, after we have examined the molecular biology of the disease.

Many, But Not All, Cancer-Causing Agents Damage DNA

The agents that can cause cancer, known as **carcinogens**, are many and varied, but the easiest to understand are those that damage DNA and so generate mutations. These cancer-causing mutagens include chemicals and various forms of radiation—UV light from sunshine, as well as ionizing radiation, such as γ rays and α particles from radioactive decay.

Many quite disparate chemicals are carcinogenic when they are fed to experimental animals or painted repeatedly on their skin. Examples include a range of aromatic hydrocarbons and derivatives of them such as aromatic amines, nitrosamines, and alkylating agents such as mustard gas. Although these chemical carcinogens have diverse structures, they have at least one common property—they cause mutations. In one common test for mutagenicity, the carcinogen is mixed with an activating extract prepared from rat liver cells (to mimic the biochemical processing that occurs in an intact animal—discussed below) and is added to a culture of specially designed test bacteria; the resulting mutation rate of the bacteria is then measured (**Figure 20–21**). Most of the compounds scored as mutagenic by this rapid and convenient assay in bacteria also cause mutations or chromosome aberrations when tested on mammalian cells. When mutagenicity data from various sources are analyzed, one finds that many carcinogens are mutagens.

A few of these carcinogens act directly on DNA, but generally the more potent ones are relatively inert chemically; they become damaging only after they have been changed to a more reactive form by metabolic processes in the liver, catalyzed by a set of intracellular enzymes known as the *cytochrome P-450 oxidases*. These enzymes normally help to convert ingested toxins into harmless and easily excreted compounds. Unfortunately, their activity on certain chemicals generates products that are highly mutagenic. Examples of carcinogens activated in this way include the fungal toxin *aflatoxin B1* and *benzo[a]pyrene*, a cancer-causing chemical present in coal tar and tobacco smoke (**Figure 20–22**).

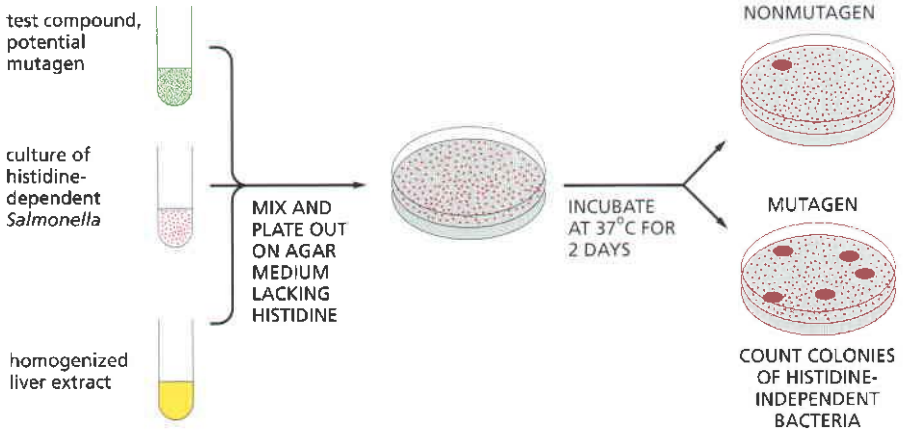
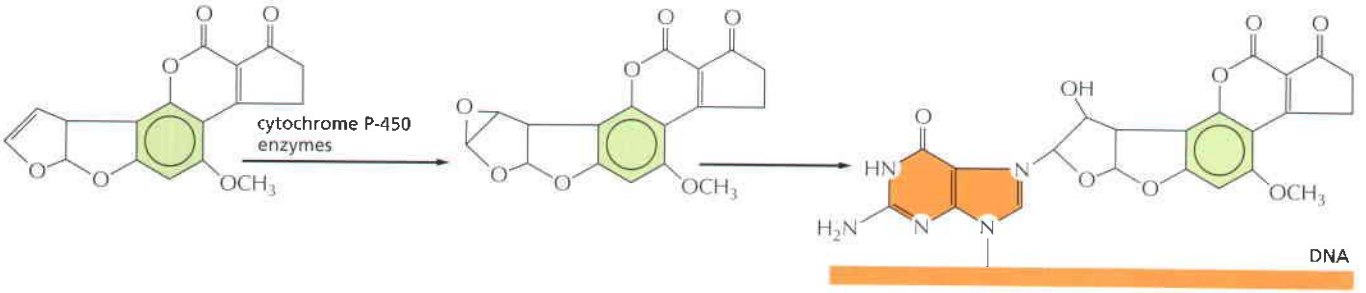


Figure 20–21 The Ames test for mutagenicity. The test uses a strain of *Salmonella* bacteria that require histidine in the medium because of a defect in a gene necessary for histidine synthesis. Mutagens can cause a further change in this gene that reverses the defect, creating revertant bacteria that do not require histidine. To increase the sensitivity of the test, the bacteria also have a defect in their DNA repair machinery that makes them especially susceptible to agents that damage DNA. A majority of compounds that are mutagenic in tests such as this are also carcinogenic.



(A) AFLATOXIN AFLATOXIN-2,3-EPOXIDE CARCINOGEN BOUND TO GUANINE IN DNA

Figure 20–22 Some known carcinogens. (A) Carcinogen activation. A metabolic transformation must activate many chemical carcinogens before they will cause mutations by reacting with DNA. The compound illustrated here is *aflatoxin B1*, a toxin from a mold *(Aspergillus flavus oryzae)* that grows on grain and peanuts when they are stored under humid tropical conditions. It is thought to be a contributory cause of liver cancer in the tropics and is associated with characteristic mutations of the *p53* tumor suppressor gene. (B) Different carcinogens cause different types of cancer. (B, data from Cancer and the Environment: Gene Environment Interactions, National Academies Press, 2002. With permission from National Academies Press.)

- VINYL CHLORIDE: liver angiosarcoma
- BENZENE: acute leukemias
- ARSENIC: skin carcinomas, bladder cancer
- ASBESTOS: mesothelioma
- RADIUM: osteosarcoma

(B)

Tumor Initiators Damage DNA; Tumor Promoters Do Not

Not all substances that favor the development of cancer are mutagens. Some of the clearest evidence comes from studies done long ago on the effects of cancer-causing chemicals on mouse skin, where it is easy to observe the stages of tumor progression. Skin cancers can be elicited in mice by repeatedly painting the skin with a DNA-damaging carcinogen such as benzo[a]pyrene or the related compound dimethylbenz[a]anthracene (DMBA). A single application of these mutagens, however, does not by itself give rise to a tumor or any other obvious lasting abnormality. Yet, it does cause latent genetic damage—mutations that set the stage for a greatly increased incidence of cancer when the cells are exposed either to further treatments with the carcinogen or to certain other, quite different, insults. A carcinogen that sows the seeds of cancer in this way acts as a **tumor initiator**.

Simply wounding skin that has been exposed once to such an initiator can, by stimulating cell proliferation, cause cancers to develop from some of the cells at the edge of the wound. Inflammatory white blood cells are attracted to a wound, and it is thought that inflammatory cells play a part in the cancer-promoting effect, as we mentioned earlier. Alternatively, repeated exposure over a period of months to substances known as **tumor promoters**, chemicals that are not themselves mutagenic, can cause cancer selectively in skin previously exposed to a tumor initiator. The most widely studied tumor promoters are *phorbol esters*, such as tetradecanoylphorbol acetate (TPA), which behave as artificial activators of *protein kinase C (PKC)* and hence activate part of the phosphatidylinositol intracellular signaling pathway (discussed in Chapter 15). These substances cause cancers at high frequency only if they are applied after a treatment with a mutagenic initiator (**Figure 20–23**).

The immediate effect of a tumor promoter is apparently to induce an inflammatory response, causing the secretion of both growth factors and proteases in the local environment. These act both directly and indirectly on cells to stimulate cell division (or to cause cells that would normally differentiate and stop dividing to continue dividing instead). In a region of skin that has previously been exposed to an initiator, this proliferation results in the development of many small, benign, wartlike tumors called *papillomas*. The greater the prior dose of initiator, the larger the number of papillomas induced; it is thought that each papilloma consists of a single clone of cells descended from a mutant cell that the initiator has produced. As one might expect for genetic damage, the hidden changes

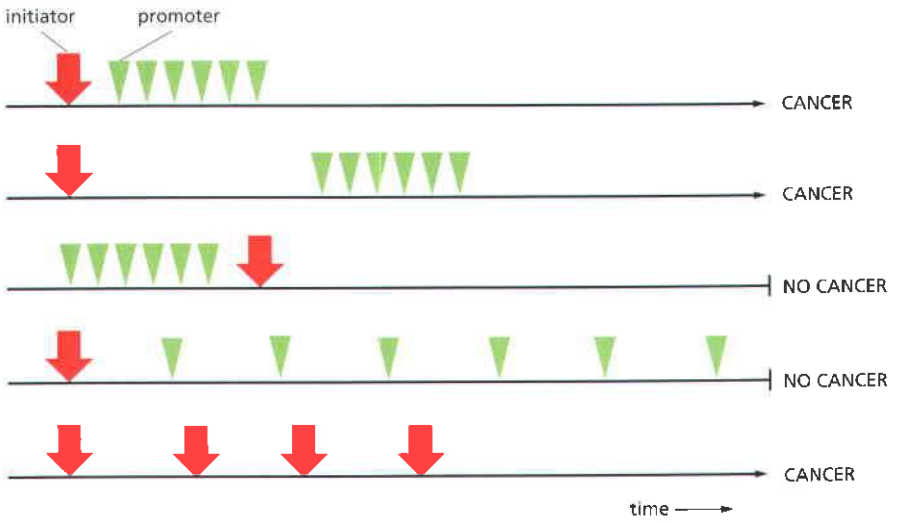


Figure 20–23 Some possible schedules of exposure to a tumor initiator (mutagenic) and a tumor promoter (nonmutagenic) and their outcomes. Cancer ensues only if the exposure to the promoter follows exposure to the initiator and only if the intensity of exposure to the promoter exceeds a certain threshold. Cancer can also occur as a result of repeated exposure to the initiator alone.

caused by a tumor initiator are irreversible; for this reason, they can be uncovered by treatment with a tumor promoter even after a long delay.

It is still not certain how tumor promoters work, and different promoters are likely to work in different ways. One possibility is that they simply provoke the expression of proliferation-inducing genes that had been mutated before the promoter was applied but were not expressed: a mutation that makes a gene product hyperactive will not show its effects until the gene is expressed. Alternatively, the tumor promoter may temporarily alter the way the cell reacts to the product of the mutated gene, either by releasing the cell from a counteracting inhibitory influence or by triggering production of a co-factor necessary for proliferation of the mutated gene product. Whichever the mechanism, the result is that the mutant cell is enabled to grow and divide and produce a large cluster of cells (**Figure 20–24**).

A typical papilloma might contain about 10^5 cells. If exposure to the tumor promoter is stopped, almost all the papillomas regress, and the skin regains a largely normal appearance. In a few of the papillomas, however, further changes occur that enable cell growth and division to continue in an uncontrolled way, even after the promoter has been withdrawn. These changes seem to originate in an occasional single papilloma cell, at about the frequency expected for spontaneous mutations. In this way, a small proportion of the papillomas progress to become cancers. Thus, the tumor promoter apparently favors the development of cancer by expanding the population of cells that carry an initial mutation: the more such cells there are and the more times they divide, the greater is the chance that at least one of them will sustain another mutation, or an epigenetic change, that carries it one step further toward malignancy.

Although naturally occurring cancers do not necessarily arise through the specific sequence of distinct initiation and promotion steps just described, their evolution must be governed by similar principles. They too will evolve at a rate that depends not only on the frequency of genetic or epigenetic changes, but also on the local influences affecting the survival, growth, proliferation, and spread of these altered cells.

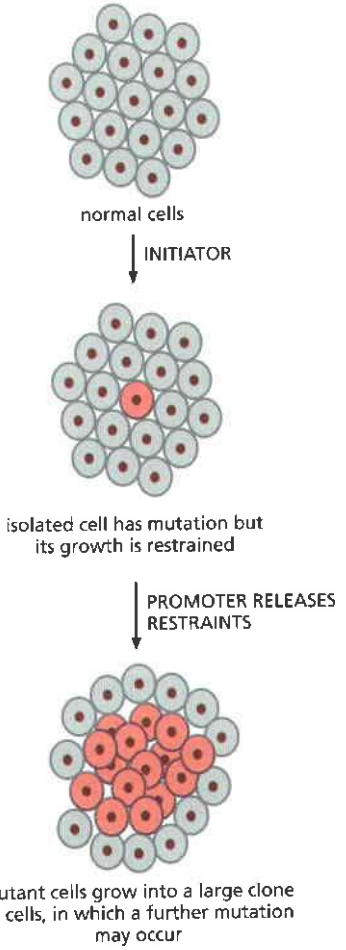


Figure 20–24 The effect of a tumor promoter. The tumor promoter creates a local environment that expands the population of mutant cells, thereby increasing the probability of tumor progression by further genetic change.

Viruses and Other Infections Contribute to a Significant Proportion of Human Cancers

A small but significant proportion of human cancers, perhaps 15% in the world as a whole, are thought to arise by mechanisms that involve viruses, bacteria, or parasites. The main culprits, as shown in **Table 20–1**, are the DNA viruses. Evidence for their involvement comes partly from the detection of viruses in cancer patients and partly from epidemiology. Liver cancer, for example, is common in parts of the world (Africa and Southeast Asia) where hepatitis-B viral infections

Table 20–1 Viruses Associated with Human Cancers

VIRUS	ASSOCIATED CANCER	AREAS OF HIGH INCIDENCE
DNA viruses		
Papovavirus family		
Papillomavirus (many distinct strains)	warts (benign)	worldwide
	carcinoma of the uterine cervix	worldwide
Hepadnavirus family		
Hepatitis-B virus	liver cancer (hepatocellular carcinoma)	Southeast Asia, tropical Africa
Hepatitis-C virus	liver cancer (hepatocellular carcinoma)	worldwide
Herpesvirus family		
Epstein–Barr virus	Burkitt's lymphoma (cancer of B lymphocytes)	West Africa, Papua New Guinea
	nasopharyngeal carcinoma	Southern China, Greenland
RNA viruses		
Retrovirus family		
Human T-cell leukemia virus type I (HTLV-1)	adult T-cell leukemia/ lymphoma	Japan, West Indies
Human immunodeficiency virus (HIV, the AIDS virus)	Kaposi's sarcoma	Central and Southern Africa

For all these viruses, the number of people infected is much larger than the numbers who develop cancer: the viruses must act in conjunction with other factors. Moreover, some of the viruses contribute to cancer only indirectly; HIV, for example, destroys helper T lymphocytes, which allows a herpes virus to transform endothelial cells. Similarly, hepatitis-C virus causes chronic hepatitis, which promotes the development of liver cancer.

are common, and in those regions the cancer occurs almost exclusively in people who show signs of chronic hepatitis-B infection. Chronic infection with hepatitis-C virus, which has infected 170 million people worldwide, is also clearly associated with the development of liver cancer.

The precise role of a cancer-associated virus is often hard to decipher because there is a delay of many years from the initial viral infection to the development of the cancer. Moreover, the virus is responsible for only one of a series of steps in the progression to cancer, and other environmental factors and genetic accidents are also involved. As we explain later, DNA viruses frequently carry genes that can subvert the control of cell division in the host cell, causing uncontrolled cell proliferation. DNA viruses that operate in this manner include the human papillomaviruses. Some of these cause harmless warts, but others infect the uterine cervix and are implicated in the development of carcinomas of the cervix.

In some cancers, viruses seem to have additional, indirect tumor-promoting actions. The hepatitis-B and C viruses may, for example, favor the development of liver cancer by causing chronic inflammation (hepatitis), which stimulates cell division in the liver, and enhances the viruses' more direct effects on cell proliferation. In AIDS, the human immunodeficiency virus (HIV) promotes development of an otherwise rare cancer called Kaposi's sarcoma by destroying the immune system, thereby permitting a secondary infection with a human herpes virus (HHV-8) that has a direct carcinogenic action. Chronic infection with parasites and bacteria can also promote the development of some cancers. For example, chronic infection of the stomach with the bacterium *Helicobacter pylori*, which causes ulcers, appears to be a major cause of stomach cancer. And bladder cancer in some parts of the world is associated with chronic infection by the blood fluke, *Schistosoma haematobium*, a parasitic flatworm.

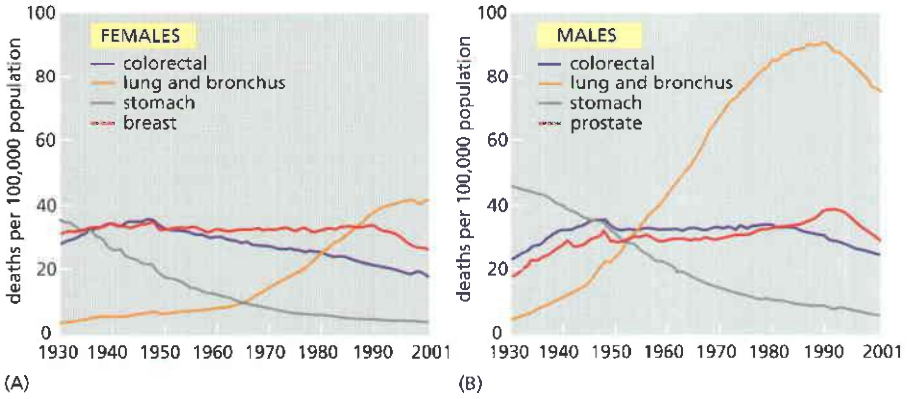


Figure 20–25 Age-adjusted cancer death rates, United States, 1930–2001. Selected death rates adjusted to the age distribution of the US population in 1970, are plotted for (A) females and (B) males. Note the dramatic rise in lung cancer for both sexes, following the pattern of tobacco smoking, and the fall in deaths from stomach cancer, possibly related to changes in diet or in patterns of infection with *Helicobacter*. Recent reductions in other cancer death rates may correspond to improvements in detection and treatment. Age-adjusted data like these are needed to compensate for the inevitable increase in cancer rates as people live longer, on average. (Adapted from Cancer Facts and Figures, 2005. With permission from American Cancer Society.)

Identification of Carcinogens Reveals Ways to Avoid Cancer

Tobacco smoke is by far the most important environmental cause of cancer in the world today. It contains both carcinogens and tumor promoters. Other comparably important chemical causes of cancer in humans remain to be identified. It is sometimes thought that the main environmental causes of cancer are the products of a highly industrialized way of life—the rise in pollution, the enhanced use of food additives, and so on—but there is little evidence to support this view. The idea may have come in part from the identification of some highly carcinogenic materials used in industry, such as 2-naphthylamine and asbestos. Except for the increase in cancers caused by smoking, however, and a remarkable decrease in stomach cancer that may reflect decreasing rates of infection with *Helicobacter pylori*, the incidence of the most common cancers for individuals of a given age has not changed very much in the past hundred years (**Figure 20–25**).

Most of the carcinogenic factors that are known to be significant are by no means specific to the modern world. The most potent known carcinogen, by certain assays at least, is aflatoxin B1 (see Figure 20–22). It is produced by fungi that naturally contaminate foods such as tropical peanuts and is an important cause of liver cancer in Africa and Asia. Some factors that accelerate cancer are intrinsic to our bodies. Thus, in women, the risk of cancer is powerfully influenced by the reproductive hormones produced at different stages of life, and there is a striking correlation between reproductive history and the occurrence of breast cancer (**Figure 20–26**). The reproductive hormones presumably affect breast cancer incidence through their influence on cell proliferation in the breast. Clearly, when attempting to identify the environmental causes of cancer, we need to keep an open mind.

Epidemiology—the analysis of disease frequency in populations—remains the principal tool for finding environmental causes of human cancer. The approach has enjoyed some notable successes, and it promises more to come. By revealing the role of smoking, for example, epidemiology has shown a way to reduce the total cancer death rate in North America and Europe by as much as 25%. The approach works best when applied to a fairly homogeneous population in which it is easy to distinguish between individuals who were exposed to the agent and those who were not, and when the agent under investigation is responsible for most of the cases of a certain kind of cancer. In the early part of the twentieth century, for example, in one British factory all of the men who had been employed in distilling 2-napthylamine (and were thereby subjected to prolonged exposure) eventually developed bladder cancer (see Figure 20–8); the connection was relatively easy to establish because both the chemical and the form of cancer were uncommon in the general population.

In contrast, it is very hard to identify, by epidemiology alone, everyday environmental factors that favor the development of common cancers: most of these factors are probably agents to which we are all exposed to some extent, and

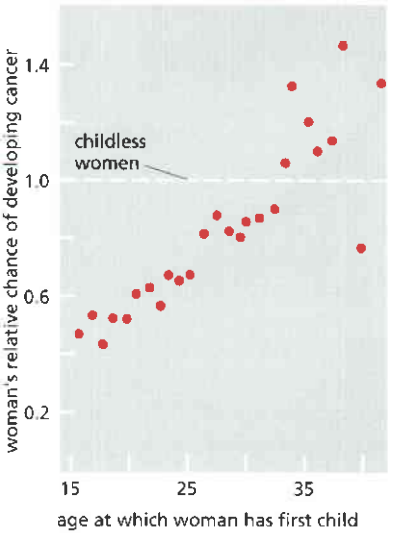


Figure 20–26 Effects of childbearing on the risk of breast cancer. The relative probability of breast cancer developing at some time in a woman's life is plotted as a function of the age at which she gives birth to her first child. The graph shows the probability relative to that for a childless woman. The longer the period of exposure to reproductive hormones up to birth of the first child, the greater the risk. It is thought that the first full-term pregnancy may result in a permanent change in the state of differentiation of the cells of the breast, altering their subsequent responses to hormones. Several other lines of epidemiological evidence also support the view that exposure to certain combinations of reproductive hormones, especially estrogen, can promote development of breast cancer. (From J. Cairns, Cancer: Science and Society. San Francisco: W.H. Freeman, 1978. After B. MacMahon, P. Cole and J. Brown, *J. Natl Cancer Inst.* 50:21–42, 1973. With permission from Oxford University Press.)

many of them probably contribute together to a given cancer's incidence. If, say, eating oranges either doubled or halved the risk of colorectal cancer, it is unlikely that we would find it out, unless we had some prior reason to suspect a connection. The same goes for the countless other substances that we eat, drink, breathe, and put on our bodies. The epidemiological evidence is, however, fairly clear on one dietary point: as mentioned earlier, too much food, leading to obesity, significantly increases the risk of cancer.

Even when evidence is obtained that a substance may be carcinogenic, either from epidemiology or from laboratory tests, it may be difficult to decide what level of human exposure is acceptable. Estimating how many cases of human cancer a certain amount of a substance is likely to cause is difficult; balancing this risk against the utility of the substance is more difficult still. Certain agricultural fungicides, for example, appear to be mildly carcinogenic at high doses in animal tests. But it has been calculated that if they were not used in agriculture, the contamination of food by fungal metabolites such as aflatoxin B1 would cause far more cases of cancer than the fungicide residues in food ever would.

Nevertheless, efforts to identify potential carcinogens still have a central place in our struggle to prevent cancer. Prevention is not only better than a cure, but for many types of cancer, it may be more readily attainable.

Summary

Environmental agents accelerate the rate of tumor development and progression. These can act either as tumor initiators or tumor promoters. Tumor promoters induce an inflammatory response, and they create a local environment that alters gene expression, stimulates cell proliferation, and increases the population of mutant cells created by the initiator. The majority of the known environmental agents that increase cancer development are mutagens, including chemical carcinogens and various forms of radiation, including UV light and ionizing radiation. Because many environmental factors contribute to the development of a given cancer and some of these are under our control, a large proportion of cancers are in principle preventable.

The finding that immigrant populations adopt the cancer patterns of their new host country strongly suggests that the majority of cancers could be avoided by changing the food we eat and other environmental exposures. Epidemiology can be a powerful tool for identifying these environmental effects on human cancer. The epidemiological approach does not require knowing how the environmental agents work, and it can uncover factors that are not chemical entities, such as lifestyle and certain patterns of child-bearing. Individuals can avoid many of the environmental risk factors that are identified in this way. These include smoking tobacco and falling prey to infection with cancer-causing viruses such as papillomaviruses or hepatitis B or C. However, because many factors that affect cancer risk are very difficult to identify, we remain largely ignorant of the principal environmental factors that affect cancer incidence.

FINDING THE CANCER-CRITICAL GENES

As we have seen, cancer depends on the accumulation of inherited changes in somatic cells. To understand it at a molecular level we need to identify the mutations and epigenetic changes involved and to discover how they give rise to cancerous cell behavior. Finding the relevant cells is often easy; they are favored by natural selection and call attention to themselves by giving rise to tumors. But how do we identify the small number of genes with the cancer-promoting changes among all the other genes in the cancerous cells? A similar needle-in-haystack problem arises in any search for a gene underlying a given mutant phenotype, but for cancer the task is particularly complex. A typical cancer depends on a whole set of mutations and epigenetic changes—usually a somewhat different set in each individual patient—and introduction of any single one of these

into a normal cell is not enough to make the cell cancerous. In addition, the cooperation between different altered genes makes it harder to test the significance of an inherited change in any individual gene. To make matters worse, a given cancer cell will also contain a large number of somatic mutations that are accidental by-products of its genetic instability, and it can be difficult to distinguish these meaningless changes from those changes that have a causative role in the disease.

Despite these difficulties, many genes that are repeatedly altered in human cancers have been identified—several hundred of them—although it is clear that many more remain to be discovered. We will call such genes, for want of a better term, **cancer-critical genes**, meaning all genes whose alteration frequently contributes to the causation of cancer. Our knowledge of these genes has accumulated piecemeal through many different and sometimes circuitous approaches, ranging from early studies of cancer-causing infections in chickens to investigations of embryonic development. Analyses of rare but highly revealing inherited forms of cancer have also added to our understanding. More recently, sequencing of DNA from multiple cases of specific types of cancers has begun to give a more systematic picture of the genetic changes that are a regular feature of those diseases.

In this section, we discuss both the methods used for identifying cancer-critical genes and the varied kinds of inherited changes that occur in them during the development of cancer.

The Identification of Gain-of-Function and Loss-of-Function Mutations Requires Different Methods

Cancer-critical genes are grouped into two broad classes, according to whether the cancer risk arises from too much activity of the gene product, or too little. Genes of the first class, in which a gain-of-function mutation can drive a cell toward cancer, are called **proto-oncogenes**; their mutant, overactive or overexpressed forms are called **oncogenes**. Genes of the second class, in which a loss-of-function mutation can contribute to cancer, are called **tumor suppressor genes**. A third class, whose effects are more indirect, are those genes whose mutation results in genomic instability, a class we describe as *DNA maintenance genes.*

As we shall see, mutations in both oncogenes and tumor suppressor genes can have similar effects in enhancing cell proliferation and cell survival, as well as in promoting tumor development. Thus, from the point of view of a cancer cell, oncogenes and tumor suppressor genes—and the mutations that affect them—are flip sides of the same coin. The techniques needed to find these genes, however, differ—depending on whether the genes are overactive or underactive in cancer.

Mutation of a single copy of a proto-oncogene that converts it to an oncogene has a dominant, growth-promoting effect on a cell (**Figure 20–27**A). Thus, we can identify the oncogene by its effect when it is *added*—by DNA transfection, for example, or through infection with a viral vector—to the genome of a suitable type of tester cell. In the case of the tumor suppressor gene, on the other hand, the cancer-causing alleles produced by the change are generally recessive: often (but not always) both copies of the normal gene must be removed or inactivated in the diploid somatic cell before an effect is seen (Figure 20–27B). This calls for a different approach, aimed at detecting what is *missing* in the cancer cell.

In some cases, a specific gross chromosomal abnormality, visible under the microscope, is repeatedly associated with a particular type of cancer. This can give a clue to the location of an oncogene that is activated as a result of the chromosomal rearrangement (as in the example of the chromosomal translocation responsible for chronic myelogenous leukemia, discussed earlier). Alternatively, a visible deletion of a chromosomal segment may reveal the site of a deleted tumor suppressor gene.

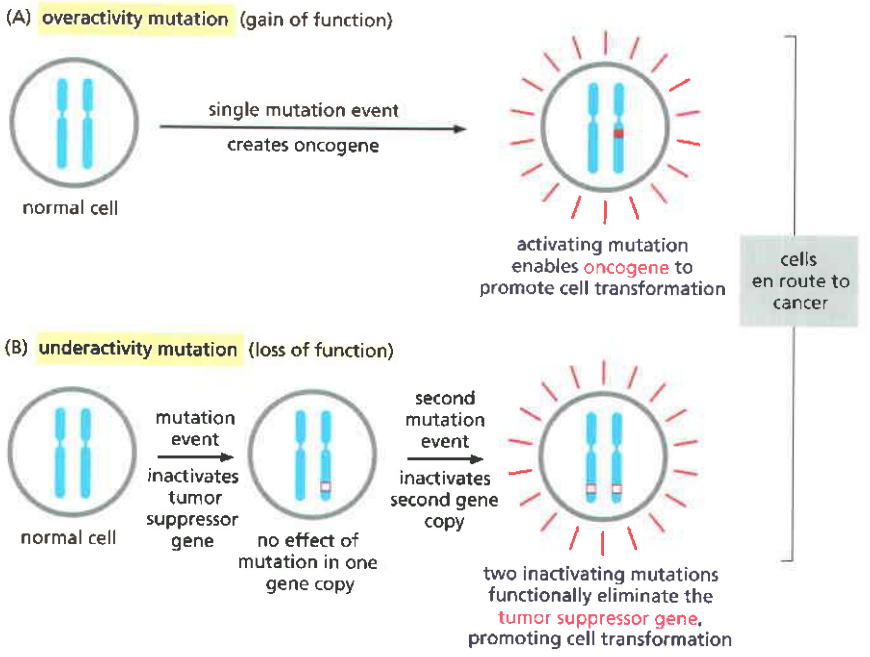


Figure 20–27 Cancer-critical mutations fall into two readily distinguishable categories, dominant and recessive. In this diagram, activating mutations are represented by *solid red boxes,* inactivating mutations by *hollow red boxes.* (A) Oncogenes act in a dominant manner: a gain-of-function mutation in a single copy of the cancer-critical gene can drive a cell toward cancer. (B) Mutations in tumor suppressor genes, on the other hand, generally act in a recessive manner: the function of both alleles of the cancer-critical gene must be lost to drive a cell toward cancer. Although in this diagram the second allele of the tumor suppressor gene is inactivated by mutation, it is often inactivated instead by loss of the second chromosome. Not shown is the fact that mutation of some tumor suppressor genes can have an effect even when only one of the two gene copies is damaged.

Retroviruses Can Act as Vectors for Oncogenes That Alter Cell Behavior

Tumor viruses have played a remarkable part in the search for the genetic causes of human cancer. Although viruses have no role in the majority of common human cancers, they are more prominent as causes of cancer in some animal species. Analysis of these animal tumor viruses provided a critical key to understanding the mechanisms of cancer in general and to the discovery of oncogenes in particular.

One of the first animal viruses to be implicated in cancer was discovered nearly 100 years ago in chickens, which are subject to infections that cause connective-tissue tumors, or sarcomas. The infectious agent was characterized as a virus—*the Rous sarcoma virus,* which we now know to be an RNA virus. Like all the other *RNA tumor viruses* discovered since, it is a **retrovirus.** When it infects a cell, its RNA is copied into DNA by reverse transcription, and the DNA is inserted into the host genome, where it can persist and be inherited by subsequent generations of cells. The Rous sarcoma virus carries an oncogene that causes cancer in chickens. This oncogene is not necessary for the virus's own survival or reproduction, as demonstrated by the discovery of mutant forms of the virus that multiply normally but no longer make their host cells cancerous—a process called cell *transformation.* Some of these mutants were found to have deleted all or part of a gene that codes for a protein called **Src** (pronounced "Sarc"). Other mutations in this gene made the transforming effect of the virus temperature-sensitive: infected cells show a transformed phenotype at 34°C, but return to the normal phenotype within a few hours when the temperature is raised to 39°C.

What was the origin of this gene? In 1975, when a radioactive DNA copy of the viral *Src* gene sequences was used as a probe to search for closely related sequences by DNA–DNA hybridization (see Figure 8–36), the genomes of vertebrate cells were found to contain sequences that are closely similar, but not identical, to this viral gene. The viral oncogene had evidently been picked up accidentally by the retrovirus from the genome of a previous infected host cell, but the host cell proto-oncogene (*c-Src*) had undergone mutation in the process to become an oncogene (*v-Src*) (**Figure 20–28**). This finding that cancer could originate from mutation of a specific gene that was present in the genome of a normal animal profoundly altered the field of cancer research.

A large number of other oncogenes have since been identified in other retroviruses and analyzed in similar ways. Each has led to the discovery of a corresponding proto-oncogene that is present in normal animal cells.

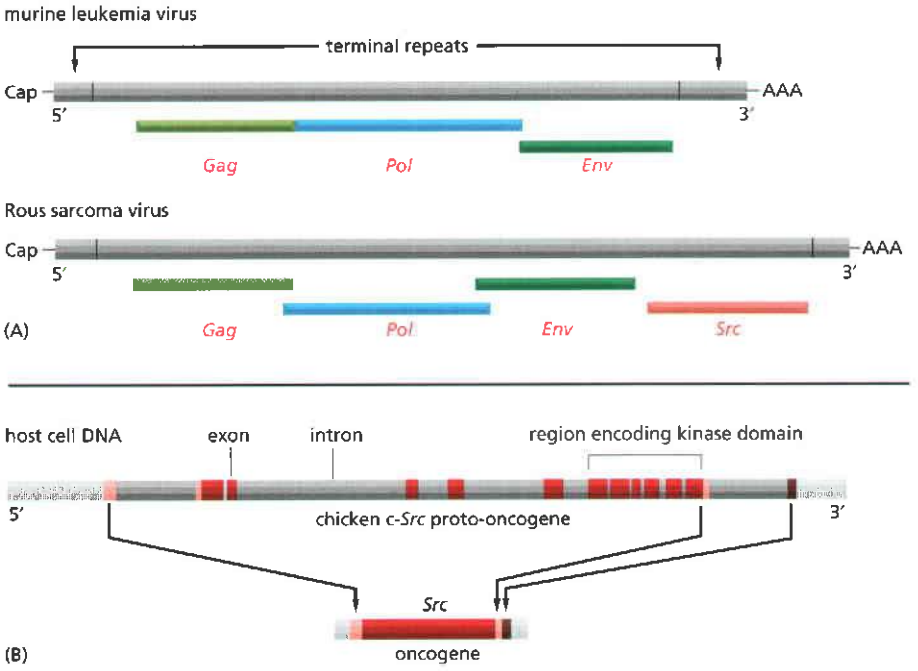


Figure 20–28 The structure of the Rous sarcoma virus. (A) The organization of the viral genome as compared with that of a more typical retrovirus (murine leukemia virus). Rous sarcoma virus is unusual among the retroviruses that carry oncogenes in that it has retained the three viral genes required for the ordinary viral life cycle: *Gag* (which produces a polyprotein that is cleaved to generate the capsid proteins), *Pol* (which produces reverse transcriptase and an enzyme involved in integrating the viral chromosome into the host genome), and *Env* (which produces the envelope glycoprotein). In other oncogenic retroviruses, one or more of these viral genes are wholly or partly lost when the retrovirus acquires the transforming oncogene. Therefore the transforming virus can generate infectious particles only in a cell that is simultaneously infected with a nondefective, nontransforming *helper virus*, which supplies the missing functions. (B) The relationship between the *v-Src* oncogene and the cellular *Src* proto-oncogene from which it has been derived. The introns present in cellular *Src* have been spliced out of *v-Src;* in addition, *v-Src* contains mutations that alter the amino acid sequence of the protein (not shown), making it hyperactive and unregulated as a tyrosine-specific protein kinase (see Figure 3–69). The v-Src protein is also greatly overproduced because the virus makes large amounts of RNA.

Different Searches for Oncogenes Have Converged on the Same Gene—Ras

While some researchers searched for oncogenes in retroviruses, others took a more direct approach and searched for DNA sequences in human cancer cells that could provoke uncontrolled proliferation when introduced into noncancerous cell lines. As tester cells for the assay, they usually used cell lines derived from mouse fibroblasts. These cells, previously selected to proliferate indefinitely in culture, are thought to already contain genetic alterations that take them part of the way toward malignancy. For this reason, the addition of a single oncogene can be enough to produce a dramatic effect.

To detect an oncogene in this way, DNA is extracted from the tumor cells, broken into fragments, and introduced into the cultured cells. If any of the fragments contains an oncogene, small colonies of abnormally proliferating cells begin to appear; because the cells forming such colonies display cancerous properties, they are said to be *transformed* or to have undergone **transformation.** <CCGA> Each colony is a clone of cells that originated from a single cell, carrying an added gene that released the cell from some of the social controls on cell proliferation; the transformed cells forming the colony outgrow the untransformed cells in the culture and pile up layer upon layer as they proliferate (**Figure 20–29**).

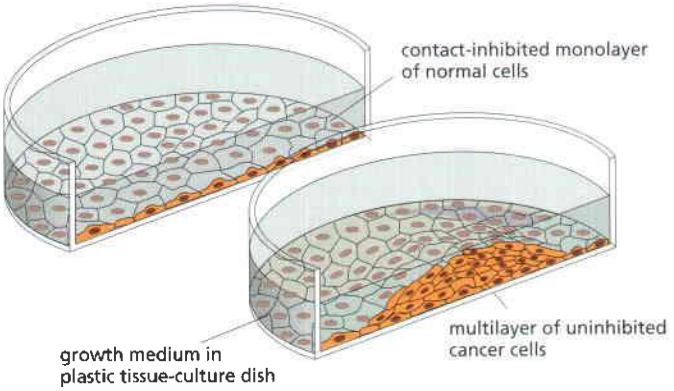


Figure 20–29 Loss of contact inhibition in cell culture. Most normal cells stop proliferating once they have carpeted the dish with a single layer of cells: proliferation seems to depend on contact with the dish, and to be inhibited by contacts with other cells—a phenomenon known as "contact inhibition." Cancer cells, in contrast, usually disregard these restraints and continue to grow, so that they pile up on top of one another.

This assay led to the isolation and sequencing of the first human oncogene, a mutant version of the proto-oncogene ***Ras***. The *Ras* gene is now known to be mutated in about one in five human cancers. The discovery of the human *Ras* oncogene was all the more dramatic because, shortly before, a mutated *Ras* gene had been found to be the tumor-causing gene in a retrovirus that causes sarcomas in rodents. This discovery in the early 1980s of the same oncogene in human tumor cells and in an animal tumor virus was electrifying. The implication that cancers are caused by mutations in a limited number of cancer-critical genes changed our understanding of the molecular biology of cancer.

As discussed in Chapter 15, normal Ras proteins are monomeric GTPases that help transmit signals from cell-surface receptors to the cell interior. <GAAC> The *Ras* oncogenes isolated from human tumors contain point mutations that create a hyperactive Ras protein that cannot shut itself off by hydrolyzing its bound GTP to GDP. Because this makes the protein hyperactive, its effect is dominant—that is, only one of the cell's two gene copies needs to change to have an effect. Because the *Ras* genes are mutated in a wide range of human cancers, they remain one of the most important examples of cancer-critical genes.

Through methods such as those we have just described, and various other routes, we now know of hundreds of proto-oncogenes—that is, genes that can be converted by activating mutations into oncogenes that help to cause cancer.

Studies of Rare Hereditary Cancer Syndromes First Identified Tumor Suppressor Genes

Identifying a gene that has been inactivated in the genome of a cancer cell requires a different strategy from finding a gene that has become hyperactive: one cannot, for example, use a cell transformation assay to identify something that simply is not there. The key insight that led to the discovery of the first tumor suppressor gene came from studies of a rare type of human cancer, **retinoblastoma**, which arises from cells in the retina of the eye that are converted to a cancerous state by an unusually small number of mutations. As often happens in biology, the discovery arose from examination of a special case, but it turned out to reveal a gene of widespread relevance. In fact, genes identified in rare familial cancer syndromes are commonly found to be relevant in more common sporadic cancers, where they often serve as tumor suppressors.

Retinoblastoma occurs in childhood, and tumors develop from neural precursor cells in the immature retina. About one child in 20,000 is afflicted. One form of the disease is hereditary, and the other is not. In the hereditary form, multiple tumors usually arise independently, affecting both eyes; in the nonhereditary form, only one eye is affected, and by only one tumor. A few individuals with retinoblastoma have a visibly abnormal karyotype, with a deletion of a specific band on chromosome 13 that, if inherited, predisposes an individual to the disease. Deletions of this same region are also encountered in tumor cells from some patients with the nonhereditary disease, which suggested that the cancer was caused by loss of a critical gene in that location.

Using the location of this chromosomal deletion, it was possible to clone and sequence the ***Rb* gene**. It was then discovered that those who suffer from the hereditary form of the disease have a deletion or loss-of-function mutation present in one copy of the *Rb* gene in every somatic cell. These cells are predisposed to becoming cancerous, but do not do so if they retain one good copy of the gene. The retinal cells that are cancerous are defective in both copies of *Rb* because of a somatic event that has eliminated the function of the previously good copy.

In patients with the nonhereditary form of the disease, by contrast, the noncancerous cells show no defect in either copy of *Rb*, while the cancerous cells have become defective in both copies. These nonhereditary retinoblastomas are very rare because they require two independent events that inactivate the same gene on two chromosomes in a single retinal cell lineage (**Figure 20–30**). The *Rb* gene is also missing in several common types of sporadic cancer, including carcinomas of

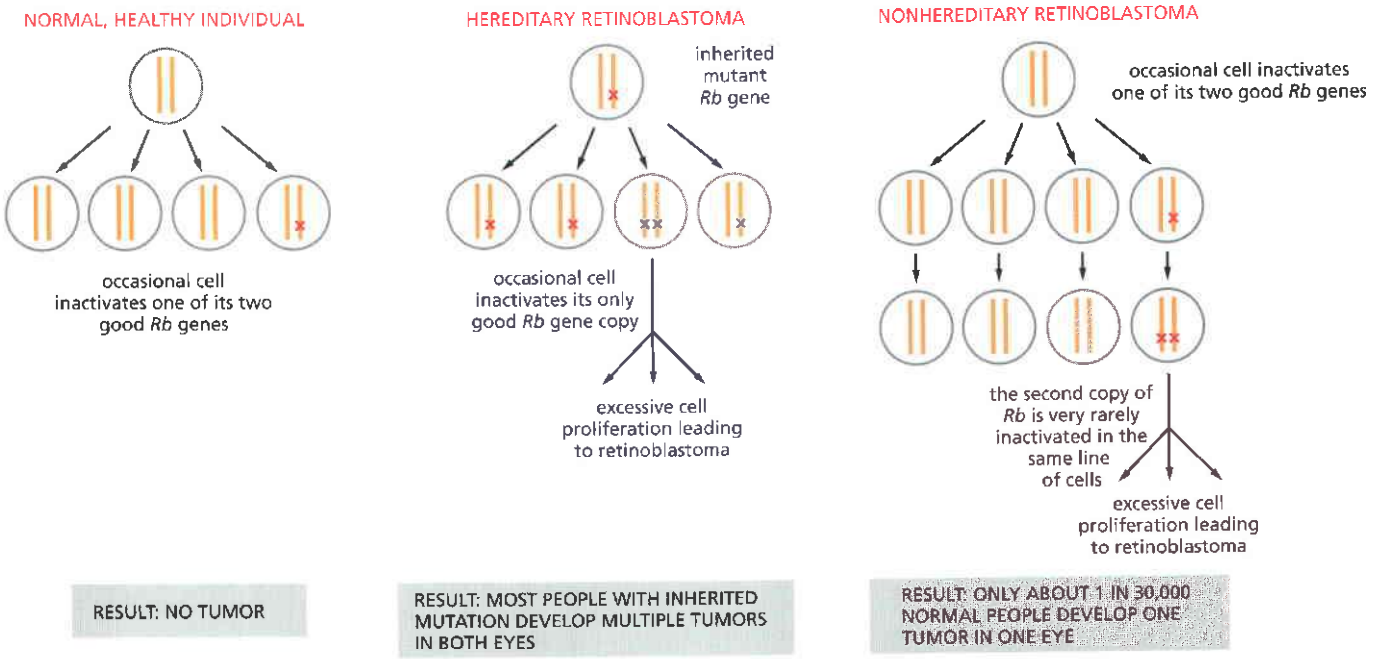


Figure 20–30 The genetic mechanisms that cause retinoblastoma. In the hereditary form, all cells in the body lack one of the normal two functional copies of the *Rb* tumor suppressor gene, and tumors occur where the remaining copy is lost or inactivated by a somatic event (either mutation or epigenetic silencing). In the nonhereditary form, all cells initially contain two functional copies of the gene, and the tumor arises because both copies are lost or inactivated through the coincidence of two somatic events in one line of cells.

lung, breast, and bladder. These more common cancers arise by a more complex series of genetic changes than does retinoblastoma, and they make their appearance much later in life. But in all of them, it seems, loss of *Rb* function is frequently a major step in the progression toward malignancy.

The *Rb* gene encodes the **Rb protein**, which is a universal regulator of the cell cycle present in almost all cells of the body (see Figure 17–62). It acts as one of the main brakes on progress through the cell-division cycle, and its loss can allow cells to enter the cell cycle inappropriately, as we discuss later.

Tumor Suppressor Genes Can Also Be Identified from Studies of Tumors

The Rb story illustrates how rare hereditary cancer syndromes can be used to uncover tumor suppressor genes that are relevant to common sporadic cancers. However, only a few of the tumor suppressor genes now known to be important have been discovered in this way. A more direct approach to the identification of such genes involves comparing tumor cells with non-cancerous cells from the same patient so as to discover what exactly, out of the 3 billion nucleotides of the human genome, is missing, functionally defective, or abnormally silenced. Because of the genetic instability of cancer cells, there is usually a great deal changed. Most of the alterations will be random, accidental by-products of the genetic instability. Therefore, a relevant tumor suppressor gene can only be identified by the criterion that it is repeatedly missing, defective, or otherwise silenced in many independent cases of a particular type of cancer. Tracking down tumor suppressor genes in this way is a hard task, but feasible with modern techniques for large-scale DNA analysis (as we explain below). Dozens of tumor suppressor genes have already been well characterized and many more are known.

Both Genetic and Epigenetic Mechanisms Can Inactivate Tumor Suppressor Genes

It is the inactivation of tumor suppressor genes that is dangerous. This inactivation can occur in many ways, with different combinations of mishaps converging to eliminate or cripple both gene copies. The first copy may, for example, be lost by a small chromosomal deletion or inactivated by a point mutation. The second copy is commonly eliminated by a less specific and more probable

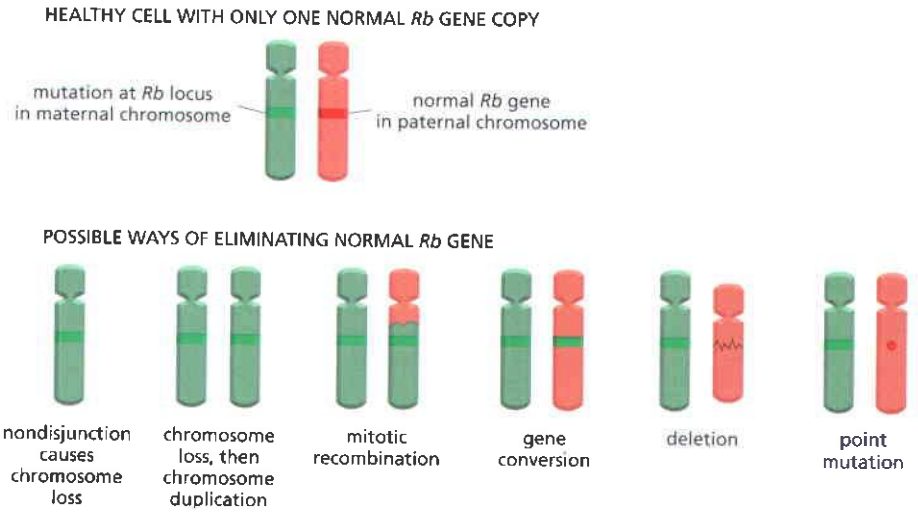


Figure 20–31 Six ways of losing the remaining good copy of a tumor suppressor gene through changes in DNA sequences. A cell that is defective in only one of its two copies of a tumor suppressor gene—for example, the *Rb* gene—usually behaves as a normal, healthy cell; the diagrams below show how this cell may lose the function of the other gene copy as well and thereby progress toward cancer. A seventh possibility, frequently encountered with some tumor suppressors, is that the gene may be silenced by an epigenetic change, without alteration of the DNA sequence, as illustrated in Figure 20–32. (After W.K. Cavenee et al., *Nature* 305:779–784, 1983. With permission from Macmillan Publishers Ltd.)

mechanism: the chromosome carrying the remaining normal copy may be lost from the cell through errors in chromosome segregation, or the normal gene may be replaced by a mutant version through either mitotic recombination or a gene conversion event.

Figure 20–31 summarizes the range of possible ways to lose the remaining good copy of a tumor suppressor gene through DNA sequence changes, using the *Rb* gene as an example. It is important to note that, except for the point mutation mechanism illustrated at the far right, these pathways all produce cells that carry only a single type of DNA sequence in the chromosomal region containing their *Rb* genes—a sequence that is identical to the sequence in the original mutant chromosome.

As discussed in Chapter 4, normal human genetic variation makes each of our maternal and paternal chromosome sets distinguishably different. On average, human DNA sequences differ—that is, we are heterozygous—at roughly one in every thousand nucleotides. Where a large segment of one chromosome has been either lost or converted to the DNA sequence in its homologous chromosome, as in Figure 20–31, there is a *loss of heterozygosity (LOH)*: only one version of each variable DNA sequence in that neighborhood remains. Millions of common sites of heterozygosity in the human genome have been mapped as part of the Human Genome Project: each of these sites is characterized by a specific DNA sequence that is known to be polymorphic—that is, to occur commonly in two or more slightly different versions in the human population. Given a sample of tumor DNA, one can check which of the versions of these polymorphic sequences are present. The same can be done with a sample of DNA from noncancerous tissue from the same patient, for comparison. A loss of heterozygosity throughout a region of the genome containing one or more polymorphic sites, or loss of a genetic marker sequence that is seen in the non-cancerous control DNA, can point the way toward a chromosomal region that contains a relevant tumor suppressor gene. However, because of their genetic instability, cancer cells often exhibit a loss of heterogeneity for many different chromosomal regions. Therefore the detection of tumor suppressor genes by this approach generally requires the subtraction of a large amount of random noise.

Epigenetic changes provide another important way to permanently inactivate a tumor suppressor gene. Most commonly, the gene may become packaged into heterochromatin and the C nucleotides in CpG sequences in its promoter may become methylated in a heritable manner (see Figure 20–12). These mechanisms can irreversibly silence the gene in a cell and in all of its progeny. Given a catalog of possible tumor suppressor genes, it is relatively easy to test their promoters for abnormal amounts of DNA methylation. Studies of this type suggest that epigenetic gene silencing is a frequent event in tumor progression, and epigenetic mechanisms are now thought to help inactivate several different tumor suppressor genes in most human cancers (Figure 20–32).

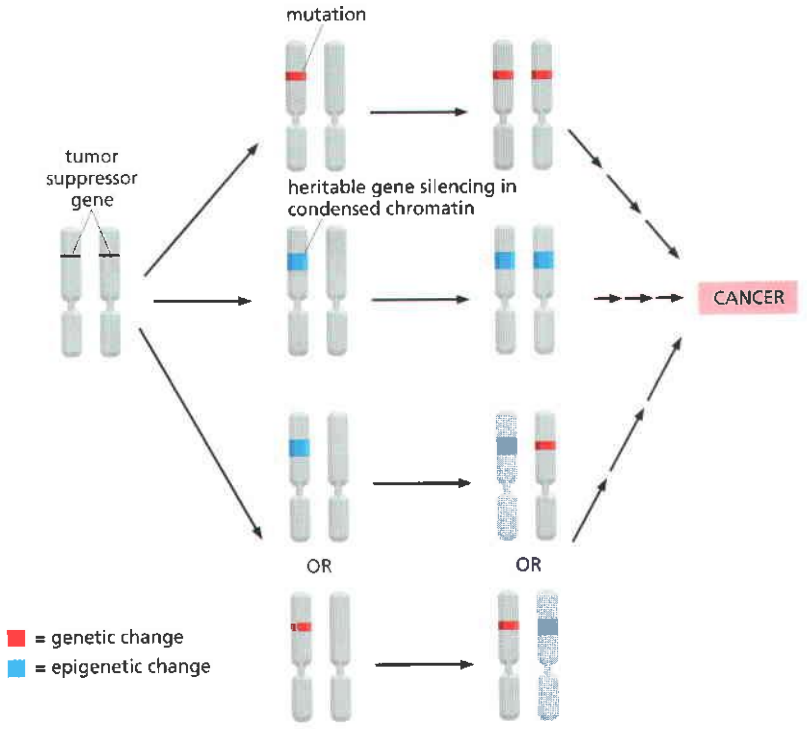


Figure 20–32 The pathways leading to loss of tumor suppressor gene function in cancer involve both genetic and epigenetic changes. As discussed in Chapter 4, the packaging of a gene into condensed chromatin can prevent its expression in a way that is inherited when a cell divides (see Figure 4–52). As indicated, the changes that silence tumor suppressor genes can occur in any order.

Genes Mutated in Cancer Can Be Made Overactive in Many Ways

In the case of a proto-oncogenes it is gene activation that leads toward cancer. Figure 20–33 summarizes the types of accidents that can convert a proto-oncogene into an oncogene. (1) A small change in DNA sequence such as a point mutation or deletion may produce a hyperactive protein when it occurs within a protein-coding sequence, or lead to protein overproduction when it occurs within a regulatory region for that gene. (2) Gene amplification events, such as those that can be caused by errors in DNA replication, may produce extra gene copies; this can lead to overproduction of the protein (Figure 20–34). (3) A chromosomal rearrangement, involving the breakage and rejoining of the DNA helix, may either change the protein-coding region, resulting in a hyperactive fusion protein, or alter the control regions for a gene so that a normal protein is overproduced.

Specific types of abnormality are characteristic of particular genes and of the responses to particular carcinogens. For example, 90% of the skin tumors

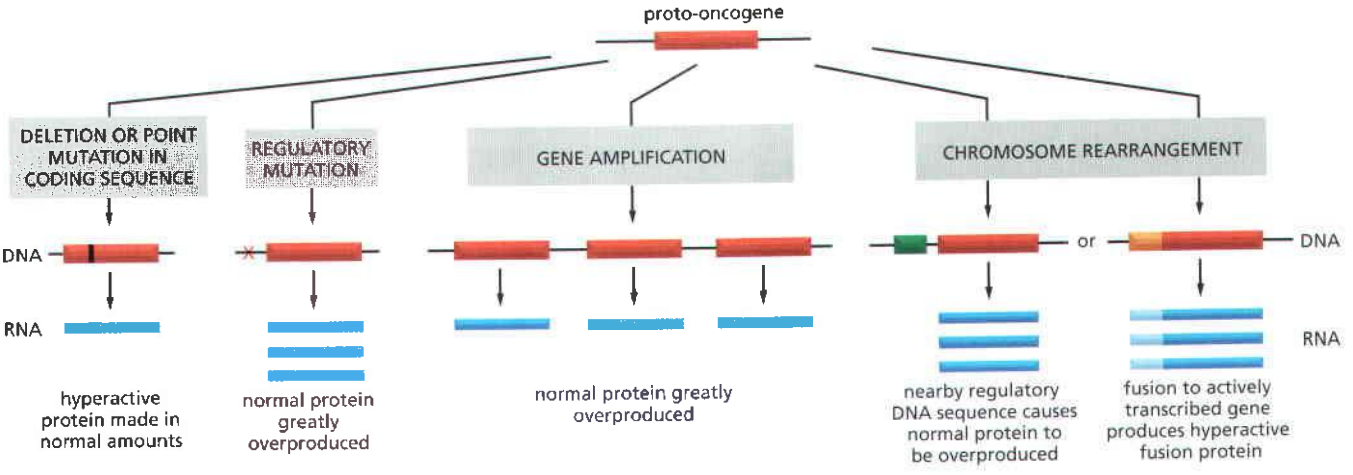


Figure 20–33 The types of accidents that can make a proto-oncogene overactive and convert it into an oncogene.

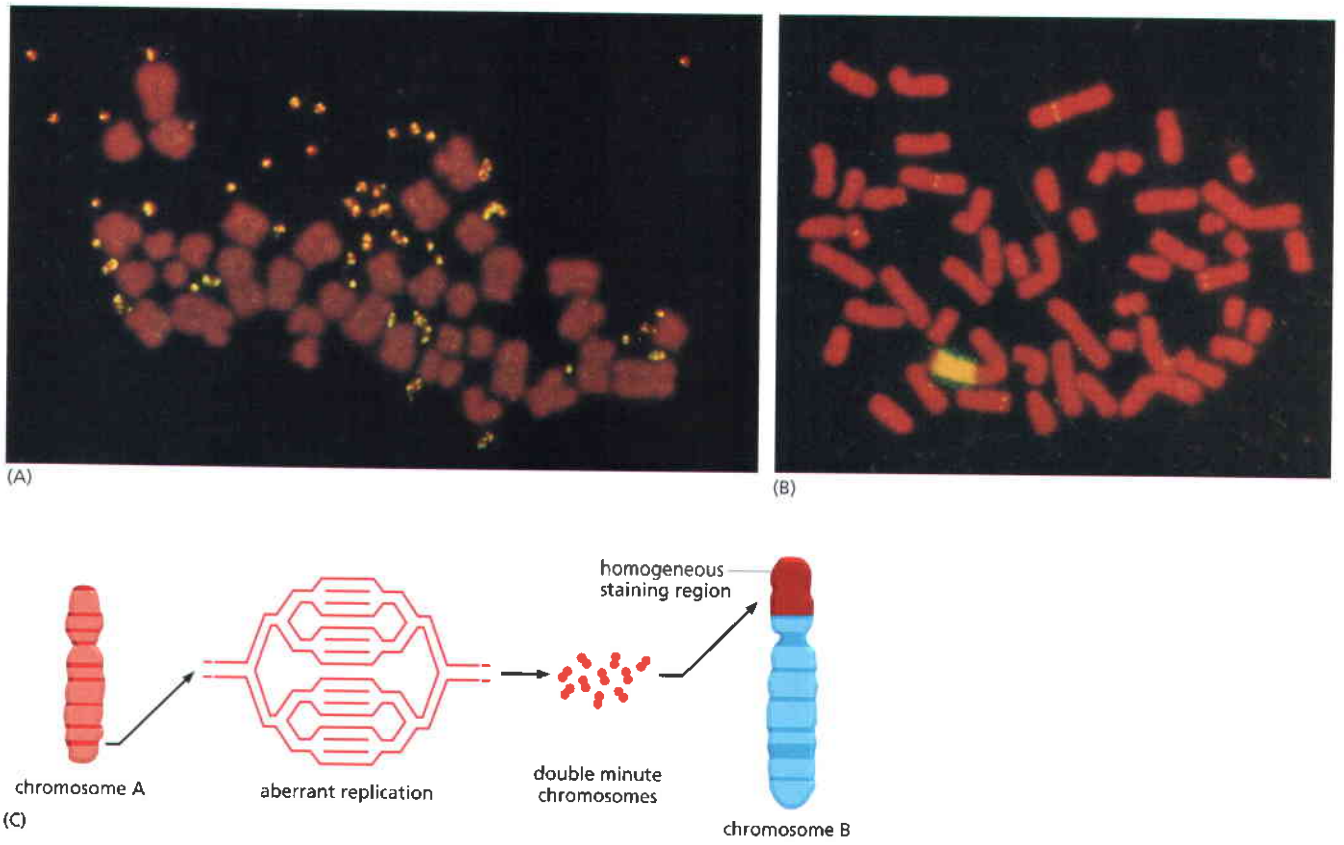


Figure 20–34 Chromosomal changes in cancer cells resulting from gene amplification. In these examples, the numbers of copies of a *Myc* proto-oncogene have been amplified. Amplification of oncogenes is common in carcinomas and is often visible as a curious change in the karyotype: the cell is seen to contain additional pairs of miniature chromosomes—so-called *double minute chromosomes*—or to have a *homogeneously staining region* interpolated in the normal banding pattern of one of its regular chromosomes. Both these aberrations consist of massively amplified numbers of copies of a small segment of the genome. The chromosomes are stained with a *red* fluorescent dye, while the multiple copies of the *Myc* gene are detected by *in situ* hybridization with a *yellow* fluorescent probe. (A) Karyotype of a cell in which the *Myc* gene copies are present as double minute chromosomes (paired *yellow specks*). (B) Karyotype of a cell in which the multiple *Myc* gene copies appear as a homogeneously staining region *(yellow)* interpolated in one of the regular chromosomes. (Ordinary single-copy *Myc* genes can be detected as tiny *yellow dots* elsewhere in the genome.) (C) Schematic of how gene amplification occurs. It is thought that a rare, abnormal DNA replication event produces a chromosome with extra copies of one chromosomal region, as shown. The repair of this structure releases DNA circles that can replicate to form long tandemly repeated sequences, producing double minute chromosomes. As the result of a second rare event, the DNA from one of these chromosomes can become integrated into a new site on a normal chromosome to produce a homogeneously staining region. Other pathways can also amplify genes, such as that described later in Figure 20–41. (A and B, courtesy of Denise Sheer.)

evoked in mice by painting the skin with the tumor initiator dimethylbenz[a]anthracene (DMBA) have an A-to-T alteration at exactly the same site in a mutant *Ras* gene; presumably, of the many mutations caused by DMBA, only those at this site efficiently stimulate skin cells to form a tumor.

The receptor for the extracellular signal protein *epidermal growth factor (EGF)*, by contrast, can be activated by a deletion that removes part of its extracellular domain. These mutant receptors are able to form active dimers even in the absence of EGF, and thus they produce an inappropriate stimulatory signal, analogous to a faulty doorbell that rings even when nobody is pressing the button. Mutations of this type are found in the most common type of human brain tumor, called glioblastoma.

The Myc protein, on the other hand, generally contributes to cancer by being overproduced in its normal form. The Myc protein acts in the nucleus to stimulate cell growth and division, as discussed in Chapter 17; thus, excessive quantities of Myc cause cells to proliferate in circumstances where a normal cell would halt. Overproduction of Myc can occur in various ways. In some cases, the gene is amplified—that is, errors of DNA replication lead to the creation of large numbers of gene copies in a single cell (see Figure 20–34). More commonly, the

overproduction appears to be due to a change in a regulatory element that acts on the gene. For example, a chromosomal translocation can inappropriately bring powerful gene regulatory sequences next to the Myc protein-coding sequence, so as to produce unusually large amounts of *Myc* mRNA. Thus, in Burkitt's lymphoma a translocation brings the *Myc* gene under the control of sequences that normally drive the expression of antibody genes in B lymphocytes. As a result, the mutant B cells tend to proliferate excessively and form a tumor. Similar specific chromosome translocations are common in other lymphomas and leukemias.

The Hunt for Cancer-critical Genes Continues

The sequencing of the human genome has opened up new avenues for the systematic discovery of cancer-critical genes. It is now possible, in principle, to examine every one of the approximately 25,000 human genes in a given cancer cell line, or in samples of tissue from a set of cases of cancer of a given type, looking for potentially significant abnormalities, using automated analysis of either their genomic DNA or the mRNAs the cells produce. By analyzing substantial numbers of different cancers, it should be possible eventually to identify all the genes that are commonly altered in human cancer. Although enormously costly, large-scale DNA sequencing efforts have already begun to identify new human oncogenes. One example is the finding that a hyperactive form of the Raf protein kinase (discussed in Chapter 15) is present in a high percentage of melanomas and, at lower frequency, in other cancers.

Straightforward DNA sequencing is not the only way to tackle the problem, however, and there are other powerful new methods for identifying new oncogenes and tumor suppressor genes that seem more efficient. Three different types of approaches appear particularly promising.

1. *Comparative genomic hybridization (CGH)* uses fluorescent labeling of fragments of DNA extracted from normal cells and from cancer cells to identify regions of the genome that are amplified or deleted in a given type of cancer (**Figure 20–35**). The labeled fragments are allowed to hybridize to DNA microarrays in which each spot corresponds to a known location in the normal genome. Different spots light up in different fluorescent colors,

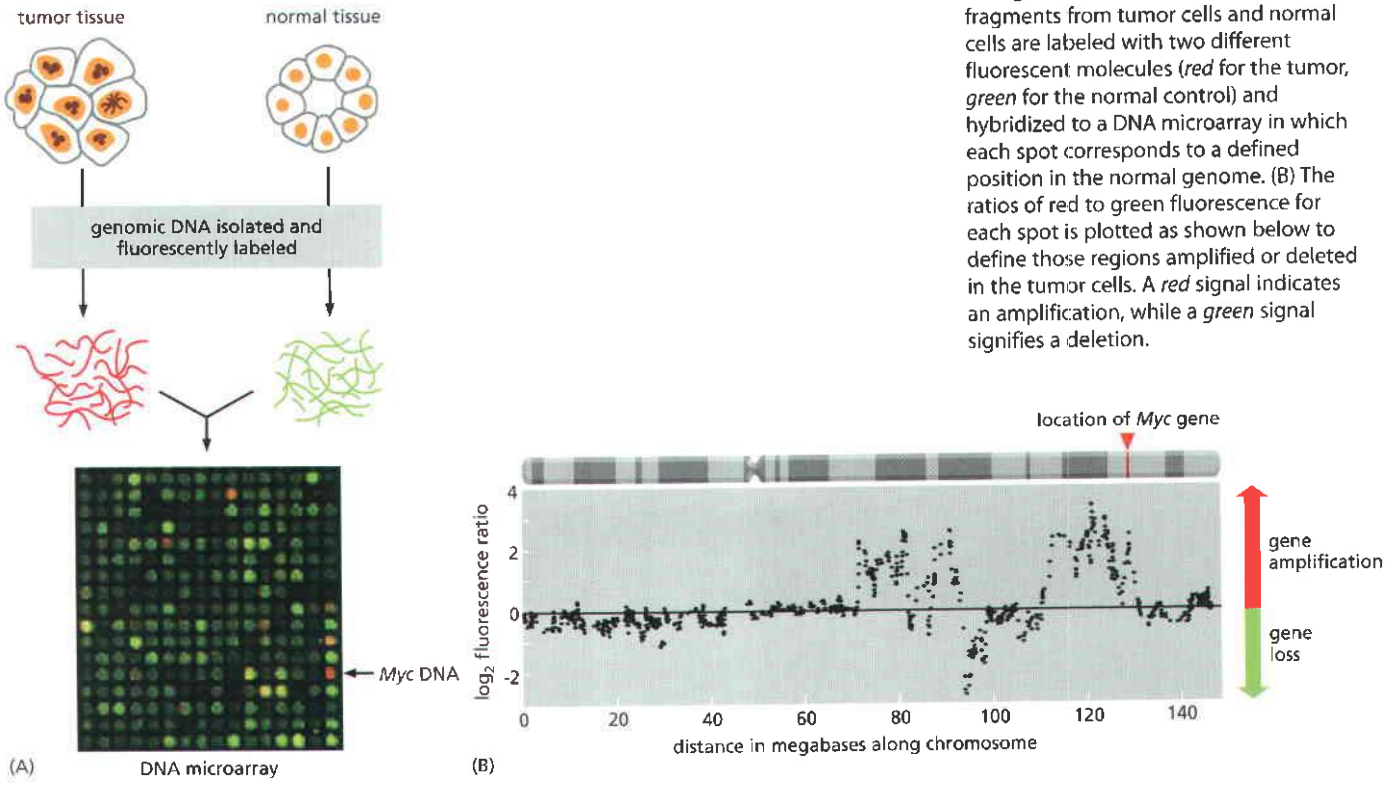


Figure 20–35 Comparative genomic hybridization for detection of the DNA changes in tumor cells. (A) DNA fragments from tumor cells and normal cells are labeled with two different fluorescent molecules (*red* for the tumor, *green* for the normal control) and hybridized to a DNA microarray in which each spot corresponds to a defined position in the normal genome. (B) The ratios of red to green fluorescence for each spot is plotted as shown below to define those regions amplified or deleted in the tumor cells. A *red* signal indicates an amplification, while a *green* signal signifies a deletion.

according to the ratio of normal to cancerous labeled fragments that bind to them. In this way, one can locate regions of the cancer cell genome that have been amplified or deleted. One then searches more narrowly for candidate genes within these regions that may contribute to cancer development.

2. DNA microarrays can also be used to reveal specific changes in gene expression associated with cancer. In this case, a cell's population of mRNAs is used to prepare the probe for hybridization, rather than its chromosomal DNA (see Figure 8–76).
3. Finally, large genetic screens based on RNA interference (RNAi) technology (discussed in Chapter 8) provide a powerful new functional approach for identifying tumor suppressor genes. The expression of small interfering RNAs (siRNA) in a cell can inactivate both copies of a gene by either destroying the corresponding mRNA or inhibiting its translation (see p. 571). This provides an efficient way, in principle, to identify any tumor suppressor gene whose loss promotes cancerous transformation in the particular animal or cell line being tested.

Whatever the approach used to identify a new candidate cancer-critical gene, it remains a challenge to determine whether the gene really contributes to cancer causation. Testing for the effects of an overexpression of candidate oncogenes or an inhibition of candidate tumor suppressor genes in cell culture assays can help. But more convincing tests use transgenic mice that overexpress candidate oncogenes and knockout mice that lack candidate tumor suppressor genes. In both types of mice, cancer development should be accelerated if the genes are real culprits.

To search for novel cancer-critical genes in the ways just described, there is in principle no need to know at the outset what the normal functions of the gene are or how altering them promotes cancer development. As our understanding of cell biology improves, however, it becomes easier to guess which genes are likely suspects and to test their culpability by directly determining if they are frequently altered in specific cancers. The task of finding cancer-critical genes is therefore closely entangled with the problem of discovering what they normally do and how, when altered, they contribute to the development of cancer. This is the topic of the next section.

Summary

Cancer-critical genes can be classified into two groups, according to whether their gain of function or their loss of function contributes to cancer development. Gain-of-function mutations that convert proto-oncogenes to oncogenes stimulate cells to increase their numbers when they should not; loss-of-function mutations of tumor suppressor genes abolish the inhibitory controls that normally help to hold cell numbers in check. Oncogenes have a dominant genetic effect, and many of them were first discovered because they cause cancer in animals when introduced by a retrovirus that originally picked up a normal form of the gene (a proto-oncogene) from a previous host cell. Oncogenes can also be identified by characteristic chromosomal aberrations that can activate a proto-oncogene.

Mutations in tumor suppressor genes are generally recessive, in that cells tend to behave normally until both gene copies are deleted, inactivated, or silenced epigenetically. Several of these genes were initially identified in rare, hereditary cancer syndromes, but their loss or inactivation is a common feature of many sporadic cancers. Individuals who inherit one defective and one functional copy of a tumor suppressor gene have an increased predisposition toward developing cancer because a single change, anywhere in the body, that eliminates or inactivates the remaining good gene copy is enough to produce a cell that totally lacks the tumor-suppressor function.

The quest for cancer-critical genes continues, with increasingly powerful tools available for systematically searching the DNA or mRNAs of cancer cells for either significant mutations or altered expression. Once a candidate gene has been identified, its importance for cancer development can be assessed in mice—by overexpressing it in the case of a candidate oncogene, or by inactivating it in the case of a candidate tumor suppressor gene.

THE MOLECULAR BASIS OF CANCER-CELL BEHAVIOR

The cancer-gene hunter views oncogenes and tumor suppressor genes—and the mutations and epigenetic changes that affect them—differently. But, from a cancer cell's point of view, they are two sides of the same coin. The same kinds of effects on cell behavior can result from mutations in either class of genes, because most cellular control mechanisms have both stimulatory (proto-oncogene) and inhibitory (tumor suppressor) components. If the aim is to understand how cancer cells function, and to do this in terms of molecular genetics, the important distinction is not between tumor-suppressor genes and proto-oncogenes, but between cancer-critical genes that act in different biochemical and regulatory pathways.

Some of the pathways important in cancer convey signals from a cell's environment (discussed in Chapter 15); others are responsible for the cell's internal programs, such as those that control the cell cycle (discussed in Chapter 17) or cell death (discussed in Chapter 18); still others govern the cell's movements (discussed in Chapter 16) or mechanical interactions with its neighbors (discussed in Chapter 19). The various pathways are linked and interdependent in complex ways. Much of what we know about them has been learned as a byproduct of cancer research; conversely, research on basic aspects of cell biology has transformed our understanding of cancer.

In the first section of this chapter, we summarized the general properties that make a cell cancerous and listed the kinds of misbehaviors that a cancer cell displays. In this section, we consider how these characteristic misbehaviors arise from mutations in cancer-critical genes and how the functions of these genes in the context of cancer can be determined. We end the section by discussing colon cancer as an extended example, showing how a succession of changes in cancer-critical genes enables a tumor to evolve from one pattern of bad behavior to another that is worse.

Studies of Both Developing Embryos and Genetically Engineered Mice Have Helped to Uncover the Function of Cancer-critical Genes

Given a gene that is mutated in a cancer, we need to understand both how the gene functions in normal cells and how mutations in the gene contribute to the aberrant behaviors characteristic of cancer cells. When the *Rb* gene was originally cloned, for example, all that was known was that it was mutated in retinoblastomas. In the case of *Ras*, the mutant gene was known to direct cells to proliferate excessively and inappropriately in culture, but how the Ras protein functioned in either normal or cancer cells was a mystery. For both Rb and Ras, cancer research was the starting point for studies that revealed the key role of these proteins in normal cells—Rb as a cell-cycle inhibitor and Ras as a central component of intracellular signaling pathways.

Today, we know much more about cells, so that when a new gene is identified as critical for cancer, it often turns out to be familiar from studies in another context. Many oncogenes and tumor suppressor genes, for example, are homologs of genes already known for their role in embryonic development. Examples include components of practically all the major signaling pathways through which cells communicate during development (discussed in Chapters 15 and 22): the Wnt, Hedgehog, TGFβ, Notch, and receptor tyrosine kinase signaling pathways all include components encoded by cancer-critical genes—with Ras being part of the last of these pathways.

With hindsight, this is no surprise. As we discuss in Chapter 23, the same signaling mechanisms that control embryonic development operate in the normal adult body to control cell turnover and maintain homeostasis. Both the development of a multicellular animal and its maintenance as an adult depend on

cell–cell communication and on regulated cell growth, cell division, cell differentiation, cell death, cell movement, and cell adhesion—in other words, on all the aspects of cell behavior that are deranged in cancer. Developmental biology, often using model animals such as *Drosophila* and *C. elegans*, thus provides a key to the normal functions of many cancer-critical genes.

Ultimately, however, we want to know what alterations in these genes do to cells in the tissues that give rise to the cancer. Some information can be obtained by studying cells in culture or by examining samples from human cancer patients. But to investigate how mutations in cancer-critical genes affect tissues in a whole organism, transgenic and knockout mice have proven particularly useful.

Transgenic mice that carry an oncogene in all their cells can be generated by methods described in Chapter 8. Oncogenes introduced in this way may be expressed in many tissues or in only a select few, depending on the regulatory DNA associated with the transgene. Studies of such mice confirm that a single oncogene is generally not sufficient to turn a normal cell into a cancer cell. Typically, in mice engineered to express a *Myc* or *Ras* oncogene, some of the tissues that express the oncogene may show enhanced cell proliferation, and, over time, occasional cells will undergo further changes to give rise to cancers. Most cells expressing the transgene, however, do not give rise to cancers. Nevertheless, from the point of view of the whole animal, the inherited oncogene is a serious menace because it creates a high risk that a cancer will arise somewhere in the body. Mice that express both *Myc* and *Ras* oncogenes (bred by mating a transgenic mouse carrying a *Myc* oncogene with one carrying a *Ras* oncogene) develop cancers earlier and at a much higher rate than either parental strain (**Figure 20–36**); but, again, the cancers originate as scattered isolated tumors among noncancerous cells. Thus, even cells expressing these two oncogenes must undergo further, randomly generated changes to become cancerous, strongly suggesting that multiple mutations are required for tumorigenesis, as supported by other evidence and discussed earlier.

Numerous tumor suppressor genes have been knocked out in mice, including *Rb*. As anticipated, many of the mutant strains that are missing one copy of a tumor suppressor gene are cancer prone, and the tumors that develop typically show a loss of heterozygosity—that is, they have lost or inactivated the second copy of the gene. Deletion of both copies of a tumor suppressor gene often leads to death at an embryonic stage, reflecting the essential roles of these genes in normal development. To overcome this problem and test the effect of homozygous mutations in an adult tissue, one can engineer mice to carry conditional mutations (see Figure 5–79), such that only one organ or tissue—say, the liver—displays the genetic defect. Alternatively, mice can be engineered so that the defect is created in response to an experimental signal (a drug) delivered at a chosen time. Through the use of these techniques, mice not only provide insights into the mechanisms of tumor formation, but also models in which to test and develop new cancer therapies.

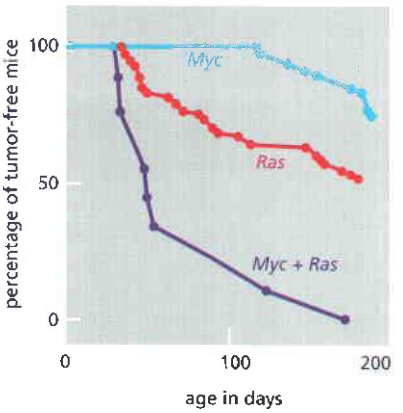


Figure 20–36 Oncogene collaboration in transgenic mice. The graphs show the incidence of tumors in three types of transgenic mouse strains, one carrying a *Myc* oncogene, one carrying a *Ras* oncogene, and one carrying both oncogenes. For these experiments two lines of transgenic mice were first generated. One carries an inserted copy of an oncogene created by fusing the proto-oncogene *Myc* with the mouse mammary tumor virus regulatory DNA (which then drives *Myc* overexpression in the mammary gland). The other line carries an inserted copy of the *Ras* oncogene under control of the same regulatory element. Both strains of mice develop tumors much more frequently than normal, most often in the mammary or salivary glands. Mice that carry both oncogenes together are obtained by crossing the two strains. These hybrids develop tumors at a far higher rate still, much greater than the sum of the rates for the two oncogenes separately. Nevertheless, the tumors arise only after a delay and only from a small proportion of the cells in the tissues where the two genes are expressed. Further accidental changes, in addition to the two oncogenes, are apparently required for the development of cancer. (After E. Sinn et al., *Cell* 49:465–475, 1987. With permission from Elsevier.)

Many Cancer-Critical Genes Regulate Cell Proliferation

Many cancer-critical genes encode components of the pathways that regulate the social behaviors of cells in an animal—in particular, components of the signaling pathways through which influences from neighboring cells control whether a cell grows, divides, differentiates, or dies (**Figure 20–37**). In fact, many of the components of these signaling pathways were first identified through searches for cancer-causing genes, and a full list of the proteins encoded by proto-oncogenes and tumor suppressor genes includes examples of practically every type of protein involved in cell signaling: secreted signal proteins, transmembrane receptors, intracellular GTP-binding proteins, protein kinases, gene regulatory proteins, and many other classes of proteins besides. Cancer mutations alter these components in a way that creates proliferative signals even when more cells are not needed, switching on cell growth, DNA replication, and cell division inappropriately. Examples include the gain-of-function mutations that inappropriately activate the EGF receptor and the Ras protein, discussed earlier.

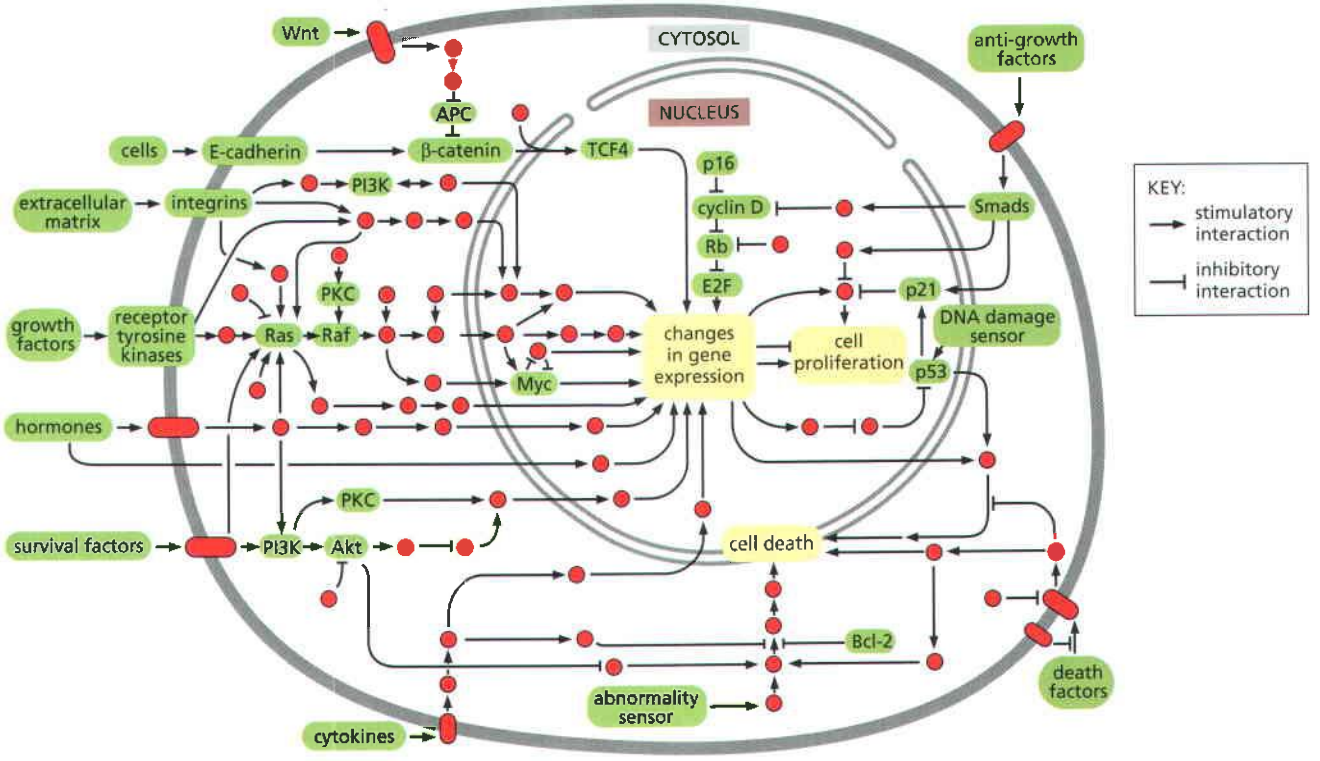


Figure 20–37 Chart of the major signaling pathways relevant to cancer in human cells, indicating the cellular locations of some of the proteins modified by mutation in cancers. Products of both oncogenes and tumor suppressor genes often act within the same pathways. Individual signaling proteins are indicated by *red circles*, with the cancer-critical components and control mechanisms discussed in this chapter in *green*. Stimulatory and inhibitory interactions between proteins are designated as shown in the key. (Adapted from D. Hanahan and R.A. Weinberg, *Cell* 100:57–70, 2000. With permission from Elsevier.)

The components in signaling pathways that normally function to inhibit cell proliferation often appear as tumor suppressors. A well-studied example is the TGFβ signaling pathway (discussed in Chapter 15). Loss-of-function mutations in this pathway contribute to the development of several types of human cancers. The receptor TGFβ-RII, for example, is mutated in some cancers of the colon, as is Smad4—a key intracellular signaling protein in the pathway—which is also often inactivated in cancers of the pancreas and some other organs. These findings reflect the normal function of the TGFβ pathway in restricting cell proliferation (see Figure 23–26).

Not surprisingly, mutations that directly affect the central cell-cycle control system feature prominently in many cancers. The tumor suppressor protein Rb, discussed earlier, controls a key point at which cells decide to enter the cell cycle and replicate their DNA. Rb serves as a brake that restricts entry into S phase by binding to and inhibiting gene regulatory proteins of the E2F family, which are needed to transcribe genes that encode proteins required for entrance into S phase. Normally, this inhibition by Rb is relieved at the appropriate time by the phosphorylation of Rb by various *cyclin-dependent kinases (Cdks)*, which cause Rb to release its inhibitory grip on the E2F proteins (discussed in Chapter 17).

Many cancer cells proliferate inappropriately by eliminating Rb entirely, as we have already seen. Others achieve the same effect by acquiring mutations that alter other components of the Rb regulatory pathway (**Figure 20–38**). In normal cells, a complex of cyclin D and the cyclin-dependent kinase Cdk4 (G_1-Cdk) is responsible for phosphorylating Rb so as to allow progress through the cycle (see Figure 17–62 and Table 17–1, p. 1063). The *p16 (INK4)* protein—which is produced when cells are stressed—inhibits cell-cycle progression by preventing the formation of an active cyclin D–Cdk4 complex; it is an important component of the normal cell-cycle-arrest response to stress. Some glioblastomas and breast cancers have amplified the genes encoding Cdk4 or cyclin D, thus favoring cell proliferation. And deletion or inactivation of the *p16* gene is common in many forms of human cancer. In cancers in which mutations do not inactivate the *p16* gene, DNA methylation of its regulatory region often silences the gene; this is an example of an epigenetic change contributing to the development of cancer. Mutation in any one component of a given pathway is sufficient to inactivate the pathway and promote cancer. As expected, therefore, a

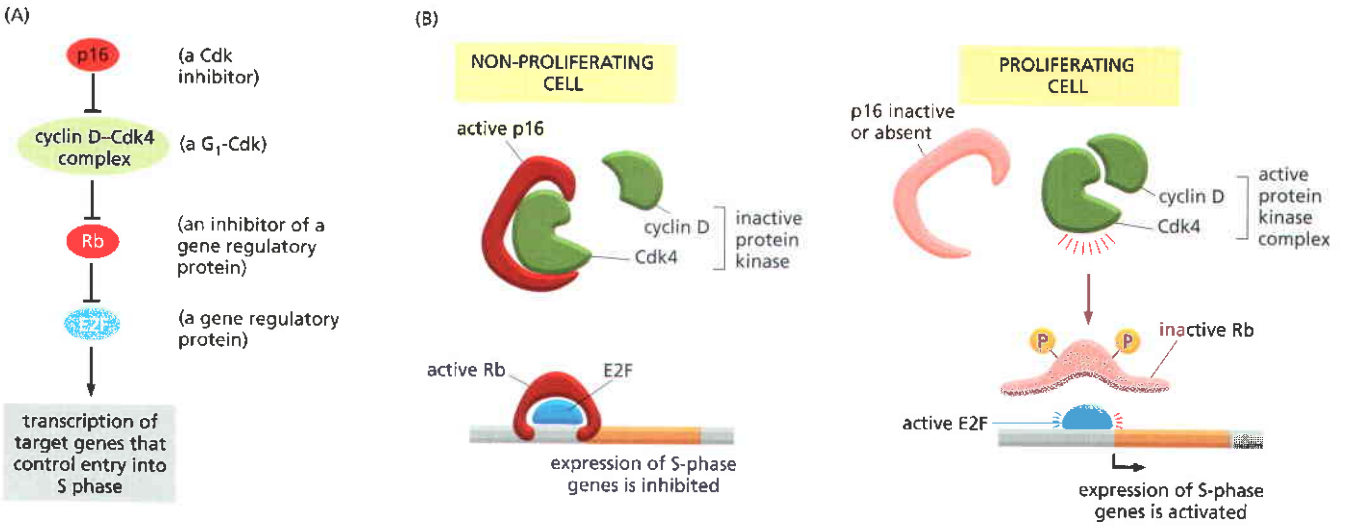


Figure 20–38 The pathway by which Rb controls cell cycle entry contains both proto-oncogenes and tumor suppressor genes. All the components of this pathway have been found to be altered by mutation in human cancers (products of proto-oncogenes, *green;* products of tumor suppressor genes, *red;* E2F is shown in *blue* because it has both inhibitory and stimulatory actions, depending on the other proteins that are bound to it). In most cases, only one of the components is altered in any individual tumor. (A) A simplified view of the dependency relationships in this pathway; see Figure 17–62 for further details. (B) The Rb protein inhibits entry into the cell-division cycle when it is unphosphorylated. The complex of Cdk4 and cyclin D phosphorylates Rb, thereby encouraging cell proliferation. When a cell is stressed, p16 inhibits the formation of an active Cdk4–cyclin D complex, preventing proliferation. Inactivation of Rb or p16 by mutation encourages cell division (thus each can be regarded as a tumor suppressor), while overactivity of Cdk4 or cyclin D encourages cell division (thus each can be regarded as a proto-oncogene).

cancer rarely inactivates more than one component in a pathway: this would bring no additional benefit for the cancer's evolution.

Distinct Pathways May Mediate the Disregulation of Cell-Cycle Progression and the Disregulation of Cell Growth in Cancer Cells

As described in Chapter 17, a special cell-cycle control system ensures that eucaryotic cells make a copy of each of their chromosomes and segregate exactly one copy into each of the two daughter cells created by cell division. The initiation of this process requires cell-cycle progression signals that are carefully regulated, as the default state for each cell in a multicellular organism is to remain in a quiescent G_0 state (see p. 1103). But cell proliferation requires more than progression through the cell cycle; it also requires cell growth, which involves complex anabolic processes through which the cell synthesizes all the necessary macromolecules from small-molecule precursors. If a cell divides inappropriately without growing first, it will get smaller at each division and will ultimately halt or die. The continued growth of a cancer therefore requires heritable changes that not only disregulate cell-cycle progression but also provoke cell growth (**Figure 20–39**).

The phosphoinositide 3-kinase (PI 3-kinase)/Akt intracellular signaling pathway is critical for cell growth control. As described in Chapter 15, various extracellular signal proteins, including insulin and insulin-like growth factors, normally stimulate this pathway. In cancer cells, however, the pathway is activated by mutation so that the cell can grow in the absence of such signals. The resulting abnormal activation of the protein kinase Akt is central to this disregulated growth process, since this activation not only stimulates protein synthesis (see Figure 17–65), but also greatly increases both glucose uptake (via mTOR activation) and the production of the acetyl CoA in the cytosol required for cell lipid synthesis (via ACL activation), as outlined in Figure 20–39. Thus, a common tumor-suppressor gene mutation, in many different cancers, is loss of the PTEN phosphatase, whose normal function is to limit Akt activation by dephosphorylating the molecules that PI 3-kinase phosphorylates.

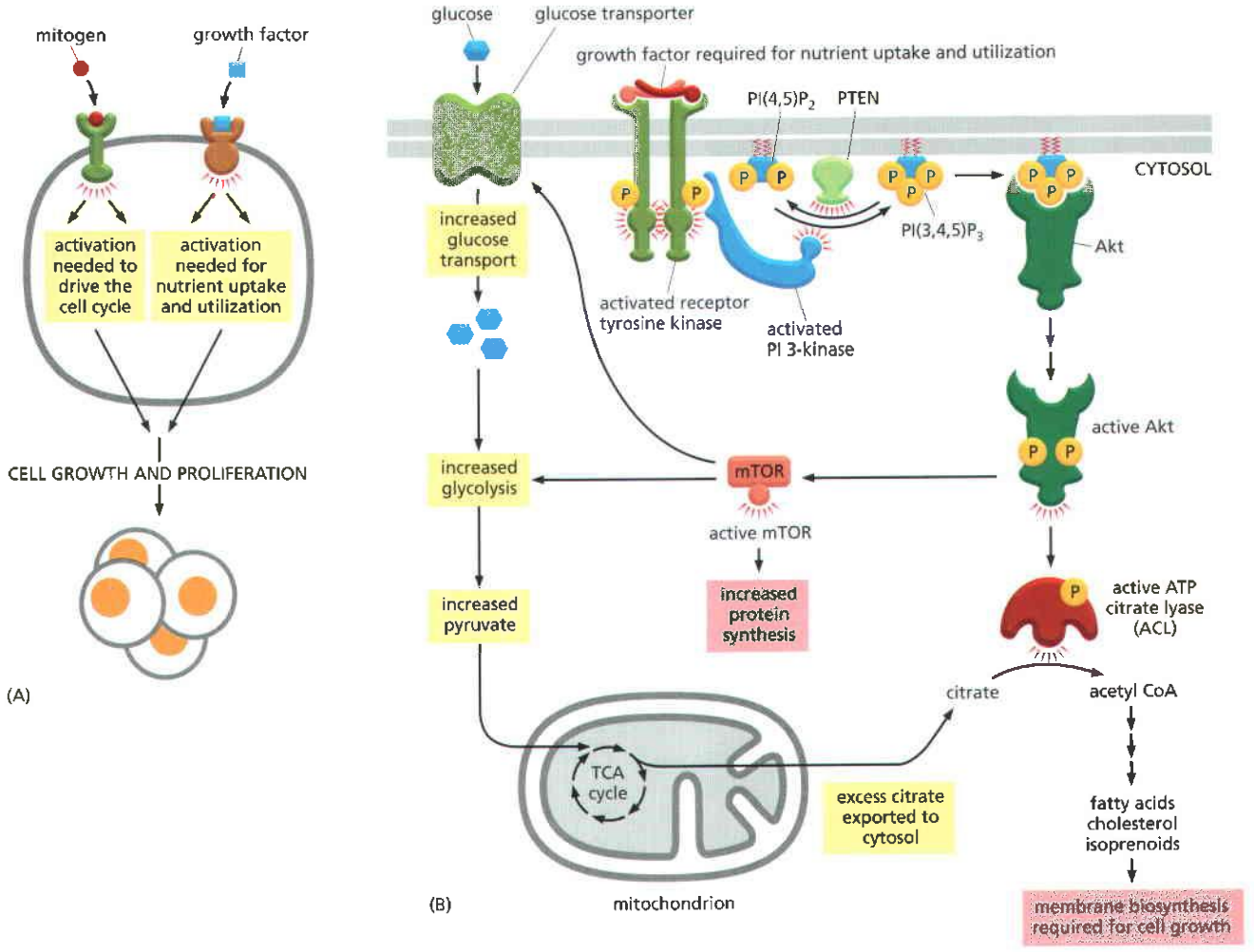


Figure 20–39 Cells may require two types of signals to proliferate. (A) In order to multiply successfully, most normal cells are suspected to require both extracellular signals that drive cell cycle progression (here *red* mitogen) and extracellular signals that drive cell growth (here *blue* growth factor). (B) Diagram of the signaling system containing Akt that drives cell growth through greatly stimulating glucose uptake and utilization, including a conversion of the excess citric acid produced from sugar intermediates in mitochondria into the acetyl CoA that is needed in the cytosol for lipid synthesis and new membrane production. As indicated, protein synthesis is also increased. This system becomes abnormally activated early in tumor progression.

The abnormal activation of the PI 3-kinase/Akt pathway, which normally occurs early in the process of tumor progression, explains the excessive rate of glycolysis that is observed in tumor cells, known as the Warburg effect. Accompanied by excretion of excess pyruvate as lactate, the excessive glucose uptake by cancer cells is used to locate tumors by modern whole-body imaging techniques (see Figure 20–1).

Mutations in Genes That Regulate Apoptosis Allow Cancer Cells to Survive When They Should Not

Control of cell numbers depends on maintaining a balance between cell proliferation and cell death. In the germinal centers of lymph nodes, for example, B cells proliferate rapidly, but most of their progeny are eliminated by apoptosis. Correct regulation of apoptosis is thus essential in maintaining the normal balance of cell birth and death in tissues that undergo cell turnover. It also has a vital role in eliminating damaged or stressed cells. As described in Chapter 18, animal cells commit suicide by undergoing apoptosis when they sense that something has gone drastically wrong—when their DNA is severely damaged, for example, or when they are deprived of extracellular survival signals that tell them they are in their proper place. As previously discussed (see Figure 20–14), cancer cells are relatively resistant to apoptosis, which allows them to increase in number and survive where they should not.

Mutations in apoptosis-control genes usually account for this resistance. One protein that normally inhibits apoptosis, called Bcl2, was discovered and named because its expression is activated by a chromosomal translocation in a B-cell lymphoma. The translocation places the *Bcl2* gene under the control of a regulatory DNA sequence that drives Bcl2 overexpression. This allows the survival of B

lymphocytes that would normally have died, greatly increasing the number of B cells and contributing to the development of the B-cell cancer.

One of the genes involved in the control of apoptosis is mutated in an exceptionally wide range of cancers. This tumor suppressor gene encodes a protein that stands at a crucial intersection in the network of intracellular control pathways governing a cell's responses to DNA damage and various other cell stresses, including low oxygen (hypoxia) and growth factor deprivation. The protein is tumor protein p53, and it is produced by the *Tp53* gene, more commonly known as *p53*. As we now explain, when *p53* is defective, genetically damaged dividing cells do not merely fail to die; they persist in proliferating, accumulating yet more genetic damage that can increase malignancy.

Mutations in the *p53* Gene Allow Many Cancer Cells to Survive and Proliferate Despite DNA Damage

The ***p53*** gene—named for the molecular mass of its protein product—may be the most important gene preventing human cancer. Either this tumor suppressor gene or other components in the p53 pathway are mutated in nearly all human cancers. Why is *p53* so critical? The answer lies in its multifaceted function in cell-cycle control, in apoptosis, and in maintenance of genetic stability—all aspects of the fundamental role of the p53 protein in protecting the organism against the consequences of cell damage and the risk of cancer.

In contrast to Rb, most body cells have very little p53 protein under normal conditions: although the protein is synthesized, it is rapidly degraded. Moreover, p53 is not essential for normal development. Mice in which both copies of the gene have been deleted or inactivated typically appear normal in all respects except one—they universally develop cancer before 10 months of age. These observations suggest that p53 may have a function that is required only in special circumstances. Indeed, when normal cells are deprived of oxygen or exposed to treatments that damage DNA, such as ultraviolet light or gamma rays, they raise their concentration of p53 protein by blocking the degradation of the molecule. The accumulation of p53 protein is seen also in cells where oncogenes such as *Ras* and *Myc* are unusually active, generating an abnormal stimulus for cell division.

In all these cases, the high level of p53 protein acts to limit the harm done. Depending on circumstances and the severity of the damage, p53 may either drive the damaged or abnormally proliferating cell to commit suicide by apoptosis—a relatively harmless event for the multicellular organism—or may trigger a mechanism that bars the cell from dividing so long as the damage remains unrepaired. Similarly, when telomeres get too short, p53 is activated and inhibits further cell division, producing the phenomenon of replicative cell senescence (**Figure 20–40**). The protection that p53 provides is part of the reason why mutations that activate oncogenes such as *Ras* and *Myc* are not sufficient to create a tumor.

The p53 protein performs its job mainly by acting as a gene regulatory protein. <TGAA> Indeed, the most common mutations observed in p53 in human tumors are in its DNA-binding domain, where they cripple the ability of p53 to bind to its DNA target sequences. As discussed in Chapter 17, the p53 protein

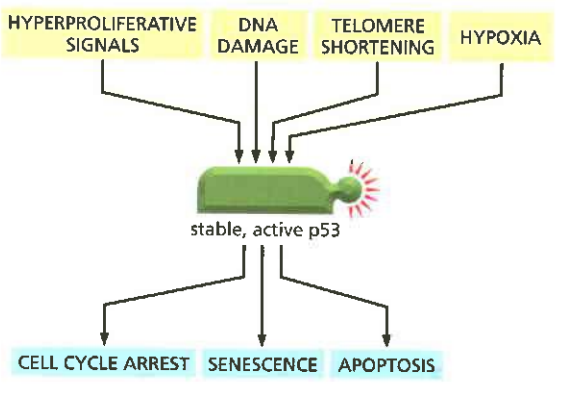


Figure 20–40 Modes of action of the p53 tumor suppressor. The p53 protein is a cellular stress sensor. In response to hyperproliferative signals, DNA damage, hypoxia, and/or telomere shortening, the p53 levels in the cell rise and cause cells to undergo cell cycle arrest, apoptosis, or replicative cell senescence. (As discussed in Chapter 17, a senescent cell progressively loses the ability to divide.) All of these outcomes keep tumor growth in check.

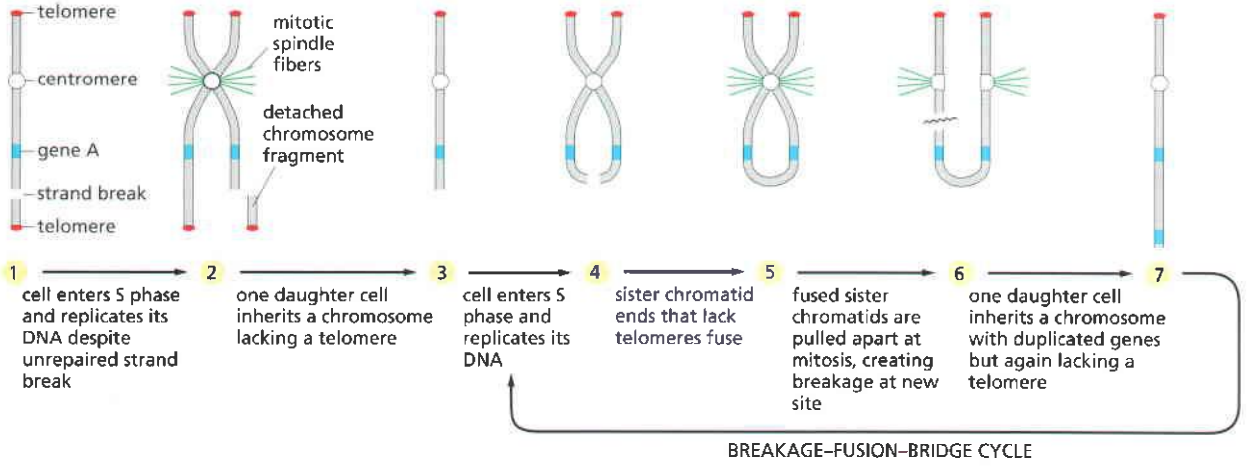


Figure 20–41 How the replication of damaged DNA can lead to chromosome abnormalities, such as gene amplification and gene loss. The diagram shows one of several possible mechanisms. The process begins with accidental DNA damage in a cell that lacks functional p53 protein. Instead of halting at the p53-dependent checkpoint in the division cycle, where a normal cell with damaged DNA would pause until it repaired the damage, the p53-defective cell enters S phase, with the consequences shown. Once a chromosome carrying a duplication and lacking a telomere has been generated, repeated rounds of replication, chromatid fusion, and unequal breakage (the so-called breakage–fusion–bridge cycle) can increase the number of copies of the duplicated region still further. Selection in favor of cells with increased numbers of copies of a gene in the affected chromosomal region will thus lead to mutants in which the gene is amplified to a high copy number. The multiple copies may—either through a recombination event or through unrepaired DNA strand breakage—become excised from their original locus and so appear as independent double minute chromosomes (see Figure 20–34). The chromosomal disorder can also lead to a loss of genes, with selection in favor of cells that have lost tumor suppressors.

exerts its inhibitory effects on the cell cycle, in part at least, by binding to DNA to induce the transcription of *p21*, which encodes a protein that binds to and inhibits the Cdk complexes required for progression through the cell cycle. By blocking the kinase activity of these Cdk complexes, the p21 protein prevents the cell from entering S phase and replicating its DNA.

The mechanisms by which p53 induces apoptosis are more complex; it can stimulate the expression of many pro-apoptotic genes, but it may also bind to and inactivate anti-apoptotic Bcl2 proteins on the surface of mitochondria, thereby promoting apoptosis (discussed in Chapter 18).

Cells defective in *p53* fail to show these p53-dependent stress responses. They tend to escape apoptosis, and, if their DNA is damaged, they may carry on dividing, plunging into DNA replication without pausing to repair the damage. As a result, they may either die or, far worse, survive and proliferate with a corrupted genome (**Figure 20–41**). Such genetic mayhem can lead both to loss of tumor suppressor genes and to activation of oncogenes. The loss of p53 also makes some cancer cells much less sensitive to irradiation and to many anticancer drugs, which would otherwise cause them to die or to halt their proliferation.

In summary, p53 helps a multicellular organism cope safely with DNA damage and other cell stresses, acting as a check on cell proliferation in circumstances where it would be dangerous. The loss of *p53* activity is unusually dangerous because it can promote cancer for four different reasons. First, it allows cells with DNA damage to continue through the cell cycle. Second, it allows them to escape apoptosis. Third, by permitting division of cells with damaged chromosomes, it leads to the genetic instability characteristic of cancer cells, allowing further cancer-promoting mutations to accumulate as they divide. Fourth, it makes cells relatively resistant to anticancer drugs and irradiation in some tumor types. Many other mutations can contribute to one or another of these types of misbehavior, but p53 mutations contribute to them all.

DNA Tumor Viruses Block the Action of Key Tumor Suppressor Proteins

DNA tumor viruses cause cancer mainly by interfering with the cell's controls on the cell cycle and apoptosis, including those that depend on p53. To understand this type of viral carcinogenesis, it is important to understand the life history of viruses. Many DNA viruses use the host cell's DNA replication machinery to replicate their own genomes. To produce numerous infectious virus particles within a single host cell, the DNA virus has to commandeer this machinery and drive it hard, breaking through the normal constraints on DNA replication and usually killing the host cell in the process. Many DNA viruses reproduce only in this way. But some of them have a second option: they can propagate their genome as a quiet, well-behaved passenger in the host cell, replicating in parallel with the host cell's DNA in the course of ordinary cell division cycles. These viruses can switch between two modes of existence according to circumstances,

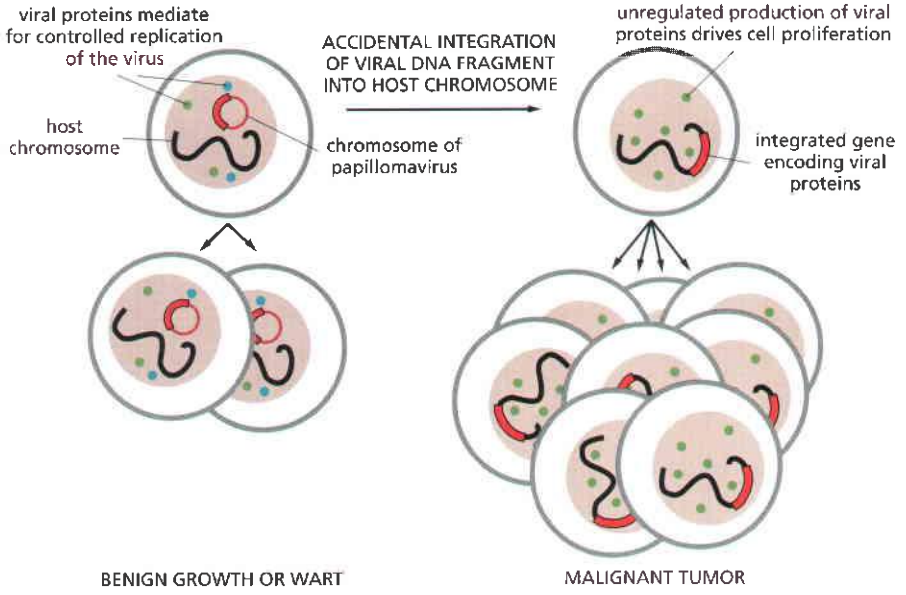


Figure 20–42 How certain papillomaviruses are thought to give rise to cancer of the uterine cervix. Papillomaviruses have double-stranded circular DNA chromosomes of about 8000 nucleotide pairs. These chromosomes are normally stably maintained in the basal cells of the epithelium as plasmids whose replication is regulated so as to keep step with the chromosomes of the host *(left)*. Rare accidents can cause the integration of a fragment of such a plasmid into a chromosome of the host, altering the environment of the viral genes in the basal cells. This (or possibly some other cause) disrupts the normal control of viral gene expression. The unregulated production of viral proteins interferes with the control of cell division in the basal cells, thereby helping to generate a cancer *(right)*.

remaining latent and harmless for a long time, but then proliferating in occasional cells in a process that kills the host cell and generates large numbers of infectious particles.

Those DNA viruses that have the ability to replicate their DNA as quiescent passengers in a latent phase of viral replication either integrate their viral genome into one or more host chromosomes, or they have the ability to form an extrachromosomal plasmid that replicates in step with the host chromosomes during the latent phase.

No matter which way of life a DNA virus is following, it is not in its interest to kill the host organism. But for viruses with a latent phase, accidents can occur that prematurely activate some of the viral proteins that the virus would normally use in its replicative phase to allow the viral DNA to replicate independently of the cell cycle. This type of accident can switch on the persistent proliferation of the host cell, leading to cancer, and the DNA viruses susceptible to such accidents have been designated as **DNA tumor viruses**.

The **papillomaviruses**, for example, are the cause of human warts and are especially important as a key causative factor in carcinomas of the uterine cervix (about 6% of all human cancers). Papillomaviruses infect the cervical epithelium and maintain themselves in a latent phase in the basal layer of cells as extrachromosomal plasmids, which replicate in step with the chromosomes. Infectious virus particles are generated through a switch to a replicative phase in the outer epithelial layers, as progeny of these cells begin to differentiate before being sloughed from the surface. Here, cell division should normally stop, but the virus interferes with this cell-cycle arrest so as to allow replication of its own genome. Usually, the effect is restricted to the outer layers of cells and is relatively harmless, as in a wart. Occasionally, however, a genetic accident causes the viral genes that encode the proteins that prevent cell-cycle arrest to integrate into the host chromosome and become active in the basal layer, where the stem cells of the epithelium reside. This can lead to cancer, with the viral genes acting as oncogenes (**Figure 20–42**).

In papillomaviruses, the viral genes that are mainly to blame are called *E6* and *E7*. The protein products of these viral oncogenes interact with many host cell proteins, but, in particular, they bind to two key tumor suppressor proteins of the host cell, putting them out of action and so permitting the cell to replicate its DNA and divide in an uncontrolled way. One of these host proteins is Rb: by binding to Rb, the viral E7 protein prevents Rb from binding to and inhibiting the E2F proteins mentioned earlier, which allows uncontrolled entry into S phase. The other is p53; by binding to p53, the viral E6 protein triggers p53 destruction (**Figure 20–43**), allowing the abnormal cell to survive, divide, and accumulate yet more abnormalities. Other DNA tumor viruses use similar mechanisms to inhibit Rb and p53, underlining the importance of inactivating

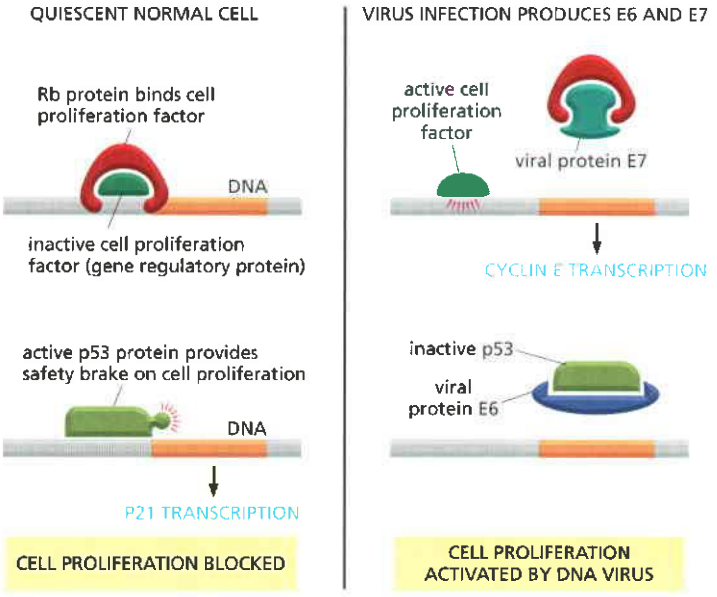


Figure 20–43 Activation of cell proliferation by a DNA tumor virus. Papillomavirus produces two viral proteins, E6 and E7, which sequester the host cell's p53 and Rb proteins, respectively. E6 protein binding leads to ubiquitylation of its p53 partner, inducing p53 proteolysis (not shown). The SV40 virus (a related virus that infects monkeys) uses a single dual-purpose protein, called large T antigen, to sequester these same two proteins.

both of these tumor suppressor pathways for releasing the normal constraints on cell proliferation. It is striking that virtually all human cancers have also inactivated both pathways, usually affecting only one component in each.

The Changes in Tumor Cells That Lead to Metastasis Are Still Largely a Mystery

Perhaps the most significant gap in our understanding of cancer concerns invasiveness and metastasis. We have yet to clearly identify mutations that specifically permit cells to invade surrounding tissues, spread through the body, and form metastases. Indeed, it is not even clear exactly what properties a cancer cell must acquire to become metastatic. One extreme view is that the ability of cancer cells to metastasize requires no further genetic changes beyond those needed to weaken the normal controls on cell growth, cell division, and cell death. Some experiments using DNA microarrays to compare the mRNAs expressed in a metastasis and in the primary tumor support this view. The alternative view, more widely held, is that metastasis is a difficult and complex task that requires additional mutations, but that these may be so varied according to circumstance that they are hard to discover individually.

One molecular change in metastatic cancer cells from a variety of tumors—including melanomas and carcinomas of breast, stomach, and liver—is an overexpression of the Rho-family GTPase *RhoC*, which helps to mediate actin-based cell motility. In some cases, inhibition of RhoC activity in these cells is sufficient to block their metastatic ability. It is not yet clear, however, exactly how RhoC contributes to the ability to metastasize.

Metastasis presents different problems for different types of cells. For a cancerous white blood cell already roaming the body via the circulation, metastasis should be easier than for a carcinoma cell that has to escape from an epithelium. As we discussed earlier, it is helpful to distinguish two phases of tumor progression required for a carcinoma to metastasize. First, the cells must escape the normal confines of their parent epithelium and begin to invade the tissue immediately beneath. Second, they must travel to distant sites via the blood or lymph and establish *metastases*.

The first step, local invasiveness, requires a loss of the mechanisms that normally hold epithelial cells together. As mentioned earlier, this step resembles the normal developmental process known as the *epithelial-to-mesenchymal transition (EMT)*, in which epithelial cells undergo a shift in character, becoming less adhesive and more migratory (discussed in Chapter 19). At the heart of this transition are changes in expression of the *E-cadherin* gene. The primary function of the transmembrane E-cadherin protein is in cell–cell adhesion, and it is especially

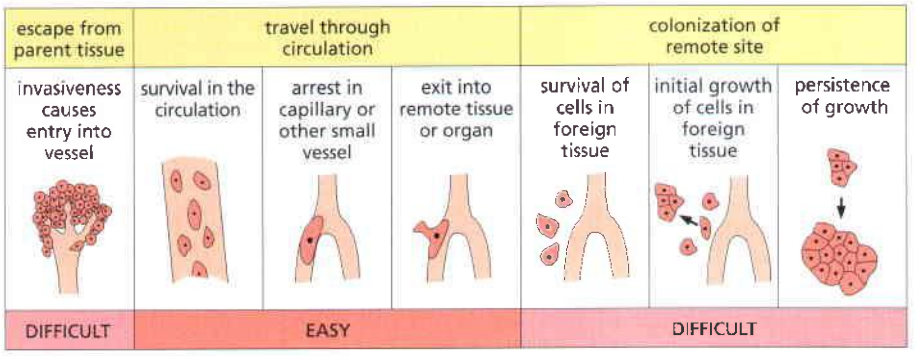


Figure 20–44 The barriers to metastasis. Studies of labeled tumor cells leaving a tumor site, entering the circulation, and establishing metastases show which steps in the metastatic process, outlined in Figure 20–17, are difficult or 'inefficient,' in the sense that they are steps in which large numbers of cells fail and are lost. It is in these difficult steps that highly metastatic cells are observed to have much greater success than nonmetastatic cells. It seems that the ability to escape from the parent tissue, and an ability to survive and grow in the foreign tissue, are key properties that cells must acquire to become metastatic. (Adapted from A.F. Chambers et al., *Breast Cancer Res.* 2:400–407, 2000. With permission from BioMed Central Ltd.)

important in binding epithelial cells together through adherens junctions (see Figure 19–15). In some carcinomas of the stomach and of the breast, the *E-cadherin* gene has been identified as a tumor suppressor gene. The abnormal behavior of tumor cells lacking the protein can be partially rescued, at least in culture: if a functional *E-cadherin* gene is put back into these cancer cells, they lose some of their invasive characteristics and begin to cohere more like normal epithelial cells. Loss of *E-cadherin*, therefore, may promote cancer development by facilitating local invasiveness.

The initial entry of tumor cells into the circulation is helped by the presence of a dense supply of blood vessels and sometimes lymphatic vessels, which tumors attract to themselves by secreting angiogenic factors such as VEGF, as discussed earlier; the abnormal fragility and leakiness of these new vessels may help the cells that have become invasive to enter and then move through the circulation with relative ease. The remaining steps in metastasis, involving exit from a blood or lymphatic vessel and the effective colonization of remote sites, are much harder to study. To discover which of the later steps in metastasis present cancer cells with the greatest difficulties, one can label the cells with a fluorescent dye or green fluorescent protein (GFP), inject them into the bloodstream of a mouse, and then monitor their fate. In such experiments one observes that many cells survive in the circulation, lodge in small vessels, and exit into the surrounding tissue, regardless of whether they come from a tumor that metastasizes or one that does not. Some cells die immediately after they enter foreign tissue; others survive entry into the foreign tissue but fail to proliferate; still others divide a few times and then stop, forming micrometastases containing ten to several thousand cells. At this point, the metastasis-competent cells outperform their nonmetastatic relatives, as only they form growing tumors, raising the possibility that the ability to continue to proliferate in a foreign tissue (the process called *colonization*) is a key property that cancer cells must acquire to become metastatic (**Figure 20–44**).

The challenge is to use assays such as this to discover at a molecular level what, if anything, distinguishes the cells that colonize from those that do not. One possibility, as discussed earlier, is that the cells capable of colonization are just those rare members of the parent tumor population that have the special character of cancer stem cells. It seems that this must be at least part of the story, as we have already explained. But although we know of markers that help identify cancer stem cells, at least for some cancers, we do not yet understand what it is that fundamentally gives them their special stem-cell properties. This is part of the wider problem of understanding stem cells in general, discussed at length in Chapter 23. In any case, it remains very hard to pinpoint what additional genetic or epigenetic changes, if any, are needed to enable cancer stem cells to colonize foreign tissue.

Colorectal Cancers Evolve Slowly Via a Succession of Visible Changes

At the beginning of this chapter, we saw that most cancers develop gradually from a single aberrant cell, progressing from benign to malignant tumors by the

accumulation of a number of independent genetic accidents and epigenetic changes. We have discussed what some of these accidents and changes are in molecular terms and seen how they contribute to cancerous behavior. We now examine one of the common human cancers more closely, using it to illustrate and enlarge upon some of the general principles and molecular mechanisms we have introduced, attempting to make sense of the natural history of the disease in these terms. We take **colorectal cancer** as our example, in which the steps of tumor progression have been followed *in vivo* with a colonoscope (a fiber-optic device for viewing the interior of the colon and rectum) and carefully studied at the molecular level. This model, based on studies of both sporadic and inherited forms of the human disease, serves as a paradigm for illustrating the stepwise nature of tumor development, and it allows us to associate specific morphological stages with inherited changes in the function of particular genes.

Colorectal cancers arise from the epithelium lining the colon (the large intestine) and rectum (the terminal segment of the gut). The organization of this tissue is broadly similar to that of the small intestine, discussed in detail in Chapter 23 (pp. 1436–1442). Both regions of gut epithelium are renewed at an extraordinarily rapid rate, taking about a week to completely replace most of the epithelial sheet. In both regions the renewal depends on stem cells that lie in deep pockets of the epithelium, called intestinal crypts. The signals that maintain the stem cells and control the normal organization and renewal of the epithelium are beginning to be quite well understood, as explained in Chapter 23. This is in large part thanks to insights that have come from cancer research. Mutations that disrupt the normal organizing signals begin the process of tumor progression for most colorectal cancers.

Colorectal cancers are common, currently causing nearly 60,000 deaths a year in the United States, or about 10% of total deaths from cancer. Like most cancers, they are not usually diagnosed until late in life (90% occur after the age of 55). However, routine examination of normal adults with a colonoscope often reveals a small benign tumor, or adenoma, of the gut epithelium in the form of a protruding mass of tissue called a *polyp* (**Figure 20–45A**; see also Figure 23–23). These adenomatous polyps are believed to be the precursors of a large proportion of colorectal cancers. Because the progression of the disease is usually very slow, there is typically a period of about 10 years in which the slowly growing tumor is detectable but has not yet turned malignant. Thus, when people are screened by colonoscopy in their fifties and the polyps are removed through the colonoscope—a quick and easy surgical procedure—the subsequent incidence of colorectal cancer is very low, according to some studies, less than a quarter of what it would be otherwise.

Colon cancer provides a clear example of the phenomenon of tumor progression discussed previously. In microscopic sections of polyps smaller than 1 cm in diameter, the cells and their arrangement in the epithelium usually appear almost normal. The larger the polyp, the more likely it is to contain cells that look abnormally undifferentiated and form abnormally organized structures. Sometimes, two or more distinct areas can be distinguished within a single polyp, with the cells in one area appearing relatively normal and those in the other appearing clearly cancerous, as though they have arisen as a mutant subclone within the original clone of adenomatous cells. At later stages in the disease, the tumor cells become invasive, first breaking through the epithelial basal lamina, then spreading through the layer of muscle that surrounds the gut (Figure 20–45B), and finally metastasizing to lymph nodes via lymphatic vessels and to liver, lung, and other organs via blood vessels.

A Few Key Genetic Lesions Are Common to a Large Fraction of Colorectal Cancers

What are the mutations that accumulate with time to produce this chain of events? Of those genes so far discovered to be involved in colorectal cancer, three stand out as most frequently mutated—the proto-oncogene *K-Ras* (a member of the *Ras* gene family) and the tumor suppressor genes *p53* and *Apc*

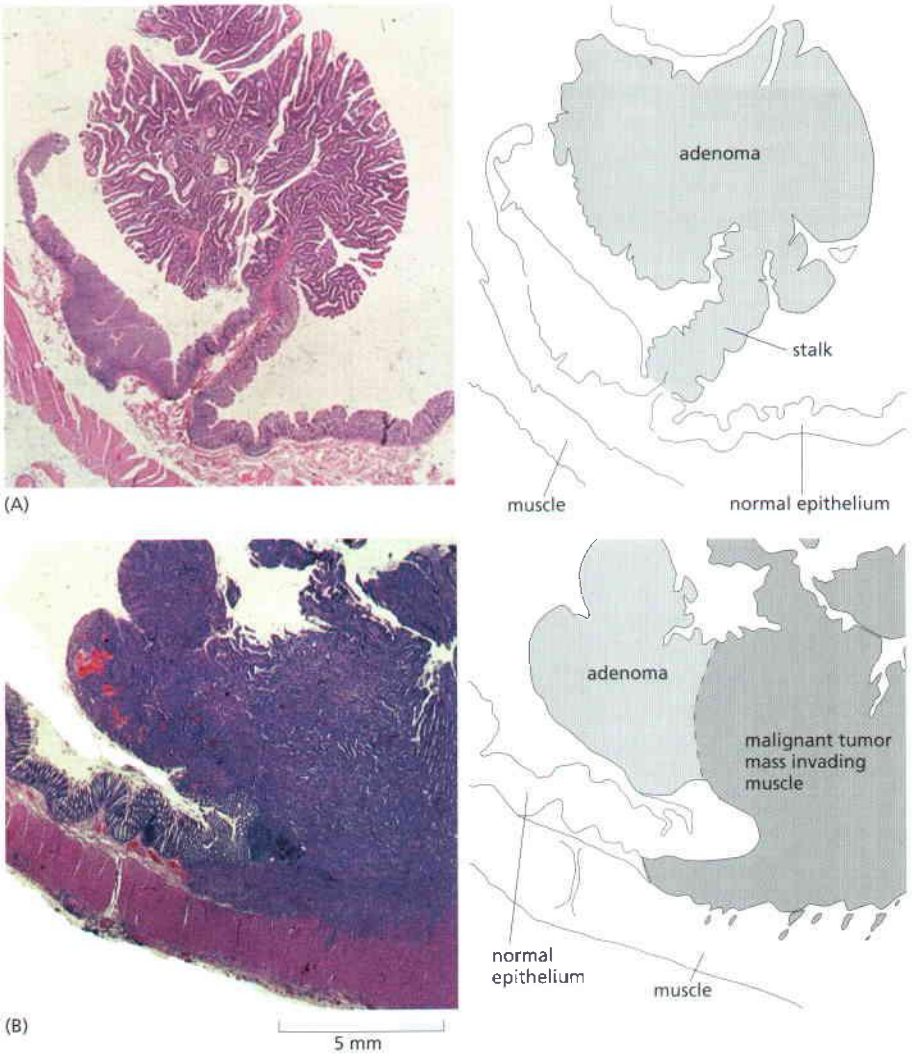


Figure 20–45 Cross sections showing the stages in development of a typical colon cancer. (A) An adenomatous polyp from the colon. The polyp protrudes into the lumen—the space inside the colon. The rest of the wall of the colon is covered with normal colonic epithelium; the epithelium on the polyp appears mildly abnormal. (B) A carcinoma that is beginning to invade the underlying muscle layer. (Courtesy of Paul Edwards.)

(discussed below). Others are involved in smaller numbers of colon cancers (**Table 20–2**). Still others remain to be identified.

One approach to discovery of the mutations responsible for colorectal cancer is to screen the cells for abnormalities in genes already known or suspected to be involved in cancers elsewhere. This type of genetic screening has revealed that about 40% of colorectal cancers have a specific point mutation in *K-Ras*, activating it to become an oncogene, and about 60% have inactivating mutations or deletions of *p53*.

As discussed earlier, another approach is to track down the genetic defects in those rare families that show a hereditary predisposition to colorectal cancer.

Table 20–2 **Some Genetic Abnormalities Detected in Colorectal Cancer Cells**

GENE	CLASS	PATHWAY AFFECTED	HUMAN COLON CANCERS (%)
K-Ras	oncogene	receptor tyrosine-kinase signaling	40
β-catenin [1]	oncogene	Wnt signaling	5–10
Apc [1]	tumor suppressor	Wnt signaling	> 80
p53	tumor suppressor	response to stress and DNA damage	60
TGFβ receptor II [2]	tumor suppressor	TGFβ signaling	10
Smad4 [2]	tumor suppressor	TGFβ signaling	30
MLH1 and other DNA mismatch repair genes	tumor suppressor (genetic stability)	DNA mismatch repair	15 (often silenced by methylation)

The genes with the same superscript act in the same pathway, and therefore only one of the components is mutated in an individual cancer.

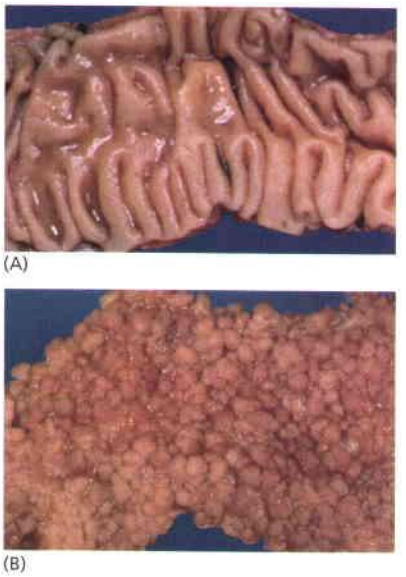


Figure 20–46 Colon of familial adenomatous polyposis coli patient compared with normal colon. (A) The normal colon wall is a gently undulating but smooth surface. (B) The polyposis colon is completely covered by hundreds of projecting polyps (shown in section in Figure 20–45), each resembling a tiny cauliflower when viewed with the naked eye. (Courtesy of Andrew Wyllie and Mark Arends.)

The first hereditary colorectal cancer syndrome to be analyzed was *familial adenomatous polyposis coli (FAP)*, in which hundreds or thousands of polyps develop along the length of the colon (**Figure 20–46**). These polyps start to appear in early adult life, and if they are not removed, one or more will almost always progress to become malignant; the average time from the first detection of polyps to the diagnosis of cancer is 12 years. The disease can be traced to a deletion or inactivation of the tumor suppressor gene *Apc*, named after the syndrome. Individuals with FAP have inactivating mutations or deletions of one copy of the *Apc* gene in all their cells and show loss of heterozygosity in tumors, even in the benign polyps. Most patients with colorectal cancer do not have the hereditary condition. But more than 80% of their cancers (but not their normal cells) have inactivated both copies of the *Apc* gene through mutations acquired during their lifetime. Thus, by a route similar to that we discussed for retinoblastoma, mutation of the *Apc* gene has been identified as one of the central ingredients of colorectal cancer.

As explained earlier, even where there is no known hereditary syndrome, tumor suppressor genes can be tracked down by searching for genetic changes acquired during the evolution of the tumor cells. A systematic scan of a large number of colorectal cancers reveals that selected regions of certain chromosomes frequently have precisely the same DNA sequences on both the maternally and paternally inherited chromosomes. This loss of heterozygosity suggests that those regions may harbor tumor suppressor genes. One of them is the region that includes *Apc*. Another includes the *Smad4* gene, which is mutated in about 30% of colon cancers; its removal blocks the growth-inhibitory effects of the TGFβ pathway. Specific parts of other chromosomes also show frequent losses or gains in colorectal cancers and are likely to be fruitful targets for finding additional cancer-critical genes.

As our knowledge of genes and their functions has expanded, another useful approach has been to look at genes that encode proteins that interact with the protein encoded by a known cancer-critical gene, in the expectation that these genes, too, may be targets for mutation. The Apc protein is now known to be an inhibitory component of the *Wnt signaling pathway* (discussed in Chapter 15). It binds to the *β-catenin* protein, another component of the Wnt pathway, and helps to induce its degradation. By inhibiting β-catenin in this way, APC prevents the β-catenin from migrating to the nucleus, where it acts as a transcriptional regulator to maintain the stem-cell state, as discussed in Chapter 23. Loss of Apc results in an excess of free β-catenin and thus to an uncontrolled expansion of the stem-cell population and a massive increase in the number and size of the intestinal crypts (see Figure 23–23).

When the *β-catenin* gene was sequenced in a collection of colorectal tumors, among the few tumors that did not have *Apc* mutations, a high proportion had activating mutations in *β-catenin* instead. Thus, it is excessive activity in the Wnt signaling pathway that is critical for the initiation of this cancer, rather than any single oncogene or tumor suppressor gene that the pathway contains. This is in accord with the general principle that an individual cancer rarely acquires mutations in more than one component in a pathway that is critical to tumorigenesis, as there is no competitive advantage for it to do so.

The being so, why is the *Apc* gene in particular so much the most common culprit in colorectal cancer? The Apc protein is large and it interacts not only with the β-catenin but also with various other cell components, including microtubules. Loss of Apc appears to increase the frequency of mitotic spindle defects, leading to chromosome abnormalities when cells divide. This additional, independent cancer-promoting effect could explain why *Apc* mutations feature so prominently in the causation of colorectal cancer.

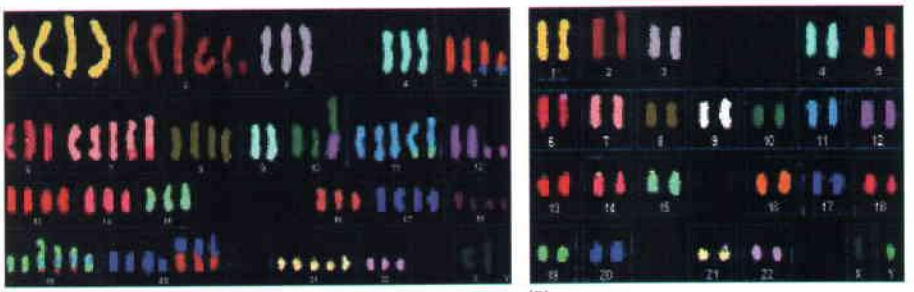


Figure 20–47 Chromosome complements (karyotypes) of colon cancers showing different kinds of genetic instability. (A) The karyotype of a typical cancer shows many gross abnormalities in chromosome number and structure. Considerable variation can also exist from cell to cell (not shown). (B) The karyotype of a tumor that has a stable chromosome complement with few chromosomal anomalies. Its defects are mostly invisible, having been created by defects in DNA mismatch repair. All of the chromosomes in this figure were stained as in Figure 20–13, the DNA of each human chromosome being stained with a different combination of fluorescent dyes. (Courtesy Wael Abdel-Rahman and Paul Edwards.)

Some Colorectal Cancers Have Defects in DNA Mismatch Repair

In addition to the hereditary disease (FAP) associated with *Apc* mutations, there is a second, more common, kind of hereditary predisposition to colon carcinoma in which the course of events differs from the one we have described for FAP. In this condition, called *hereditary nonpolyposis colorectal cancer (HNPCC)*, the probability of colon cancer is increased without any increase in the number of colorectal polyps (adenomas). Moreover, the cancer cells are unusual, in that they have a normal (or almost normal) karyotype; the majority of colorectal tumors in non-HNPCC patients, in contrast, have gross chromosomal abnormalities, with multiple translocations, deletions, and other aberrations, as well as having many more chromosomes than normal (**Figure 20–47**).

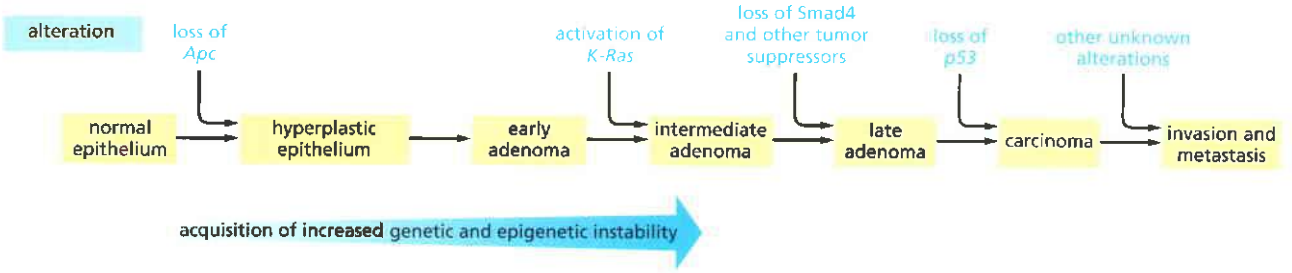
The mutations that predispose HNPCC individuals to colorectal cancer occur in one of several genes that code for central components of the DNA *mismatch repair system*, which are homologous in structure and function to the *MutL* and *MutS* genes in bacteria and yeast (see Figure 5–20). Only one of the two copies of the involved gene is defective, so the repair system is still able to remove the inevitable DNA replication errors that occur in the patient's cells. However, as discussed previously, these individuals are at risk, because the accidental loss or inactivation of the remaining good gene copy will immediately elevate the spontaneous mutation rate by a hundredfold or more (discussed in Chapter 5). These genetically unstable cells now can presumably speed through the standard processes of mutation and natural selection that allow clones of cells to progress to malignancy.

This particular type of genetic instability produces invisible changes in the chromosomes—most notably changes in individual nucleotides and short expansions and contractions of mono- and dinucleotide repeats such as AAAA… or CACACA…. Once the defect in HNPCC patients was recognized, mutations in mismatch repair genes were found in about 15% of the colorectal cancers occurring in people with no inherited predisposing mutation.

Thus, the genetic instability found in many colorectal cancers can be acquired in at least two ways. The majority of the cancers display a form of chromosomal instability that leads to visibly altered chromosomes, whereas in the others the instability occurs on a much smaller scale and reflects a defect in DNA mismatch repair. Indeed, many carcinomas show either chromosomal instability or defective mismatch repair—but rarely both. These findings clearly demonstrate that genetic instability is not an accidental byproduct of malignant behavior but a contributory cause—and that cancer cells can acquire this instability in multiple ways.

The Steps of Tumor Progression Can Often Be Correlated with Specific Mutations

In what order do *K-Ras*, *p53*, *Apc*, and the other identified colorectal cancer-critical genes mutate, and what contribution does each of them make to the asocial behavior of the cancer cell? There is no single answer, because colorectal cancer can arise by more than one route: thus, we know that in some cases, the first mutation can be in a DNA mismatch repair gene; in others, it can be in a gene regulating cell proliferation. Moreover, as previously discussed, a general feature



such as genetic instability or a tendency to proliferate abnormally can arise in a variety of ways, through mutations in different genes.

Nevertheless, certain sets of mutations are particularly common in colorectal cancer, and they occur in a characteristic order. Thus, in most cases, mutations inactivating the *Apc* gene appear to be the first, or at least a very early step, as they are detected at the same high frequency in small benign polyps as in large malignant tumors. Inherited changes that lead to genetic and epigenetic instability are likely also to arise early in tumor progression, since they are needed to drive the later steps.

Activating mutations in the *K-Ras* gene occur later, as they are rare in small polyps but common in larger ones that show disturbances in cell differentiation and histological pattern. Cultured colorectal carcinoma cells at this stage show the typical features of transformed cells, such as the ability to proliferate without anchorage to a substratum. Loss of the Smad4 tumor suppressor gene and inactivating mutations in *p53* come later still, as they are rare in polyps but common in carcinomas (**Figure 20–48**). As discussed earlier, loss of *p53* function allows cancer cells to accumulate additional mutations and to avoid apoptosis and cell-cycle arrest.

In summary, although the exact set of mutations will vary from one colorectal cancer to the next, colorectal carcinogenesis requires certain genetic alterations that disable particular control mechanisms, and the order in which these alterations occur is not random (**Figure 20–49**).

Figure 20–48 Suggested typical sequence of genetic changes underlying the development of a colorectal carcinoma. This oversimplified diagram provides a general idea of the way mutation and tumor development can fit together. But there are certainly other mutant genes of which we are not yet aware, and different colon cancers can progress through different sequences of mutations.

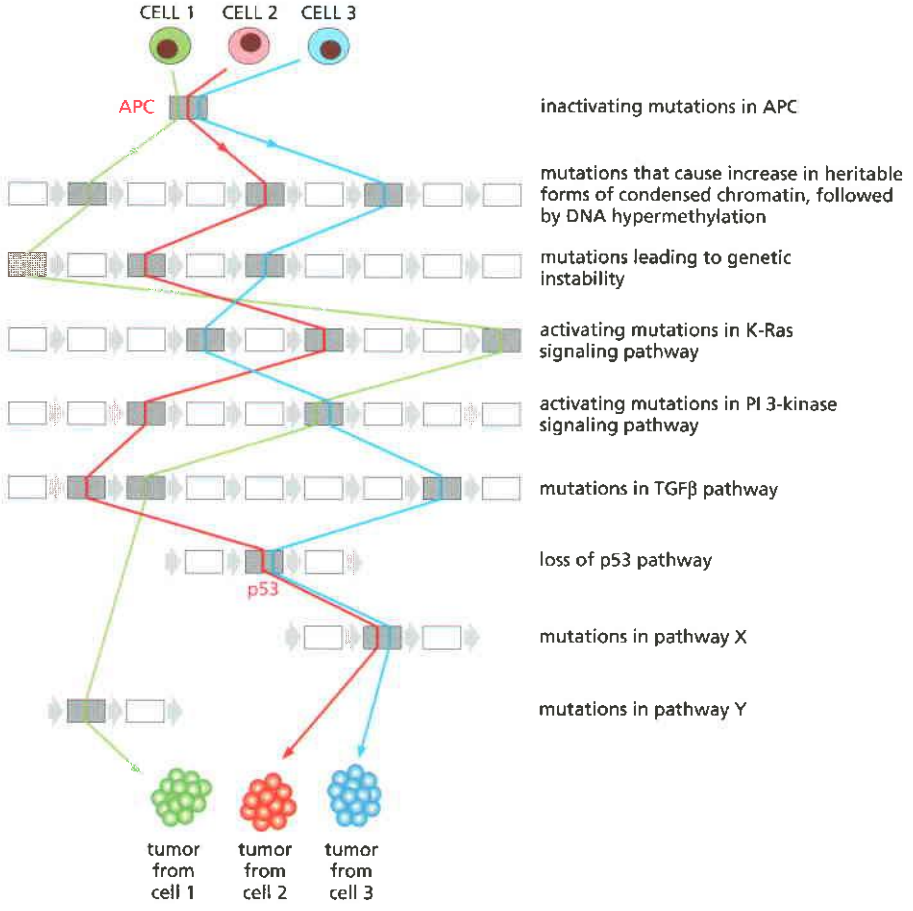


Figure 20–49 The general nature of tumor progression at the molecular level. Each row in this diagram represents a biochemical pathway, so that changing any one protein *(rectangular box)* by altering its gene is often enough to gain the full effect. The three paths *(colored vertical lines)* have been drawn to indicate the mutations that one evolving clone of tumor cells accumulates as it progresses toward cancer. This diagram is modeled after some of the known events in the development of colon cancer, but it should not be taken literally. The major points are that, while many different genes will be altered in individual colon tumors as they develop to metastasis, so that tumors are not expected to be exactly alike, there are some common patterns. In particular, both the signaling pathways that are altered and the general order in which these changes arise are often similar for different cancers.

Each Case of Cancer Is Characterized by Its Own Array of Genetic Lesions

As illustrated by colorectal cancer, the traditional classification of cancers is simplistic. Even a single type of cancer is a heterogeneous collection of disorders; despite some common features, each disorder will be characterized by its own array of genetic lesions. In the form of lung cancer known as *small-cell carcinoma,* for example, one finds mutations in *p53, Rb, Myc,* and at least five other known proto-oncogenes and tumor suppressor genes. Different patients have different combinations of mutations, and these genetic differences often correspond to different responses to treatment.

In principle, molecular biology provides the tools to find out precisely which genes are amplified, deleted, mutated, or misregulated by epigenetic mechanisms in the tumor cells of any given patient. As we discuss next, such information may soon prove to be as important for the diagnosis and treatment of cancer as was the identification of microorganisms for patients with infectious diseases.

Summary

Studies of developing embryos and genetically engineered mice have helped to reveal the functions of a large number of cancer-critical genes. Many of the oncogenes and tumor suppressor genes mutated in cancer code for components of the social control pathways that regulate when cells grow, divide, differentiate, or die. Other genes, loosely categorized as "DNA maintenance genes," help maintain the integrity of the genome. The molecular changes that allow cancers to metastasize are still largely unknown.

DNA tumor viruses such as papillomaviruses can promote the development of cancer by encoding proteins that inhibit the products of some tumor suppressor genes. The viral proteins bind and inhibit the Rb protein, which normally acts as a brake on cell division, and the p53 protein, which normally induces cell-cycle arrest or apoptosis in response to DNA damage and other cell stresses. Loss or inactivation of the p53 pathway, which occurs in nearly all spontaneous human cancers, is especially dangerous, as it allows genetically damaged cells to escape apoptosis and continue to proliferate. Inactivation of the Rb pathway also occurs in most human cancers, illustrating how fundamental each of these pathways is in protecting us against cancer.

We can often correlate the steps of tumor progression with mutations that activate specific oncogenes and inactivate specific tumor suppressor genes, with colon cancer providing the best-understood example. But different combinations of mutations and epigenetic changes are found in different types of cancer and even in different patients with the same type of cancer, reflecting the random way in which these inherited changes occur. Nevertheless, many of the same types of changes are encountered repeatedly, suggesting that there is only a limited number of ways in which our defenses against cancer can be breached.

CANCER TREATMENT: PRESENT AND FUTURE

We can apply our growing understanding of the biology of cancer to sharpen our attack on the disease at three levels: prevention, diagnosis, and treatment. Prevention is always better than cure, and, as discussed earlier, many cancers can be prevented, especially by avoiding smoking. Moreover, cancers can often be nipped in the bud by screening for small primary tumors, which can be removed before they have metastasized, as we saw for cervical cancers. Highly sensitive molecular assays promise to create new opportunities for better prevention and treatment through earlier and more precise diagnosis. Advances in screening methods along with changes in lifestyle probably offer the most immediate prospects for reducing the cancer death rate, but they can never be perfectly effective. Thus, for many years to come, full-blown malignant disease will continue to be common and cancer treatments will continue to be needed.

The Search for Cancer Cures Is Difficult but Not Hopeless

The difficulty of curing a cancer is similar to the difficulty of getting rid of weeds. Cancer cells can be removed surgically or destroyed with toxic chemicals or radiation, but it is hard to eradicate every single one of them. Surgery can rarely ferret out every metastasis, and treatments that kill cancer cells are generally toxic to normal cells as well. In addition, as described earlier, whereas the great majority of the cancer cells can often be killed by irradiation or chemotherapy, the small population of slowly dividing cancer stem cells may be harder to eliminate in these ways; if even a few cancer stem cells remain, they can regenerate the tumor. Moreover, unlike normal cells, cancer cells mutate rapidly and will often evolve resistance to the poisons and irradiation used against them.

In spite of these difficulties, effective cures using anticancer drugs (alone or in combination with other treatments) have already been found for some formerly highly lethal cancers, including Hodgkin's lymphoma, testicular cancer, choriocarcinoma, and some leukemias and other cancers of childhood. Even for types of cancer where a cure at present seems beyond our reach, there are treatments that will prolong life or at least relieve distress. But what prospect is there of doing better and finding cures for the most common forms of cancer, which still cause great suffering and so many tragic deaths?

Traditional Therapies Exploit the Genetic Instability and Loss of Cell-Cycle Checkpoint Responses in Cancer Cells

Anticancer therapies need to take advantage of some molecular abnormality of cancer cells that distinguishes them from normal cells. One such property is genetic instability, caused by an abnormality in chromosome maintenance, cell cycle checkpoints, or DNA repair. Remarkably, most current cancer therapies work by exploiting these abnormalities, although this was not known by the scientists who developed the treatments. Most anticancer drugs and ionizing radiation damage DNA. They preferentially kill certain kinds of cancer cells because these mutant cells have a diminished ability to survive the damage. Normal cells treated with radiation, for example, arrest their cell cycle until they have repaired the damage to their DNA. This cell-cycle arrest is an example of a *cell-cycle checkpoint response* discussed in Chapter 17. Cancer cells generally have defects in many of these checkpoint responses and often continue to divide after irradiation; this causes many to die after a few days because of the severe genetic damage they experience.

Unfortunately, while some of the molecular defects present in cancer cells enhance their sensitivity to cytotoxic agents, others increase their resistance. Some of the cell death that DNA damage induces, for example, occurs by apoptosis, and cancer cells often acquire defects in the control systems that activate apoptosis in response to such damage. For example, as we discussed earlier, DNA damage induced by anticancer drugs or irradiation normally activates p53, which can then trigger apoptosis. Thus the inactivation of the *p53* pathway that occurs in most cancers makes certain types of cancer cells less sensitive to these agents. Cancer cells vary widely in their response to various treatments, probably reflecting the particular kinds of defects they have in DNA repair, cell-cycle checkpoints, and the control of apoptosis.

New Drugs Can Exploit the Specific Cause of a Tumor's Genetic Instability

As we become increasingly able to pinpoint the specific alterations in cancer cells that make them different from their normal neighbors, we can use this knowledge to design therapies that kill the cells in the cancer without killing normal cells. One of the characteristics of cancer cells is their genetic instability; as we have explained, this is one of the features that helps them to evolve

and proliferate dangerously. But it is at the same time a defect—a vulnerability that we can exploit to kill them.

We have seen that cancer cells are forced to walk a difficult line as they evolve toward metastasis: they need to have a defect in their DNA maintenance processes that is severe enough to allow them to accumulate new mutations at a significantly increased rate, but not so severe that they kill themselves by frequent loss or genes necessary for cell survival. Since there are hundreds of different genes required to maintain DNA sequences and chromosome structures with high fidelity (discussed in Chapters 4 and 5), we would expect there to be at least dozens of different ways for a particular tumor cell to acquire its genetic instability. Moreover, these ways should be mutually exclusive: once a cell has become genetically unstable to a modest extent, it is likely to increase its risk of death if it inactivates additional DNA maintenance genes. Those cells that do so will die and be lost from the tumor population.

Detailed studies of the mechanisms for DNA maintenance discussed in Chapter 5 reveal a surprising amount of apparent redundancy, with multiple pathways for repairing each type of DNA damage. Thus, knocking out a particular pathway for DNA repair is generally less disastrous than one might expect, because alternate repair pathways exist. We have seen, for example, how stalled DNA replication forks can arise when the fork encounters a single-strand break in a template strand, and how cells avoid the disaster that would otherwise result. First, they have machinery to escape the problem by directly repairing single-strand breaks; and then, if that fails, they can repair stalled forks by homologous recombination (see Figure 5–53). Suppose that the cells in a particular cancer have become genetically unstable by acquiring a mutation that reduces their ability to repair stalled forks by homologous recombination. Might it be possible to eradicate that cancer by treating it with a drug that inhibits the repair of single-strand breaks, thereby greatly increasing the number of forks that stall? The consequences would be expected to be harmless for normal cells, which can repair stalled forks, but lethal for the cancer cells, which cannot.

While such a possibility might seem too good to be true, precisely this strategy appears to work to kill the cells in cancers that have inactivated both of their *Brca1* or *Brca2* tumor suppressor genes. As described in Chapter 5, Brca2 is an accessory protein that interacts with the Rad51 protein (the RecA analog in humans) in the initiation of general recombination events. Brca2 is another protein that is also required for this repair process. Like *Rb*, *Brca1* and *Brca2* were discovered as mutations that predispose humans to cancer—in this case, cancers of the breast and ovaries (though unlike *Rb*, they seem to be involved in only a small proportion of such cancers). Individuals who inherit one mutant copy of *Brca1* or *Brca2* develop tumors that have inactivated the second copy of the same gene, presumably because this change makes the cells genetically unstable and speeds tumor progression.

Drugs that inhibit an enzyme called PARP (poly ADP-ribose polymerase) have a dramatic effect on cells of such tumors, killing them with enormous selectivity. This is attributed to the fact that PARP is required for the repair of single-strand breaks in DNA. Perhaps surprisingly, PARP inhibition has very little effect on normal cells; in fact, mice that have been engineered to lack PARP1—the major PARP family member involved in DNA repair—remain healthy under laboratory conditions. This result suggests that, while the repair pathway requiring PARP provides a first line of defense against breaks in a DNA strand, these breaks can easily be repaired by a genetic recombination pathway in normal cells (**Figure 20–50**). In contrast, tumor cells that have acquired their genetic instability by the loss of Brca1 or Brca2 have lost this second line of defense, and they are therefore uniquely sensitive to PARP inhibitors (that is, they are missing repair pathway 2 in Figure 20–50).

PARP inhibition is still only under trial as a cancer treatment in humans, and is likely to be applicable to only a small proportion of cancer cases. But it exemplifies the type of rational, highly selective approach to cancer therapy that is beginning to be possible, and thus it offers hope for many other cancers. To extend this approach widely, we will need new tools for determining the precise

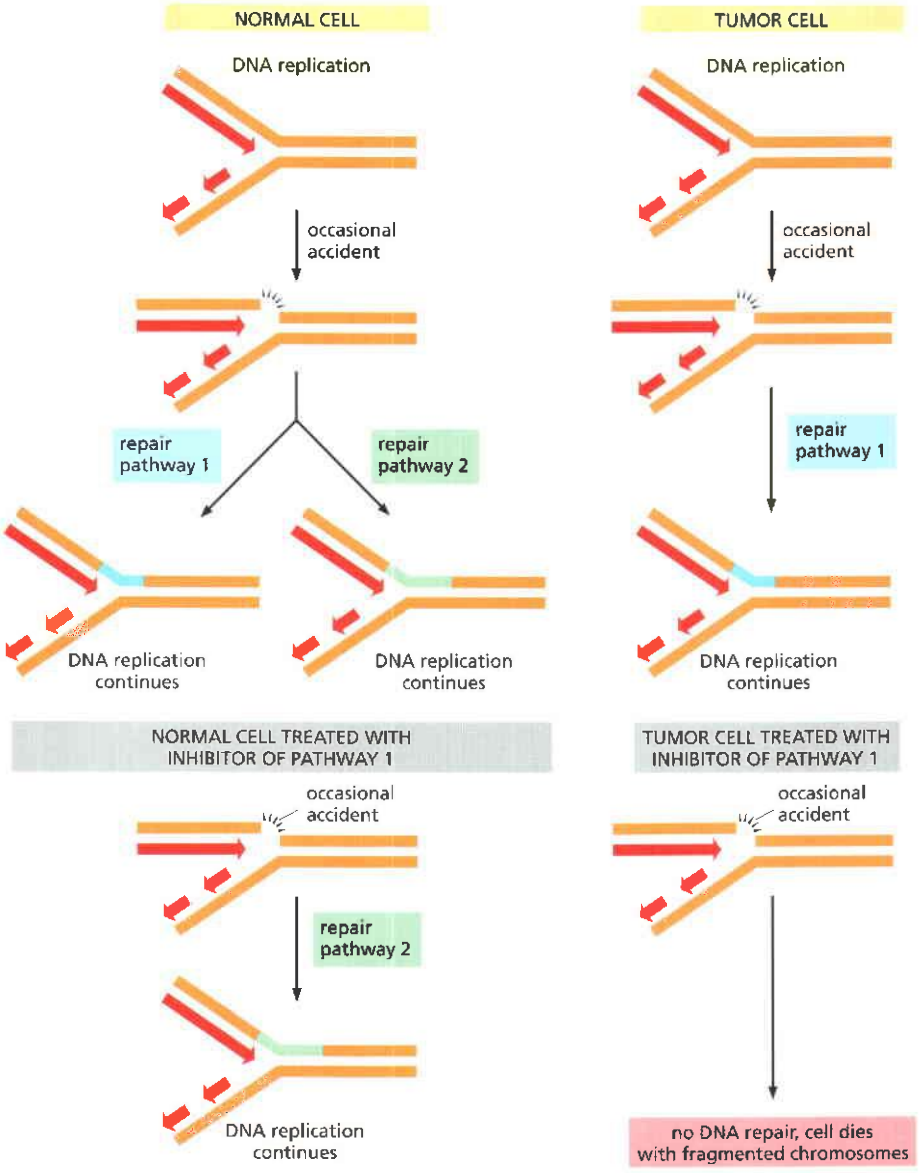


Figure 20–50 How a tumor's genetic instability can be exploited for cancer therapy. As explained in Chapter 5, the maintenance of DNA sequences is so critical for life that cells have evolved multiple pathways for repairing DNA damage and avoiding DNA replication errors. As illustrated, a replication fork will stall whenever it encounters a break in a DNA template strand. In this example, normal cells have two different repair pathways that can fix the problem and thereby prevent a mutation from arising in the newly synthesized DNA sequences. They are therefore not harmed by treatment with a drug that blocks repair pathway 1. In contrast, the inactivation of repair pathway 2 was selected for in the evolution of the tumor cells (because it made them genetically unstable). Consequently, only the tumor cells are killed by a drug treatment that blocks repair pathway 1. If a treated cell does not die by apoptosis, its daughters will die because they inherit fragmented, incomplete chromosome sets.

In the actual case that underlies this more general schematic example, the function of repair pathway 1 (requiring the PARP protein discussed in the text) is to remove accidental breaks that occur in a DNA single strand *before* they are encountered by a moving replication fork, and pathway 2 is the recombination-dependent process for repairing stalled replication forks illustrated in Figure 5–53 (requiring the Brca2 and Brca1 proteins; for details, see H.E. Bryant et al., *Nature* 434:913–916, 2005 and H. Farmer et al., *Nature* 434:917–921, 2005.)

cause of the genetic instability in individual tumors, as well as the development of many more drugs that target alternate DNA repair pathways.

Genetic Instability Helps Cancers Become Progressively More Resistant to Therapies

Genetic instability itself can be both good and bad for anticancer therapy. Although it seems to provide an Achilles' heel that therapies can exploit, it can also make eradicating cancer more difficult. An abnormally high mutation rate tends to make cancer cell populations heterogeneous, which may make it difficult to kill off the entire population with a single type of treatment. Moreover, it allows many cancers to evolve resistance to therapeutic drugs at an alarming rate.

To make matters worse, cells that are exposed to one anticancer drug often develop a resistance not only to that drug but also to other drugs to which they have never been exposed. This phenomenon of **multidrug resistance** frequently correlates with amplification of a part of the genome that contains a gene called *Mdr1*. This gene encodes a plasma-membrane-bound transport ATPase of the ABC transporter superfamily (discussed in Chapter 11), which pumps lipophilic drugs out of the cell. The overproduction of this protein (or some of its other family members) by a cancer cell can prevent the intracellular accumulation of

many cytotoxic drugs, making the cell insensitive to them. The amplification of other types of genes can likewise give the cancer cell a selective advantage. The gene that encodes the enzyme dihydrofolate reductase (DHFR), for example, can become amplified in cancer cells treated with the anticancer drug *methotrexate*. Methotrexate binds to and inhibits the ability of DHFR to bind folic acid, and the amplification greatly increases the amount of enzyme, reducing the cells' sensitivity to the drug.

New Therapies Are Emerging from Our Knowledge of Cancer Biology

Our growing understanding of cancer cell biology and tumor progression is gradually leading to better methods for treating the disease, and not only by targeting defects in cell-cycle checkpoints and DNA repair processes. As an example, estrogen receptor antagonists (such as *tamoxifen*) and drugs that block estrogen synthesis are widely used to prevent or delay recurrence of breast cancers that have been screened and found to express estrogen receptors. These antiestrogen treatments are also being tested for their ability to prevent the development of new breast cancers. These drugs do not directly kill the tumor cells, but instead prevent estrogen from promoting their proliferation.

The greatest hopes lie, however, in finding more powerful and selective ways to exterminate cancer cells directly. A variety of adventurous new ways to attack tumor cells have been shown to work in model systems—typically reducing or preventing the growth of human tumors transplanted into immunodeficient mice. Many of these strategies will turn out to be of no medical use, because they do not work in humans, have bad side effects, or are simply too difficult to implement. But some have turned out to be highly successful in the clinic. One strategy depends on the reliance of some cancer cells on a particular hyperactive protein they produce, a phenomenon known as *oncogene addiction*. Blocking the activity of the protein may be an effective means of treating the cancer if it does not unduly damage normal tissues. About 25% of breast cancers, for example, express unusually high levels of the Her2 protein, a receptor tyrosine kinase related to the EGF receptor that plays a part in the normal development of mammary epithelium. Monoclonal antibodies that inhibit Her2 function slow the growth of breast tumors in humans that overexpress Her2, and they are now an approved therapy for these cancers.

A similar approach uses antibodies to deliver toxic molecules directly to the cancer cells. Antibodies against proteins like Her2 that are abundant on the cancer cell surface can be armed with a toxin or made to carry an enzyme that cleaves a harmless 'prodrug' and converts this into a toxic molecule. In the latter case, one molecule of enzyme can then generate a large number of toxic molecules on the surface of a tumor cell; these molecules can also diffuse to neighboring tumor cells, increasing the odds that they too will be killed, even if the antibody did not bind to them directly.

Antibodies are hard to make in large amounts, are very expensive to produce and buy, and must be given by injection. The ultimate goal in cancer therapy is to develop small molecules that kill cancer cells specifically. The PARP inhibitors discussed above are one example; but to tackle the majority of cancers with simple drug treatments, we shall need a large collection of different small molecules, tailored to the many different types of cancer.

Small Molecules Can Be Designed to Inhibit Specific Oncogenic Proteins

As our knowledge of specific molecules involved in the genesis of particular cancers builds, there is an increasing effort to devise targeted therapies directed against the oncogenic proteins that are essential for a cancer cell to survive and proliferate—thereby exploiting the phenomenon of oncogene

addiction mentioned previously. A particular dramatic success of this type has raised high hopes for the general utility of such targeted therapies in the future.

As we saw earlier, chronic myelogenous leukemia (CML) is usually associated with a particular chromosomal translocation, visible as the Philadelphia chromosome (see Figure 20–5). This is the consequence of chromosome breakage and rejoining at the sites of two specific genes, called *Abl* and *Bcr*. The fusion of these genes creates a hybrid gene that codes for a chimeric protein called *Bcr-Abl*, consisting of the N-terminal fragment of Bcr fused to the C-terminal portion of Abl (**Figure 20–51**). Abl is a tyrosine kinase involved in cell signaling. The substitution of the Bcr fragment for the normal N-terminus of Abl makes it hyperactive, so that it stimulates inappropriate proliferation of the hemopoietic precursor cells that contain it and prevents these cells from dying by apoptosis—which many of them would normally do. As a result, excessive numbers of white blood cells accumulate in the bloodstream, producing CML.

The chimeric Bcr-Abl protein is an obvious target for therapeutic attack. Searches for synthetic drug molecules that can inhibit the activity of tyrosine kinases discovered one, called **Gleevec**, that blocks Bcr-Abl (**Figure 20–52**). When the drug was first given to patients with CML, nearly all of them showed a dramatic response, with an apparent disappearance of the cells carrying the Philadelphia chromosome in over 80% of patients. The response appears relatively durable: after years of continuous treatment, many patients have not progressed to later stages of the disease—although Gleevec-resistant cancers emerge with a probability of at least 5% per year.

Results are not so good for those patients who have already progressed to the acute phase of myeloid leukemia, known as blast crisis, where genetic instability has set in and the march of the disease is far more rapid. These patients show a response at first and then relapse because the cancer cells develop a resistance to Gleevec. This resistance is usually associated with secondary mutations in the part of the *Bcr-Abl* gene that encodes the kinase domain, disrupting the ability of Gleevec to bind to Bcr-Abl kinase. Second-generation inhibitors that function effectively against these Gleevec-resistant mutants have now been developed. Ultimately a cocktail of multiple agents that cooperatively block Bcr-Abl action may provide the key to successful treatment, by preventing the selection for drug-resistant cancer cells at all stages of the disease.

Despite the complications with resistance, the extraordinary success of Gleevec for the patients in the chronic (early) stage of the disease is enough to prove the principle: once we understand precisely what genetic lesions have occurred in a cancer, we can begin to design effective rational methods to treat it. This success story has fueled efforts to identify small-molecule inhibitors for other oncogenic protein kinases that could be effective targets for new anticancer drugs. A second example of such a therapy is provided by a small-molecule inhibitor of the EGF receptor, currently approved for the treatment of some lung cancers.

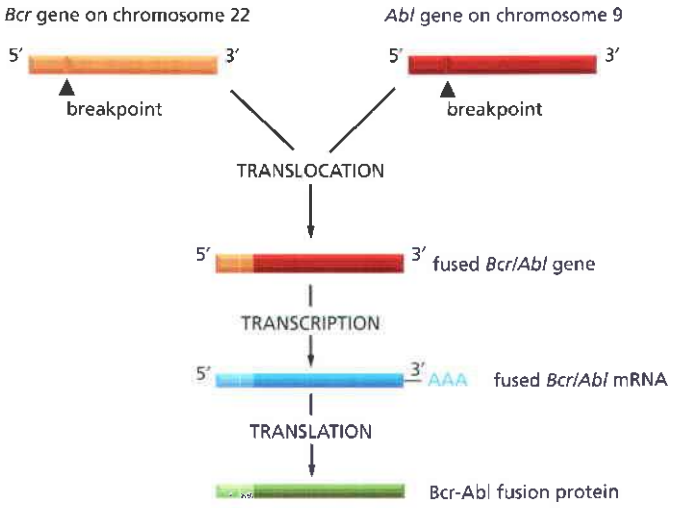


Figure 20–51 The conversion of the *Abl* proto-oncogene into an oncogene in patients with chronic myelogenous leukemia. The chromosome translocation responsible joins the *Bcr* gene on chromosome 22 to the *Abl* gene from chromosome 9, thereby generating a Philadelphia chromosome (see Figure 20–5). The resulting fusion protein has the N-terminus of the Bcr protein joined to the C-terminus of the Abl tyrosine protein kinase; in consequence, the Abl kinase domain becomes inappropriately active, driving excessive proliferation of a clone of hemopoietic cells in the bone marrow.

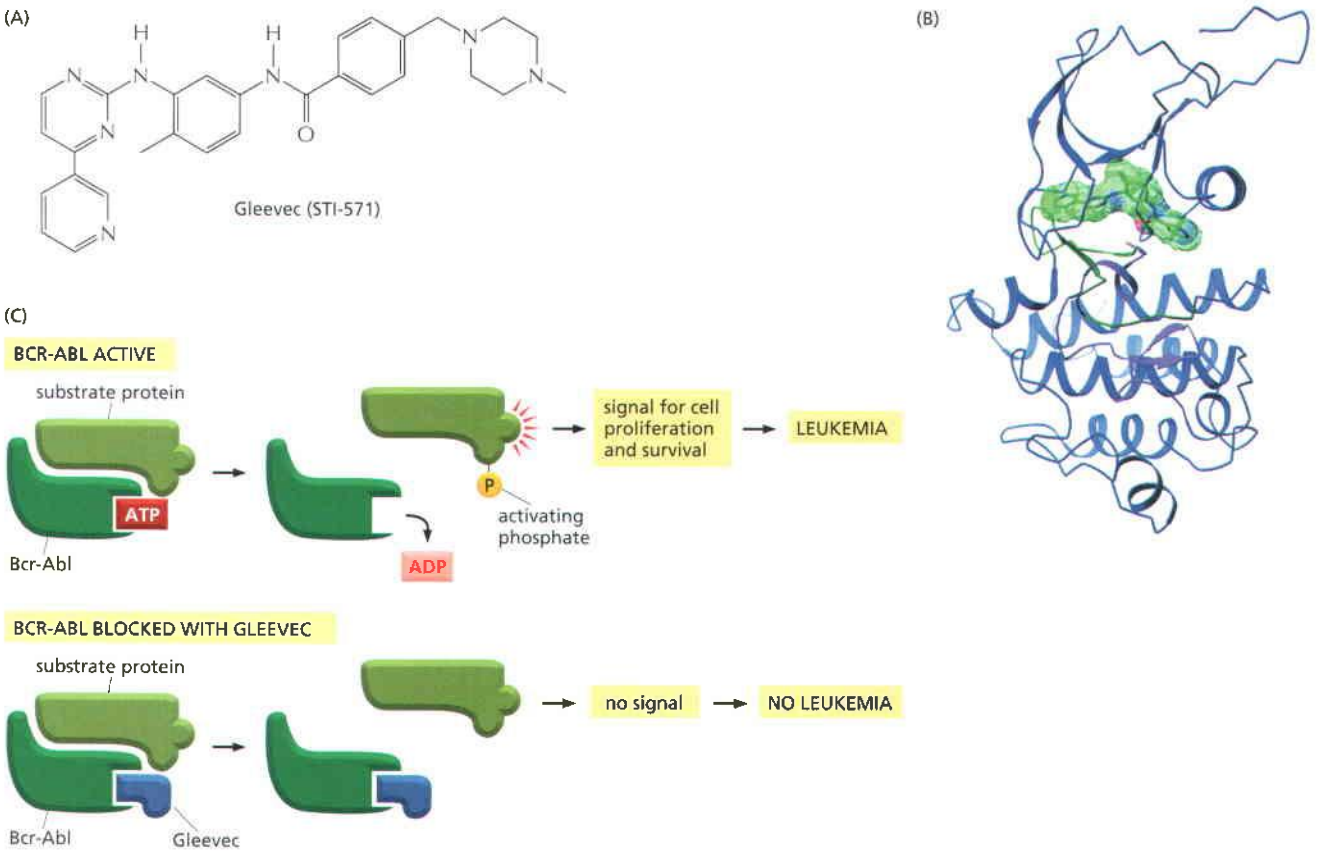


Figure 20–52 How Gleevec blocks the activity of Bcr-Abl protein and halts chronic myeloid leukemia. (A) The chemical structure of Gleevec (imatinib). The drug can be given by mouth; it has side effects but they are usually quite tolerable. (B) The structure of the complex of Gleevec (solid *green* object) with the tyrosine kinase domain of the Abl protein (ribbon diagram), as determined by X-ray crystallography. (C) Gleevec sits in the ATP-binding pocket of the tyrosine kinase domain of Bcr-Abl and thereby prevents Bcr-Abl from transferring a phosphate group from ATP onto a tyrosine residue in a substrate protein. This blocks onward transmission of a signal for cell proliferation and survival. (B, from T. Schindler et al., *Science* 289:1938–1942, 2000. With permission from AAAS.)

Tumor Blood Vessels Are Logical Targets for Cancer Therapy

Another approach to destroying tumors does not directly target cancer cells at all but instead targets the blood vessels on which large tumors depend. As discussed earlier, the growth of these vessels requires angiogenic signals such as VEGF, which, in animal models at least, can be blocked to prevent further tumor growth. Moreover, endothelial cells in the process of forming new vessels express distinctive cell-surface markers, which might provide an opportunity to attack developing tumor vessels specifically, without harming existing blood vessels in normal tissues. Clinical trials with various inhibitors of angiogenic signals are now taking place, and several drugs that inhibit the VEGF receptor have recently been approved for treating kidney cancer. Similarly, a monoclonal antibody against VEGF has been approved for treatment of colon cancer in combination with chemotherapy, even though the added benefit is only modest.

Many Cancers May Be Treatable by Enhancing the Immune Response Against a Specific Tumor

In recent years, we have realized that the body's normal immune responses help to protect us against cancer. Mice lacking important parts of their immune system have elevated levels of several types of cancer. Similarly, humans who are

immunocompromised have two- to three-fold higher rates of certain solid tumors. The self-antigens that are recognized in these cases may attract the attention of the immune system because they are expressed in unusually large amounts in the particular tumor cells. Alternatively, it is possible that these proteins provoke an autoimmune response because they are normally expressed only in embryos or in another immunologically privileged area such as the brain, so that the immune system has never had the opportunity to develop tolerance towards them.

Variant tumor cells with a reduced exposure to immune attack can defeat tumor immunosurveillance, and the process of tumor progression is thought to include heritable changes that reduce the antigenicity of the tumor. However, through our increased understanding of the complex mechanisms of both adaptive and innate immunity (see Chapters 24 and 25), oncologists are learning how to manipulate the immune response to intensify the attack on specific tumors. The hope is that this will allow these cancer cells to be eliminated as if they were foreign tissue.

Treating Patients with Several Drugs Simultaneously Has Potential Advantages for Cancer Therapy

As already described, some cancers can be dramatically reduced in size by an initial drug treatment, with all of the detectable tumor cells seeming to disappear. But months or years later the cancer will reappear in an altered form that is resistant to the drug that was at first so successful. Given the hypermutable nature of tumor cells, this is not surprising. The initial drug treatment has presumably failed to affect a tiny fraction of mutant cells in the original tumor cell population (some of the cancer stem cells, according to that view of tumors), and these cells eventually recreate the tumor by continuing to proliferate.

In some cases, this problem can be prevented by treating patients with two or more drugs simultaneously. The logic is the same as that behind the current treatment of HIV-AIDS with a cocktail of three different protease inhibitors: whereas there may always be some cells in the initial populations resistant to any one drug treatment, there should be no cell resistant to two drugs that are different. In contrast, sequential drug treatments can allow the few cells resistant to the first drug to multiply to large numbers, thereby making it likely that cells with double resistance will arise (**Figure 20–53**).

As our armament of possible drugs against cancer cells increases, it becomes increasingly likely that effective treatments involving simultaneous treatment with multiple drugs can be designed for many cancers.

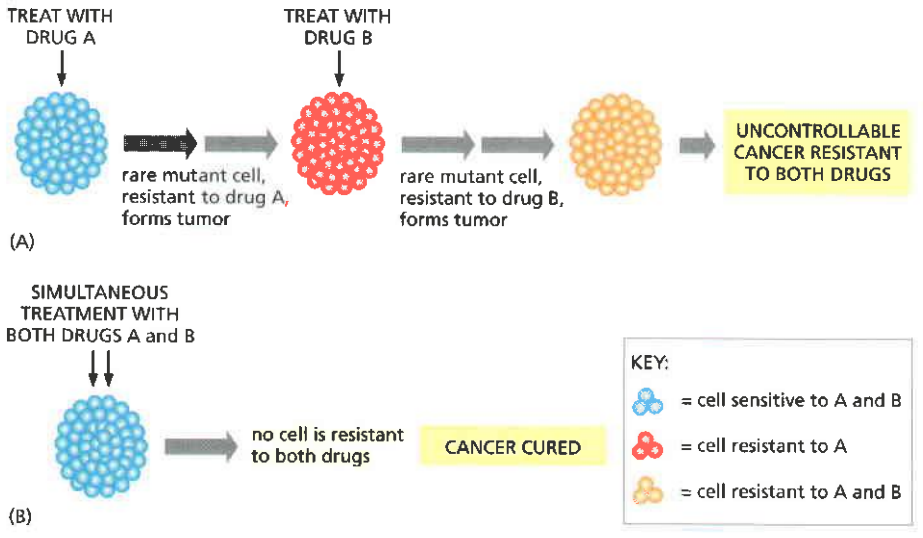


Figure 20–53 Why multidrug treatments may be more effective than sequential treatments for cancer therapy. Because tumor cells are hypermutable, single drug treatments that are given sequentially often allow for the selection of mutant cell clones that are resistant to both drugs.

Gene Expression Profiling Can Help Classify Cancers into Clinically Meaningful Subgroups

It is now possible to characterize each individual cancer at a molecular level in unprecedented detail. For example, DNA microarray technology (discussed in Chapter 8) can be used to determine the pattern of expression of thousands of genes simultaneously in a single sample of a cancer and compare it with the pattern of expression of the same genes in normal control tissue. Each case of a given type of cancer, such as breast cancer, will have its own **gene expression profile**. However, comparing the profiles of many patients shows that they can be grouped into a small number of distinct classes, whose members share common profile features.

Why might such molecular profiles be useful? All medical progress depends on accurate diagnosis. If we cannot identify a disease correctly, we cannot discover its causes, predict its outcome, select the appropriate treatment for a given patient, or conduct effective trials on a population of patients to judge whether a treatment is effective. Cancers, as we have seen, are an extraordinarily heterogeneous collection of diseases. Gene expression profiling represents a tool to make diagnosis more precise through better tumor classification. Standard classification schemes mainly rely on histological analysis and we now know that they can lump together cancers that behave differently. Thus, gene expression profiling can not only help to identify a tumor in uncertain cases, but also help to classify tumors more precisely, with important consequences for the patient. Analysis of a collection of diffuse large B-cell lymphomas, for example, demonstrated that there are two separate classes based on gene expression profiles, even though the classes could not be distinguished histologically. One class was associated with a poor prognosis and the other with a better prognosis, establishing the usefulness of the classification and providing an explanation for the previous observation that some patients with this disease respond effectively to therapy and survive, whereas others do not. Gene expression profiling is still being developed for most cancers, but it is likely to find widespread use in the future for guiding treatments.

There Is Still Much More to Do

The molecular analysis of cancer promises to transform cancer treatment by enabling us to tailor therapy much more accurately to the individual patient. The discovery of a large number of cancer-critical genes brought an end to an era of groping in the dark for clues to the molecular basis of cancer. It has been encouraging to find that there are, after all, some general principles and that many forms of the disease share some key genetic abnormalities. Because we know the identities of many cancer-critical genes and their normal functions, it is beginning to be possible to devise precisely targeted, rational treatments. But we are still far from fully understanding the most common human cancers, and we still need a much better understanding of nearly all of the processes described in this chapter before we can conquer this lethal cell-biological disease. Among other things, we need better ways to define cancer stem cells and isolate them, so that we can determine their special characteristics and the molecular basis of their "stemness;" most crucially, we need to find ways to kill them—otherwise, our therapies will ultimately fail.

Cancer survivors and their families have long been important advocates for public investment in basic biomedical research. These groups occasionally express understandable frustration with the slow pace at which our greatly increased knowledge of cancer has been translated into effective cancer therapies. But the accumulation of knowledge brings an accelerating improvement in understanding, and there is a real sense in the cancer research community today that major progress in the treatment of this group of diseases is imminent. To take one example, in recent years we have obtained a much deeper understanding of the detailed molecular processes that underlie the all-or-none decision that cells make to commit suicide through apoptosis. How long will it be before

we can analyze the abnormal signaling network in an individual tumor well enough to allow us to select a tailor-made cocktail of drugs and growth factors that specifically causes all cells in that cancer to die?

Looking back on the history of cell biology and contemplating the speed of recent progress, we can be hopeful. The desire to understand, which drives basic research, will surely reveal new ways to use our knowledge of the cell for humanitarian goals, not only in relation to cancer, but also with regard to infectious disease, mental illness, agriculture, and other areas that we can scarcely foresee.

Summary

Our growing understanding of the cell biology of cancers has already begun to lead to better ways of preventing, diagnosing and treating these diseases. Anticancer therapies can be designed to destroy cancer cells preferentially by exploiting the properties that distinguish them from normal cells, including their dependence on oncogenic proteins and the defects they harbor in their DNA repair mechanisms, cell-cycle checkpoint mechanisms, and apoptosis control pathways. It may also be possible to control the growth of tumors by attacking their blood supply and depriving them of the help they require from stromal cells. We now have proof that, by understanding normal cell control mechanisms and exactly how they are subverted in specific cancers, we can devise drugs to kill cancers precisely by attacking specific molecules critical for the growth and survival of the cancer cells. And, as we become better able to determine which genes are altered in the cells of any given tumor, we can begin to tailor treatments more accurately to each individual patient.

PROBLEMS

Which statements are true? Explain why or why not.

20–1 Genetic instability in the form of point mutations, chromosome rearrangements, and epigenetic changes needs to be maximal to allow the development of cancer.

20–2 Cancer therapies directed solely at killing the rapidly dividing cells that make up the bulk of a tumor are unlikely to eliminate the cancer from many patients.

20–3 The main environmental causes of cancer are the products of our highly industrialized way of life such as pollution and food additives.

20–4 Dimethylbenz[a]anthracene (DMBA) must be an extraordinarily specific mutagen since 90% of the skin tumors it causes have an A-to-T alteration at exactly the same site in the mutant *Ras* gene.

20–5 In the cellular regulatory pathways that control cell growth and proliferation, the products of oncogenes are stimulatory components and the products of tumor suppressor genes are inhibitory components.

Discuss the following problems.

20–6 In contrast to colon cancer, whose incidence increases dramatically with age, osteosarcoma—a tumor that occurs most commonly in the long bones—peaks during adolescence. Osteosarcomas are relatively rare in young children (up to age 9) and in adults (over 20). Why do you suppose that the incidence of osteosarcoma does not show the same sort of age-dependence as colon cancer?

20–7 As shown in **Figure Q20–1**, plots of deaths due to breast cancer and cervical cancer in women differ dramatically from the same plot for colon cancer. At around age 50 the age-dependent increase in death rates for breast and cervical cancer slows markedly, whereas death rates due to colon cancer (and most other cancers) continue to increase. Why do you suppose that the age-dependent increase in death rates for breast and cervical cancer slows after age 50?

20–8 By analogy with automobiles, defects in cancer-critical genes have been likened to broken brakes and stuck accelerators, which are caused in some cases through faulty service by bad mechanics. Using this analogy, decide how oncogenes, tumor suppressor genes, and DNA maintenance genes relate to broken brakes, stuck accelerators, and bad mechanics. Explain the basis for each of your choices.

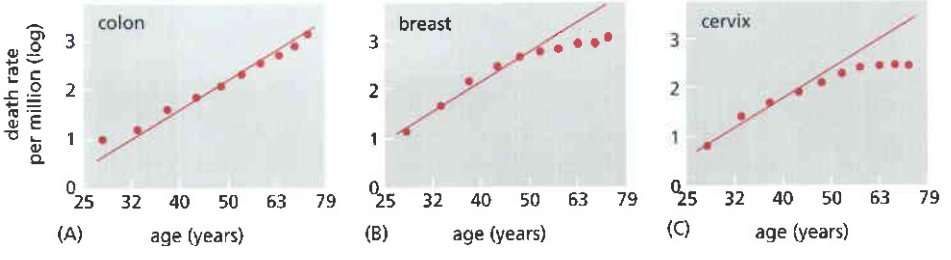


Figure Q20–1 Cancer death rates as a function of age (Problem 20–7). (A) Death rates for colon cancer in females. (B) Death rates for breast cancer in females. (C) Death rates for cervical cancer. The data in all cases are plotted as log of the death rate versus the patient age (on a log scale) at death. The straight lines in B and C are fitted to the data for the earlier age groups, whereas the line in A is fitted to all the data points. (Data from P. Armitage and R. Doll, *Br. J. Cancer* 91:1983–1989, 2004. With permission from Macmillan Publishers Ltd.)

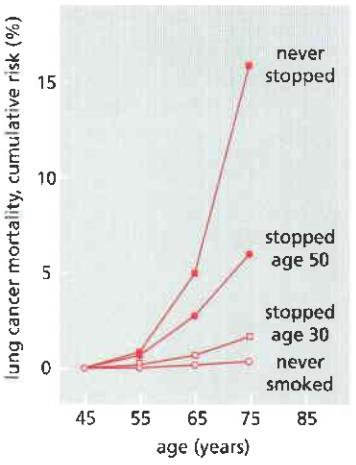


Figure Q20–2 Cumulative risk of lung cancer mortality for nonsmokers, smokers, and former smokers (Problem 20–9). Cumulative risk is the running total of deaths, as a percentage, for each group. Thus, for continuing smokers, 1% died of lung cancer between ages 45 and 55; an additional 4% died between 55 and 65 (giving a cumulative risk of 5%); and 11% more died between 65 and 75 (for a cumulative risk of 16%).

20–9 Mortality due to lung cancer was followed in groups of males in the United Kingdom for 50 years. **Figure Q20–2** shows the cumulative risk of dying from lung cancer as a function of age and smoking habits for four groups of males: those who never smoked, those who stopped at age 30, those who stopped at age 50, and those who continued to smoke. These data show clearly that an individual can substantially reduce his cumulative risk of dying from lung cancer by stopping smoking. What do you suppose is the biological basis for this observation?

20–10 The Tasmanian devil, a carnivorous Australian marsupial, is threatened with extinction by the spread of a fatal disease in which a malignant oral–facial tumor interferes with the animal's ability to feed. You have been called in to analyze the source of this unusual cancer. It seems clear to you that the cancer is somehow spread from devil to devil, very likely by their frequent fighting, which is accompanied by biting around the face and mouth. To uncover the source of the cancer, you isolate tumors from 11 devils captured in widely separated regions and examine them. As might be expected, the karyotypes of the tumor cells are highly rearranged relative to that of the wild-type devil (**Figure Q20–3**). Surprisingly, you find that the karyotypes from all 11 tumor samples are very similar. Moreover, one of the Tasmanian devils has an inversion on chromosome 5 that is not present in its facial tumor. How do you suppose this cancer is transmitted from devil to devil? Is it likely to arise as a consequence of an infection by a virus or microorganism? Explain your reasoning.

20–11 Now that DNA sequencing is so inexpensive, reliable, and fast, your mentor has set up a consortium of investigators to pursue the ambitious goal of tracking down *all* the mutations in a set of human tumors. He has decided to focus on breast cancer and colorectal cancer because they cause 14% of all cancer deaths. For each of 11 breast cancers and 11 colorectal cancers, you design primers to amplify 120,839 exons in 14,661 transcripts from 13,023 genes. As controls, you amplify the same regions from DNA samples taken from two normal individuals. You sequence the PCR products and use analytical software to compare the 456 Mb of tumor sequence with the published human genome sequence. You are astounded to find 816,986 putative mutations. This represents more than 37,000 mutations per tumor! Surely that can't be right.

Once you think about it for a while, you realize the computer sometimes makes mistakes in calling bases. To test for that source of error, you visually inspect every sequencing read and find that you can exclude 353,738 changes, leaving you with 463,248, or about 21,000 mutations per tumor. Still a lot!

A. Can you suggest at least three other sources of apparent mutations that do not actually contribute to the tumor?

B. After applying a number of criteria to filter out irrelevant sequence changes, you find a total of 1307 mutations in the 22 breast and colorectal cancers, or about 59 mutations per tumor. How might you go about deciding which of these sequence changes are likely to be cancer-related mutations and which are probably "passenger" mutations that occurred in genes with nothing to do with cancer (but were found in the tumors because they happened to occur in the same cells with true cancer mutations)?

C. Will your comprehensive sequencing strategy detect all possible genetic changes that affect the targeted genes in the cancer cells?

20–12 Virtually all cancer treatments are designed to kill cancer cells, usually by inducing apoptosis. However, one particular cancer—acute promyelocytic leukemia (APL)—has been successfully treated with all-*trans*-retinoic acid, which causes the promyelocytes to differentiate into neutrophils. How might a change in the state of differentiation of APL cancer cells help the patient?

20–13 One major goal of modern cancer therapy is to identify small molecules—anticancer drugs—that can be used to inhibit the products of specific cancer-critical genes. If you were searching for such molecules, would you design inhibitors for the products of oncogenes, tumor suppressor genes, or DNA maintenance genes? Explain why you would (or would not) select each type of gene.



(A) Tasmanian devil (*Sarcophilus harrisii*)

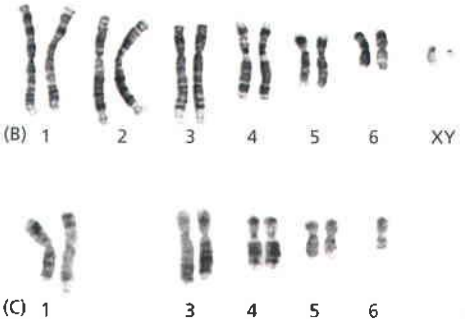


Figure Q20–3 Karyotypes of cells from Tasmanian devils (Problem 20–10). (A) A Tasmanian devil. (B) Normal karyotype for a male Tasmanian devil. The karyotype has 14 chromosomes, including XY. (C) Karyotype of cancer cells found in each of the 11 facial tumors studied. The karyotype has 13 chromosomes, no sex chromosomes, no chromosome-2 pair, one chromosome 6, two chromosomes 1 with deleted long arms, and four highly rearranged marker chromosomes (M1–M4). (From A.M. Pearse and K. Swift, *Nature* 439:549, 2006. With permission from Macmillan Publishers Ltd.)

REFERENCES

General

Bishop JM (2004) How to Win the Nobel Prize: An Unexpected Life in Science. Harvard University Press: Cambridge MA.

Stillman B & Stewart D (eds) (2005) Molecular Approaches to Controlling Cancer. Cold Spring Harbor Laboratory Press: Cold Spring Harbor.

Weinberg RA (2007) The Biology of Cancer. Garland Science: New York.

Cancer as a Microevolutionary Process

Al-Hajj M, Becker MW, Wicha M et al (2004) Therapeutic implications of cancer stem cells. *Curr Opin Genet Dev* 14:43–47.

Bielas JH, Loeb KR, Rubin BP et al (2006) Human cancers express a mutator phenotype. *Proc Natl Acad Sci USA* 103:18238–18242.

Cairns J (1975) Mutation, selection and the natural history of cancer. *Nature* 255:197–200.

Campesi J (2005) Senescent cells, tumor supression, and organismal aging: good citizens, bad neighbors. *Cell* 120:513–522.

Evan GI & Vousden KH (2001) Proliferation, cell cycle and apoptosis in cancer. *Nature* 411:342–348.

Fialkow PJ (1976) Clonal origin of human tumors. *Biochim Biophys Acta* 458:283–321.

Fidler IJ (2003) The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nature Rev Cancer* 3:453–458.

Jones PA & Baylin SB (2002) The fundamental role of epigenetic events in cancer. *Nature Rev Genet* 3:415–428.

Loeb LA, Loeb KR & Anderson JP (2003) Multiple mutations and cancer. *Proc Natl Acad Sci USA* 100:776–781.

Lowe SW, Cepero E & Evan G (2004) Intrinsic tumour suppression. *Nature* 432:307–315.

Nowell PC (1976) The clonal evolution of tumor cell populations. *Science* 194:23–28.

Pardal R, Clarke MF & Morrison SJ (2003) Applying the principles of stem-cell biology to cancer. *Nature Rev Cancer* 3:895–902.

Sjoblom T, Jones S, Wood LD et al (2006) The consensus coding sequences of human breast and colorectal cancers. *Science* 314:268–274.

Thiery JP (2002) Epithelial-mesenchymal transitions in tumour progression. *Nature Rev Cancer* 2:442–454.

Tlsty TD & Coussens LM (2006) Tumor Stroma and Regulation of Cancer Development. *Annu Rev Pathol Mech Dis* 1:119–150.

Ward RJ & Dirks PB (2007) Cancer stem cells: at the headwaters of tumor development. *Annu Rev Pathol Mech Dis* 2:175–189.

The Preventable Causes of Cancer

Ames B, Durston WE, Yamasaki E & Lee FD (1973) Carcinogens are mutagens: a simple test system combining liver homogenates for activation and bacteria for detection. *Proc Natl Acad Sci USA* 70:2281–2285.

Berenblum I (1954) A speculative review: the probable nature of promoting action and its significance in the understanding of the mechanism of carcinogenesis. *Cancer Res* 14:471–477.

Cairns J (1985) The treatment of diseases and the war against cancer. *Sci Am* 253:51–59.

Doll R (1977) Strategy for detection of cancer hazards to man. *Nature* 265:589–597.

Doll R & Peto R (1981) The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *J Natl Cancer Inst* 66:1191–1308.

Hsu IC, Metcalf RA, Sun T et al (1991) Mutational hotspot in the p53 gene in human hepatocellular carcinomas. *Nature* 350:427–428.

Newton R, Beral V & Weiss RA (eds) (1999) Infections and Human Cancer. Cold Spring Harbor Laboratory Press: Cold Spring Harbor.

Peto J (2001) Cancer epidemiology in the last century and the next decade. *Nature* 411:390–395.

Finding the Cancer-Critical Genes

Davies H, Bignell GR, Cox C et al (2002) Mutations of the BRAF gene in human cancer. *Nature* 417:949–954.

Easton DF, Pooley KA, Dunning AM et al (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 447:1087–1095.

Feinberg AP (2007) Phenotypic plasticity and the epigenetics of human disease. *Nature* 447:433–440.

Herman JG & Baylin SB (2003) Gene silencing in cancer in association with promoter hypermethylation. *N Engl J Med* 349:2042–2054.

Lowe SW, Cepero E & Evan G (2004) Intrinsic tumour suppression. *Nature* 432:307–315.

Mitelman F, Johansson B & Mertens F (2007) The impact of translocations and gene fusions on cancer causation. *Nature Rev Cancer* 7:233–245.

Pinkel D & Albertson DG (2005) Comparative genomic hybridization. *Annu Rev Genomics Hum Genet* 6:331–354.

Rowley JD (2001) Chromosome translocations: dangerous liasisons revisited. *Nature Rev Cancer* 1:245–250.

The Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661–683.

Westbrook TF, Stegmeier F & Elledge SJ (2005) Dissecting cancer pathways and vulnerabilities with RNAi. *Cold Spring Harb Symp Quant Biol* 70:435–444.

The Molecular Basis of Cancer Cell Behavior

Artandi SE & DePinho RA (2000) Mice without telomerase: what can they teach us about human cancer? *Nature Med* 6:852–855.

Brown JM & Attardi LD (2005) The role of apoptosis in cancer development and treatment response. *Nature Rev Cancer* 5:231–237.

Chambers AF, Naumov GN, Vantyghem S & Tuck AB (2000) Molecular biology of breast cancer metastasis: clinical implications of experimental studies on metastatic inefficiency. *Breast Cancer Res* 2:400–407.

Edwards PAW (1999) The impact of developmental biology on cancer research: an overview. *Cancer Metastasis Rev* 18:175–180.

Futreal PA, Coin L, Marshall M et al (2004) A census of human cancer genes. *Nature Rev Cancer* 4:177–183.

Hanahan D & Weinberg RA (2000) The hallmarks of cancer. *Cell* 100:57–70.

Hoffman RM (2005) The multiple uses of fluorescent proteins to visualize cancer *in vivo*. *Nature Rev Cancer* 5:796–806.

Kinzler KW & Vogelstein B (1996) Lessons from hereditary colorectal cancer. *Cell* 87:159–170.

Levine AJ, Hu W & Feng Z (2006) The P53 pathway: what questions remain to be explored? *Cell Death Differ* 13:1027–1036.

Macleod KF & Jacks T (1999) Insights into cancer from transgenic mouse models. *J Pathol* 187:43–60.

Mani SA, Yang J, Brooks M et al (2007) Mesenchyme Forkhead 1 (FOXC2) plays a key role in metastasis and is associated with aggressive basal-like breast cancers. *Proc Natl Acad Sci USA* 104:10069–10074.

Mueller MM & Fusenig NE (2004) Friends or foes—bipolar effects of the tumour stroma in cancer. *Nature Rev Cancer* 4:839–849.

Plas DR & Thompson CB (2005) Akt-dependent transformation: there is more to growth than just surviving. *Oncogene* 24:7435–7442.

Ridley A (2000) Molecular switches in metastasis. *Nature* 406:466–467.

Shaw RJ & Cantley LC (2006) Ras, PI(3)K and mTOR signalling controls tumour cell growth. *Nature* 441:424–430.

Sherr CJ (2006) Divorcing ARF and p53: an unsettled case. *Nature Rev Cancer* 6:663–673.

Vogelstein B & Kinzler KW (2004) Cancer genes and the pathways they control. *Nature Med* 10:789–799.

Weinberg RA (1995) The retinoblastoma protein and cell-cycle control. *Cell* 81:323–330.

Cancer Treatment: Present and Future

Alizadeh AA, Eisen MB, Davis RE et al (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403:503–511.

Baselga J & Albanell J (2001) Mechanism of action of anti-HER2 monoclonal antibodies. *Annu Oncol* 12:S35–S41.

Druker BJ & Lydon NB (2000) Lessons learned from the development of an abl tyrosine kinase inhibitor for chronic myelogenous leukemia. *J Clin Invest* 105:3–7.

Folkman J (1996) Fighting cancer by attacking its blood supply. *Sci Am* 275:150–154.

Golub TR, Slonim DK, Tamayo P et al (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286:531–537.

Huang P & Oliff A (2001) Signaling pathways in apoptosis as potential targets for cancer therapy. *Trends Cell Biol* 11:343–348.

Garraway LA & Sellers WR (2006) Lineage dependency and lineage-survival oncogenes in human cancer. *Nature Rev Cancer* 6:593–602.

Jonkers J & Berns A (2004) Oncogene addiction: sometimes a temporary slavery. *Cancer Cell* 6:535–538.

Jain RK (2005) Normalization of tumor vasculature: an emerging concept in antiangiogenic therapy. *Science* 307:58–62.

Kim R, Emi M & Tanabe K (2007) Cancer immunoediting from immune surveillance to immune escape. *Immunology* 121:1–14.

Madhusudan S & Middleton MR (2005) The emerging role of DNA repair proteins as predictive, prognostic and therapeutic targets in cancer. *Cancer Treat Rev* 31:603–617.

Roninson IB, Abelson HT, Housman DE et al (1984) Amplification of specific DNA sequences correlates with multi-drug resistance in Chinese hamster cells. *Nature* 309:626–628.

Sarkadi B, Homolya L, Szakács G & Váradi A (2006) Human multidrug resistance ABCB and ABCG transporters: participation in a chemoimmunity defense system. *Physiol Rev* 86:1179–236.

Sawyers C (2004) Targeted cancer therapy. *Nature* 432:294–297.

van 't Veer LJ, Dai H, van de Vijver MJ et al (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415:530–536.

Varmus H, Pao W, Politi K et al (2005) Oncogenes come of age. *Cold Spring Harb Symp Quant Biol* 70:1–9.

Ventura A, Kirsch DG, McLaughlin ME et al (2007) Restoration of p53 function leads to tumour regression *in vivo*. *Nature* 445:661–665.

Sexual Reproduction: Meiosis, Germ Cells, and Fertilization

21

Sex is not absolutely necessary. Single-celled organisms can reproduce by simple mitotic division, and many plants propagate vegetatively by forming multicellular offshoots that later detach from the parent. Likewise, in the animal kingdom, a solitary multicellular *Hydra* can produce offspring by budding (**Figure 21–1**), and sea anemones and marine worms can split into two half-organisms, each of which then regenerates its missing half. There are even some lizard species that consist only of females that reproduce without mating. Although such **asexual reproduction** is simple and direct, it gives rise to offspring that are genetically identical to their parent. **Sexual reproduction**, by contrast, mixes the genomes from two individuals to produce offspring that differ genetically from one another and from both parents. This mode of reproduction apparently has great advantages, as the vast majority of plants and animals have adopted it. Even many procaryotes and eucaryotes that normally reproduce asexually engage in occasional bouts of genetic exchange, thereby producing offspring with new combinations of genes. This chapter describes the cellular machinery of sexual reproduction. Before discussing in detail how the machinery works, however, we will briefly consider what sexual reproduction involves and what its benefits might be.

In This Chapter

OVERVIEW OF SEXUAL REPRODUCTION 1269

MEIOSIS 1272

PRIMORDIAL GERM CELLS AND SEX DETERMINATION IN MAMMALS 1282

EGGS 1287

SPERM 1292

FERTILIZATION 1297

OVERVIEW OF SEXUAL REPRODUCTION

Sexual reproduction occurs in **diploid** organisms, in which each cell contains two sets of chromosomes, one inherited from each parent. The specialized cells that carry out sexual reproduction, however, are **haploid**; that is, they each contain only one set of chromosomes. In the final step of sexual reproduction, a haploid cell of one individual fuses with a haploid cell of another, mixing the two genomes and restoring the diploid state. Sexual reproduction, therefore, requires a specialized type of cell division called *meiosis*, in which a diploid precursor cell gives rise to haploid progeny cells, rather than to diploid cells as occurs in ordinary mitotic cell division.

In sexually reproducing multicellular organisms, the haploid cells produced by meiosis develop into highly specialized **gametes**—*eggs* (or *ova*), *sperm* (or *spermatozoa*), pollen, or spores. In animals, females typically produce large and nonmotile eggs, whereas males typically produce small and motile sperm (**Figure 21–2**). At *fertilization*, a haploid sperm fuses with a haploid egg to form a diploid cell (a fertilized egg, or *zygote*), which contains a new combination of chromosomes. The zygote then develops into a new multicellular organism through repeated rounds of ordinary mitosis, followed by cell specialization, which includes the production of gametes (**Figure 21–3**A).

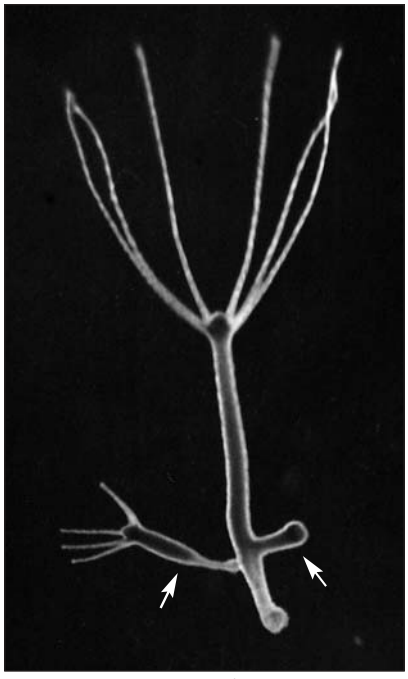


Figure 21–1 Photograph of a *Hydra* from which two new organisms are budding *(arrows).* The offspring, which are genetically identical to their parent, will eventually detach and live independently. (Courtesy of Amata Hornbruch.)

The Haploid Phase in Higher Eucaryotes Is Brief

In most organisms that reproduce sexually, diploid cells proliferate by mitotic cell division, and the haploid cells that form by meiosis do not proliferate. Some simple organisms, such as fission yeasts, are exceptional in that haploid cells

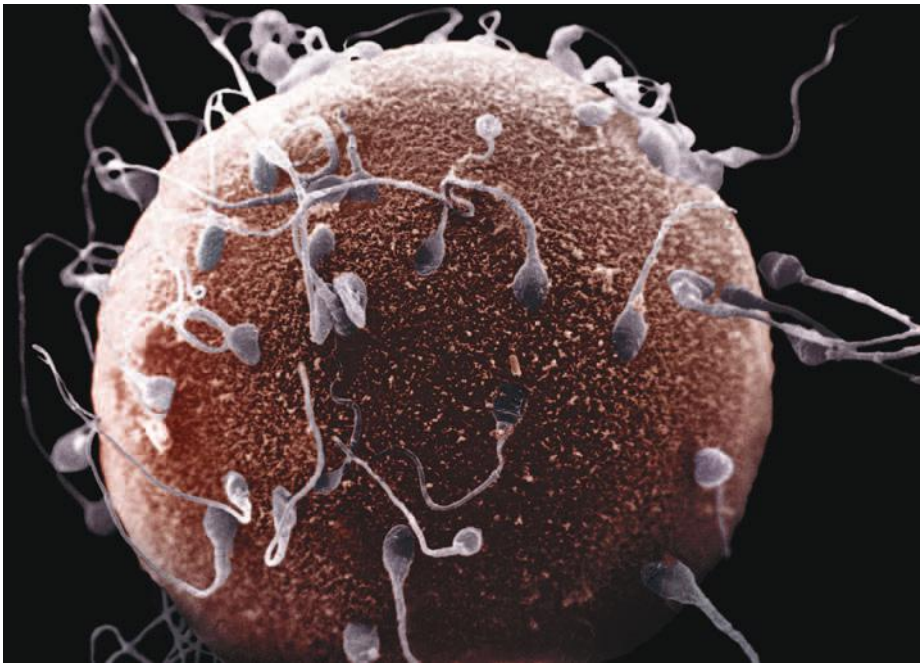


Figure 21–2 Scanning electron micrograph of an egg with many human sperm bound to its surface. Whereas the egg is immotile, the sperm are highly motile. Although many sperm are bound to the egg, only one will fertilize it, as we discuss later. (Courtesy of D. Phillips/ Science Photo Library.)

proliferate by mitotic cell division, and the diploid cells formed by the fusion of haploid cells proceed directly to meiosis to produce new haploid cells (Figure 21–3B). A less extreme exception occurs in plants, where both haploid and diploid cells proliferate. In all but the most primitive plants, such as mosses and ferns, however, the haploid phase is very brief and simple, while the diploid phase is extended into a long period of development and cell proliferation.

For almost all animals, including all vertebrates, only the diploid cells proliferate: the haploid gametes exist only briefly, do not divide at all, and are highly specialized for sexual fusion. In these organisms, it is useful to distinguish

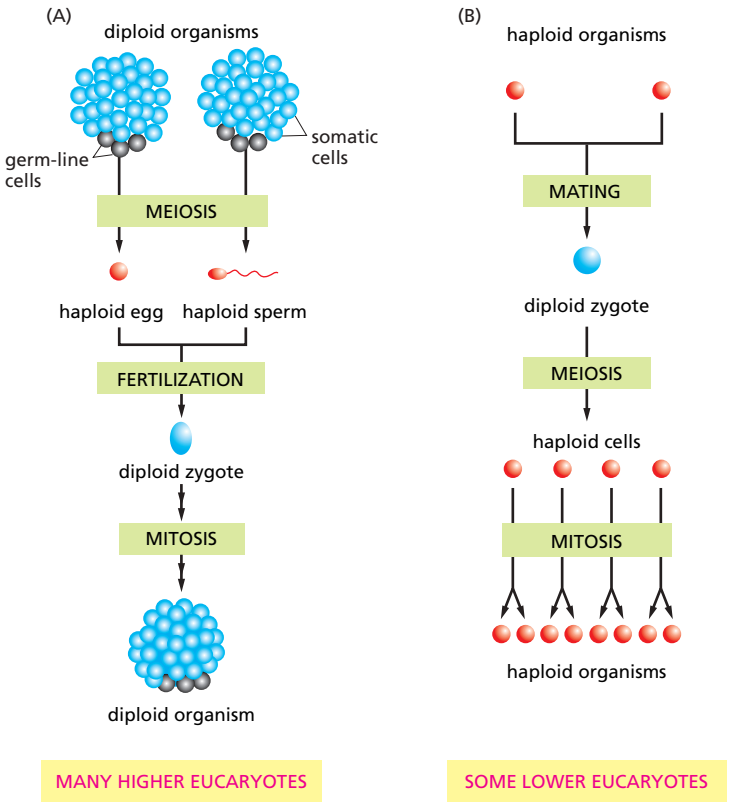


Figure 21–3 Haploid and diploid cells in the life cycles of some complex and simple eucaryotes. The haploid cells are shown in *red* and the diploid cells in *blue*. (A) Cells in most animals and plants usually proliferate in the diploid phase to form a multicellular organism; only the gametes (eggs and sperm in animals) are haploid, and they fuse at fertilization to form a diploid zygote, which develops into a new individual. The gametes develop from diploid germ-line cells *(gray)* in the gonads; all the rest of the cells in the organism are somatic cells. (B) In some simple eucaryotes such as fission yeast and the green alga *Chlamydomonas*, by contrast, the haploid cells proliferate, and the only diploid cell is the zygote, which exists transiently after mating.

between the cells of the **germ line** (or **germ cells**), which include gametes and their specified diploid precursor cells, and the **somatic cells**, which form the rest of the body and ultimately leave no progeny (see Figure 21–3A). In a sense, the somatic cells exist only to help the germ-line cells survive, develop, and transmit their DNA to the next generation.

Meiosis Creates Genetic Diversity

Sexually reproducing organisms inherit two full sets of chromosomes, one from each parent. Each set contains *autosomes,* which are common to all members of the species, and *sex chromosomes,* which are differently distributed according to the sex of the individual. Therefore, each diploid nucleus contains two closely similar versions of each autosome, plus a set of sex chromosomes appropriate to the sex of the individual. The two copies of each autosome, one from the mother and one from the father, are called **homologous chromosomes**, or **homologs**, and in most cells they maintain a separate existence as independent chromosomes. During meiosis, however, each chromosome must communicate with its unique homologous partner by physically pairing and undergoing genetic recombination. This communication is essential to enable the homologs to segregate accurately into different daughter cells during meiosis.

A crucial feature of meiosis is that it generates haploid cells that are genetically different from one another and from the two haploid cells that formed the organism in the first place. The genetic differences arise by two mechanisms. First, an individual gamete contains either the maternal or paternal version of each chromosome; because the choice of maternal or paternal occurs independently and randomly for each pair of homologs, the original maternal and paternal chromosomes are reshuffled into novel combinations in the haploid cells. Second, although the maternal and paternal versions of each chromosome have similar DNA sequences, they are not identical, and they undergo genetic recombination during meiosis—a process called *crossing-over* (discussed in Chapter 5) to produce novel hybrid versions of each chromosome; thus, each chromosome in a gamete contains a unique mixture of genetic information from both parents. We discuss these two mechanisms in more detail later (see Figure 21–13).

Sexual Reproduction Gives Organisms a Competitive Advantage

The machinery of sexual reproduction is elaborate, and the resources spent on it are large (**Figure 21–4**). What are its benefits, and why did it evolve? Sexually reproducing individuals produce varied offspring, whose varied genotypes are at least as likely to represent a change for the worse as a change for the better. Why, then, should they have a competitive advantage over individuals that breed true, by an asexual process? This question continues to perplex evolutionary biologists.

One advantage of sexual reproduction seems to be that the reshuffling of genes helps a species to survive in an unpredictably variable environment. If two parents produce many offspring with a wide variety of gene combinations, the chance that at least one of their progeny will have the combination of features necessary for survival in a changing environment is increased. Indeed, a population of budding yeast genetically engineered so that it cannot undergo meiotic genetic recombination and therefore cannot reproduce sexually adapts much less well over time to harsh environmental conditions than does the wild-type population, which can reproduce sexually.

Another advantage of sexual reproduction seems to be that it can help eliminate deleterious genes from a population: females generally mate with the fittest males, so that the least fit males leave no progeny and serve only as a sort of genetic trashcan. This stringent selection among males means that "good" genes are transmitted and "bad" genes are lost from the population more efficiently than they would otherwise be. As a result, members of the sexually reproducing population are expected to have much higher average fitness than members of an equivalent population that reproduces asexually.

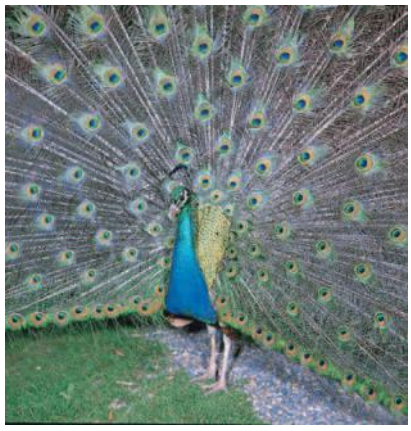


Figure 21–4 A peacock displaying his elaborate tail. This extravagant plumage serves to attract females for the purpose of sexual reproduction. It has evolved because only the fittest and most handsome males leave progeny. (Courtesy of Cyril Laubscher.)

Whatever the benefits of sexual reproduction may be, it is striking that practically all complex present-day organisms have evolved largely through generations of sexual, rather than asexual, reproduction. Asexual organisms, although plentiful, seem mostly to have remained comparatively simple and primitive.

We now turn to the cellular mechanisms of sex, beginning with the events of meiosis. We then focus our discussion mainly on mammals. We first consider the diploid cells of the germ line that give rise to the gametes and how the sex of a mammal is determined. We then discuss the nature of the gametes themselves. Finally, we consider the process of fertilization, in which an egg and a sperm fuse to form a new diploid organism.

Summary

The sexual reproductive cycle involves an alternation of diploid and haploid states: diploid cells divide by meiosis to form haploid cells, and the haploid cells from two individuals fuse in pairs to form new diploid zygotes. In the process, genomes are mixed and recombined to produce individuals with novel genetic combinations. In most higher eucaryotes, diploid cells proliferate by mitosis, and only a small proportion of them (those of the germ line) undergo meiosis to produce haploid cells; the haploid cells develop into gametes, which are specialized for sexual reproduction, exist only briefly, and do not divide. Sexual reproduction is thought to be advantageous both because it produces individuals with novel genetic combinations, some of which can survive and procreate in an unpredictably variable environment, and because it provides an efficient way to eliminate harmful mutations from a population.

MEIOSIS

The realization that gametes are haploid came from an observation that also suggested that chromosomes carry genetic information. In 1883, it was discovered in a study of roundworms that the nucleus of an unfertilized egg and that of a sperm each contain two chromosomes, whereas the fertilized egg (zygote) contains four. This led to the chromosome theory of heredity, which explained the long-standing paradox that the maternal and paternal contributions to the character of the progeny seem to be equal, despite the enormous difference in size between the egg and sperm (see Figure 21–2).

The finding also implied that haploid germ cells arise from a special kind of cell division in which the number of chromosomes is precisely halved. This type of division, called **meiosis**—the Greek word for diminution or lessening—begins in animals in diploid germ-line cells in the ovaries or testes. It might seem as if meiosis could occur by a simple modification of mitosis, in which DNA synthesis (S phase) is omitted and a single cell division produces two haploid cells directly. Meiosis, however, is more complex than this and involves two cell divisions rather than one, but with only one round of DNA synthesis. It was not until the early 1930s, as a result of painstaking cytological and genetic studies, that the basic features of meiosis were established. More recently, genetic and molecular studies have begun to identify the various meiosis-specific proteins that cause meiotic chromosomes to behave differently from mitotic chromosomes and help mediate the crucial genetic recombination events that occur in meiosis. We will see that the recombination events are important not only for genetic mixing, but also for accurate chromosome segregation during meiosis.

Gametes Are Produced by Two Meiotic Cell Divisions

Meiosis uses much of the same molecular machinery and control systems that operate in ordinary mitosis. In this chapter, however, we focus on the special features of meiosis that distinguish it from mitosis. At the beginning of meiosis, as in mitosis, the chromosomes have replicated their DNA (in meiotic S phase), and the two copies are tightly bound together by *cohesin complexes* along their entire

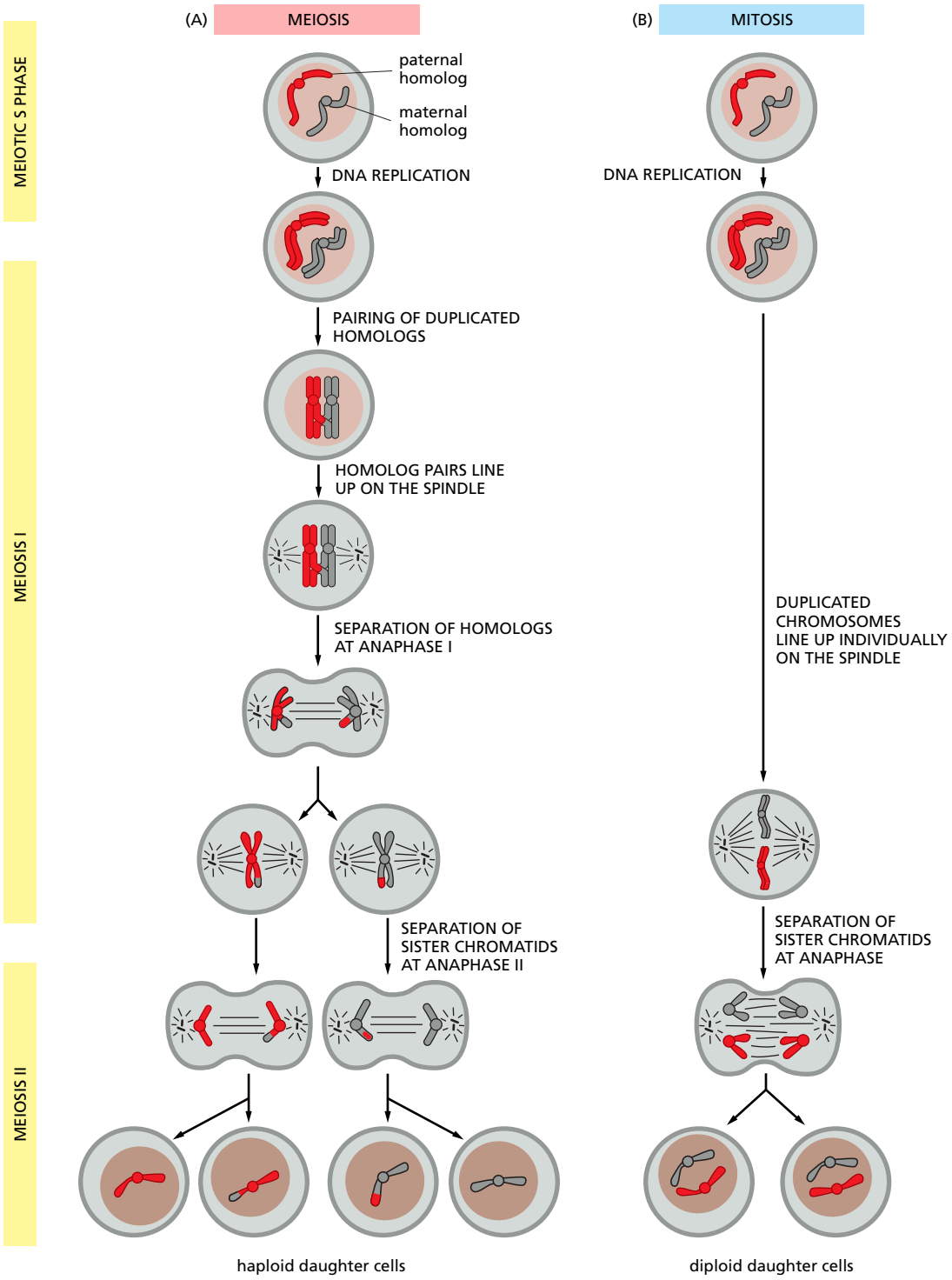


Figure 21–5 Comparison of meiosis and mitotic cell division. For clarity, only one pair of homologous chromosomes (homologs) is shown. (A) In meiosis, after DNA replication, two nuclear (and cell) divisions are required to produce the haploid gametes. The duplicated homologs, each consisting of tightly bound sister chromatids, pair up and are segregated into different daughter cells in meiosis I; the sister chromatids separate only in meiosis II. As indicated by the formation of chromosomes that are partly *red* and partly *gray*, homolog pairing in meiosis leads to genetic recombination (crossing-over) during meiosis I, as discussed later. Each diploid cell that enters meiosis therefore produces four genetically different haploid cells. <AGTG> (B) In mitosis, by contrast, homologs do not pair up, and the sister chromatids separate during the single division. Thus, each diploid cell that divides by mitosis produces two genetically identical diploid daughter cells.

length (see Figure 17–24) and are called **sister chromatids**. Unlike mitosis, however, meiosis has to produce gametes with half as many chromosomes as their diploid precursor cells. This is achieved by modifying the mitotic program so that a single round of DNA replication is followed by two successive rounds of chromosome segregation (**Figure 21–5**A). Recall that in mitosis (discussed in Chapter 17), the duplicated chromosomes line up in random order at the equator of the mitotic spindle, and the sister chromatids are pulled apart and segregated into the two daughter cells, so that each daughter inherits a complete diploid set of chromosomes and is genetically identical to the parent cell (Figure 21–5B). In division I of meiosis (**meiosis I**), by contrast, the duplicated paternal and maternal homologs (including the two replicated sex chromosomes) pair up along side each other and exchange genetic information through the process of genetic

recombination. They then line up at the equator of the meiotic spindle, after which the duplicated homologs rather than the sister chromatids are pulled apart and segregated into the two daughter cells. Only in division II of meiosis (**meiosis II**), which occurs without further DNA replication, are the sister chromatids pulled apart and segregated to produce haploid daughter cells. In this way, each diploid cell that enters meiosis produces four haploid cells, each of which inherits either the maternal or paternal copy of each chromosome, but not both (see Figure 21–5A).

Duplicated Homologs (and Sex Chromosomes) Pair During Early Prophase I

During mitosis in most organisms, homologous chromosomes behave independently of each other. During meiosis I, however, it is crucial that homologs recognize each other and associate physically in order for the maternal and paternal homologs to undergo genetic recombination and to segregate to different daughter cells at anaphase I. Special mechanisms mediate these intimate interactions between homologs.

The progressive juxtaposition of homologs occurs during a very prolonged meiotic prophase (prophase I), which can take hours in yeasts, days in mice, and weeks in higher plants. Like their mitotic counterparts, duplicated meiotic prophase chromosomes initially appear as long threadlike structures, in which the sister chromatids are so tightly glued together that they appear as one. It is during early prophase I that the homologs begin to associate along their length in a process called **pairing**, which, in some organisms at least, occurs initially through interactions between complementary DNA sequences (called *pairing sites*) in the two homologs; in most organisms, stable pairing requires genetic recombination between the homologs. As prophase I progresses, the homologs become more closely juxtaposed, forming a four-chromatid structure called a **bivalent** (**Figure 21–6**A). As we discuss later, genetic recombination begins during pairing in early prophase I, with the production of programmed double-strand breaks in chromatid DNA; some of these recombination events will later resolve into *crossovers*, where a fragment of a maternal chromatid is exchanged for a corresponding fragment of a homologous paternal chromatid (Figure 21–6B; also see Figure 5–64).

The pairing of homologs requires chromosome movements, but it is not known what drives these movements. The replicated chromosomes undergo major rearrangements within the nucleus during prophase I. The ends of the chromosomes (the *telomeres*) are tightly bound to the inner surface of the nuclear envelope. They are initially distributed diffusely there, but they then cluster transiently at one spot on the envelope and, later still, disperse again (**Figure 21–7**). Neither the mechanism nor the roles of these rearrangements are known, although they are thought to make prophase I faster and more efficient. One possibility is that they help prevent chromosome entanglements during prophase I. In fission yeast, telomere clustering is required for homolog pairing and crossing-over, but in some organisms it occurs after pairing is well underway.

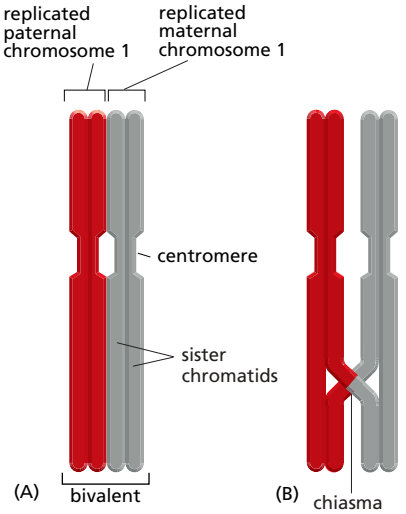


Figure 21–6 Homolog alignment and crossing-over. (A) The structure formed by two closely aligned duplicated homologs is called a *bivalent*. As in mitosis, the sister chromatids in each homolog are tightly connected along their entire lengths, as well as at their centromeres. At this stage, the homologs are usually joined together by a protein complex called the *synaptonemal complex* (not shown; see Figure 21–9). (B) A later-stage bivalent in which a single crossover event has occurred between non-sister chromatids. It is only when the synaptonemal complex disassembles and the paired homologs separate a little at the end of prophase I, as shown, that the crossover is seen microscopically as a thin connection between the homologs called a *chiasma*.

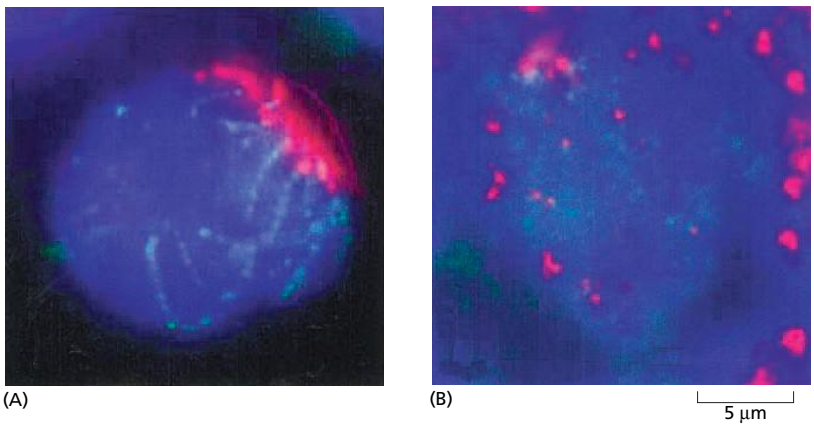


Figure 21–7 Rearrangements of telomeres during prophase in developing bovine eggs. The nucleus is stained *blue* and the telomeres are stained *red*. During prophase I, the telomeres are bound to the inner surface of the nuclear envelope. At first, they are dispersed around the nuclear envelope (not shown). Then they become clustered at one region of the envelope (A); finally, toward the end of prophase I, they disperse again (B). (From C. Pfeifer et al., *Dev. Biol.* 255:206–215, 2003. With permission from Elsevier.)

We have described the pairing of homologous autosomes during prophase I, but what happens to the sex chromosomes? This varies between different organisms. Female mammals have two X chromosomes, which pair and segregate like other homologs. But the males have one X and one Y chromosome. Although these chromosomes are not homologous, they too must pair and undergo crossing-over during prophase if they are to segregate normally at anaphase I. Pairing, crossing-over, and segregation are possible because of a small region of homology between the X and the Y at one or both ends of these chromosomes. The two chromosomes pair and crossover in this region during prophase I, ensuring that each sperm receives either one Y or one X chromosome and not both or neither. Thus, only two types of sperm are normally produced: those containing one Y chromosome, which will give rise to male embryos, and those containing one X chromosome, which will give rise to female embryos.

Homolog Pairing Culminates in the Formation of a Synaptonemal Complex

The paired homologs are brought into closer juxtaposition, with their structural axes (*axial cores*) about 400 nm apart, by a mechanism that depends in most species on the programmed double-strand DNA breaks that occur in sister chromatids. What pulls the axes together? One possibility is that the large protein machine, called a *recombination complex*, which assembles on a double-strand break in a chromatid, binds the matching DNA sequence in the nearby homolog and helps reel in this partner. This so-called *presynaptic alignment* of the homologs is followed by *synapsis,* in which the axial core of a homolog becomes tightly linked to the axial core of its partner by a closely packed array of *transverse filaments* to create a **synaptonemal complex**, which bridges the gap, now only 100 nm, between the homologs (**Figure 21–8**). Although crossing-over begins before the synaptonemal complex assembles, the final steps occur while the DNA is held in the complex (discussed in Chapter 5).

The morphological changes that occur during the pairing of meiotic chromosomes are the basis for dividing prophase I into five sequential stages—leptotene, zygotene, pachytene, diplotene, and diakinesis. As shown in **Figure 21–9**, prophase I starts with *leptotene,* when homologs condense and pair, and genetic recombination begins. At *zygotene,* the synaptonemal complex begins to assemble in local regions along the homologs; assembly initiates at sites where the homologs are closely associated and recombination events are occurring. At *pachytene,* the assembly process is complete, and the homologs are synapsed along their entire lengths. The pachytene stage can persist for days or longer, until desynapsis begins at *diplotene* with the disassembly of the synaptonemal complexes and the concomitant condensation and shortening of the chromosomes. It is only at this stage, after the complexes have disassembled, that the

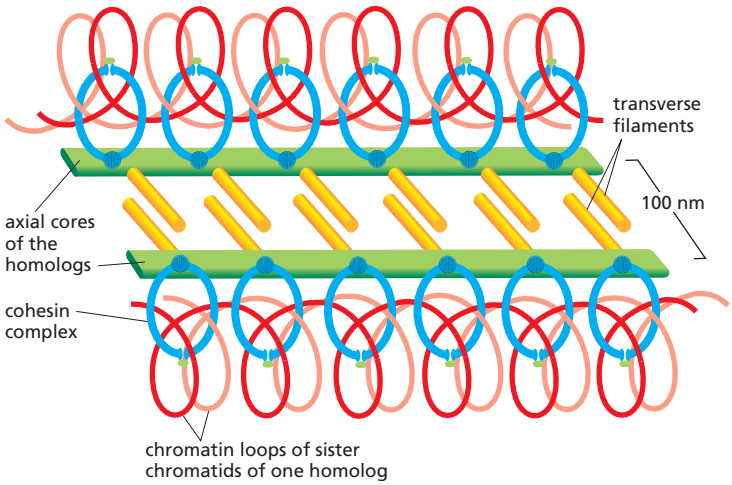
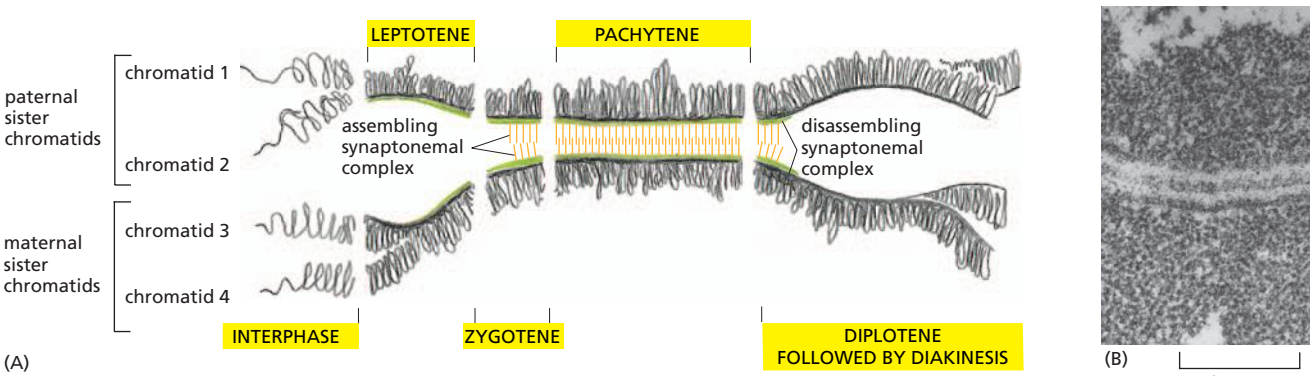


Figure 21–8 Simplified schematic drawing of a synaptonemal complex. Before the synaptonemal complex forms, recombination complexes assemble on double-strand DNA breaks on sister chromatids and help catalyze crossing-over between nonsister chromatid loops from opposite sides of the complex (not shown). (Modified from K. Nasmyth, *Annu. Rev. Genet.* 35:673–745, 2001. With permission from Annual Reviews.)



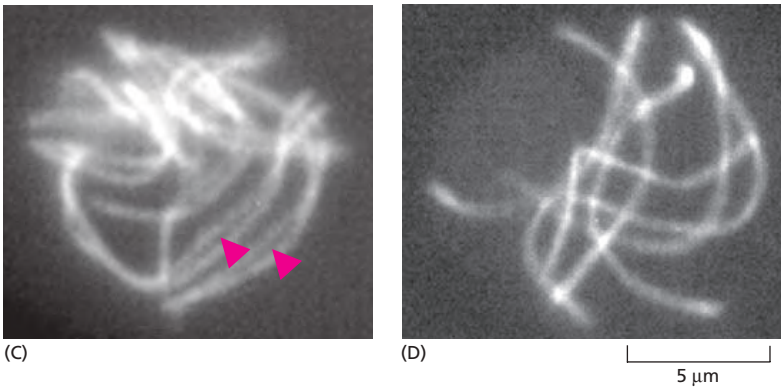


Figure 21–9 Homolog synapsis and desynapsis during the different stages of prophase I. (A) A single bivalent is shown schematically. At leptotene, the two sister chromatids coalesce, and their chromatid loops extend out together from a common axial core. The synaptonemal complex begins to assemble focally in early zygotene. Assembly continues through zygotene and is complete in pachytene. The complex disassembles in diplotene. (B) An electron micrograph of a synaptonemal complex from a meiotic cell at pachytene in a lily flower. (C and D) Immunofluorescence micrographs of prophase I cells of the fungus *Sordaria*. Partially synapsed bivalents at zygotene are shown in (C) and fully synapsed bivalents ore shown in (D). *Red arrowheads* in (C) point to regions where synapsis is still incomplete. (B, courtesy of Brian Wells; C and D, from A. Storlazzi et al., *Genes Dev.* 17:2675–2687, 2003. With permission from Cold Spring Harbor Laboratory Press.)

individual crossover events between nonsister chromatids can be seen as interhomolog connections called **chiasmata** (singular **chiasma**), which now play a crucial part in holding the compact homologs together (**Figure 21–10**). The homologs are now ready to begin the process of segregation. Prophase I ends with *diakinesis*—the stage of transition to metaphase I.

The proteins that form the transverse filaments that bridge between the axial cores of the homologs have been identified in several species, including yeasts, worms, flies, and mammals. They form homodimers that interact with each other across the 100 nm gap between the homologs, as illustrated in **Figure 21–11**. In most eucaryotes, these proteins are important for crossing-over, as mutants that lack them fail to form crossovers. The cohesin complexes that assemble on the DNA during S phase and bind the sister chromatids together during meiosis are major components of the axial core of each homolog (see Figure 21–8). Some of the cohesin subunits that operate in meiosis are the same as those that function in mitosis, whereas others are specific for meiosis. Both the crossovers and the cohesin complexes play crucial parts in segregating the homologs during meiotic division I, as we now discuss.

Homolog Segregation Depends on Meiosis-Specific, Kinetochore-Associated Proteins

One fundamental difference between meiosis I and mitosis (and meiosis II) is that in meiosis I homologs rather than sister chromatids separate and then segregate

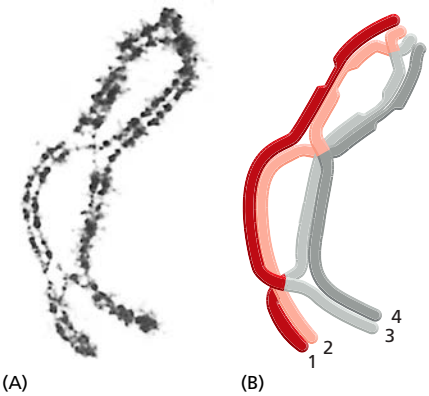


Figure 21–10 A bivalent with three chiasmata resulting from three crossover events. (A) Light micrograph of a grasshopper bivalent. (B) Drawing showing the arrangement of the crossovers in (A). Note that chromatid 1 has undergone an exchange with chromatid 3, and chromatid 2 has undergone exchanges with chromatids 3 and 4. Note also how the combination of the chiasmata and the tight attachment of the sister chromatid arms to each other (mediated by cohesin complexes) hold the two homologs together after the synaptonemal complex has disassembled; if either the chiasmata or the sister-chromatid adhesion failed to form, the homologs would come apart at this stage and not be segregated properly when the cell divides at the end of meiosis I. (A, courtesy of Bernard John.)

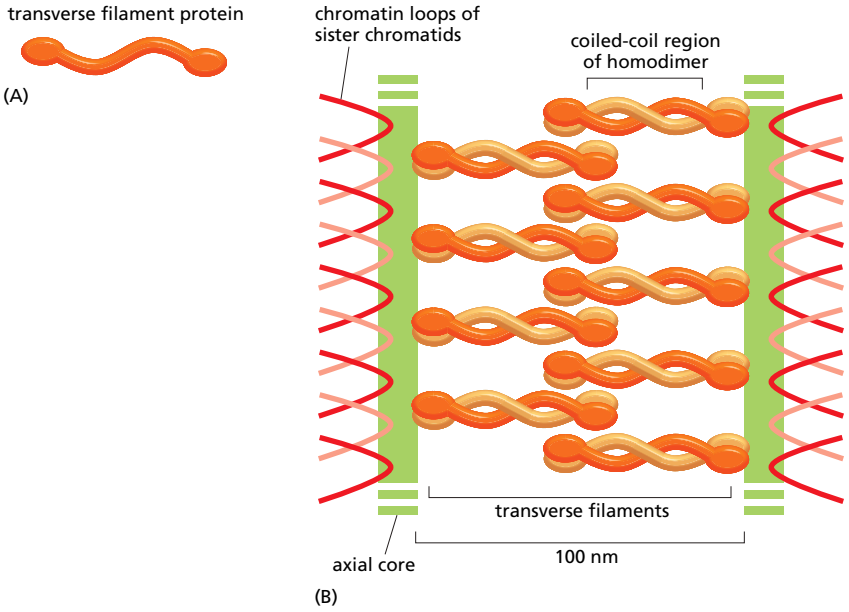


Figure 21–11 A molecular model of how transverse filaments may be formed by a single type of protein. (A) A diagram of the polypeptide chain showing the N- and C-terminal globular domains, connected by a coiled-coil region. (B) It is proposed that the protein forms homodimers, which then interact across the 100 nm gap separating the axial cores of the two homologs. (Adapted from S.L. Page and R.S. Hawley, *Science* 301:785–789, 2003. With permission from AAAS.)

into the two daughter cells (see Figure 21–5). This difference depends on three features of meiosis I that distinguish it from mitosis (**Figure 21–12**). First, the *kinetochores* (protein complexes associated with the centromeres, discussed in

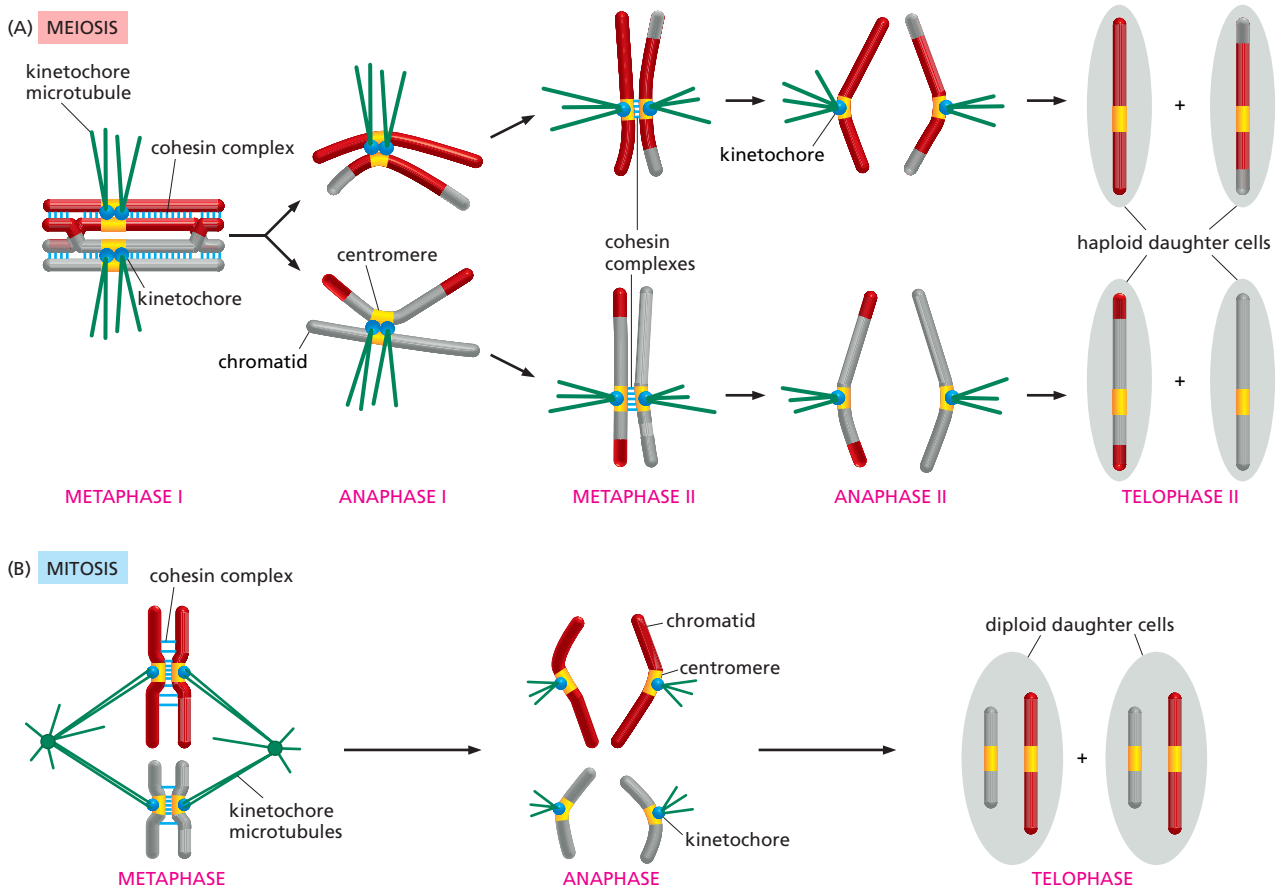


Figure 21–12 Comparison of chromosome behavior in meiosis I, meiosis II, and mitosis. Chromosomes behave similarly in mitosis and meiosis II, but they behave very differently in meiosis I. (A) In meiosis I, the two sister kinetochores are located side-by-side on each homolog at the sister centromeres and attach to microtubules emanating from the same spindle pole. The proteolytic destruction of the cohesin complexes along the sister chromatid arms unglues the arms and resolves the crossovers, allowing the duplicated homologs to separate at anaphase I, while the residual cohesin complexes at the centromeres keep the sisters together. The proteolytic destruction of the residual cohesin complexes at the centromeres allows the sister chromatids to separate at anaphase II. (B) In mitosis, by contrast, the two sister kinetochores attach to microtubules emanating from different spindle poles, and the two sister chromatids come apart at the start of anaphase and segregate into separate daughter cells (discussed in Chapter 17).

Chapters 4 and 17) on the two sister chromatids of a homolog attach to microtubules emanating from the same pole of the meiosis I spindle and thus segregate together into the same daughter cell at anaphase I; this contrasts with mitosis (and meiosis II), in which the kinetochores on the two sister chromatids of a chromosome attach to opposite poles of the spindle and therefore segregate into different daughter cells at anaphase. Second, a strong physical linkage is maintained between the homologs that resists the pulling forces of the meiosis I spindle until the bivalents are aligned at the equator of the spindle and the homologs separate at anaphase I. The chiasmata formed between nonsister chromatids and the cohesion between sister-chromatid arms cooperate in holding the homologs together (see Figure 21–10). Third, the arms of the sister chromatids separate at anaphase I, resolving the chiasmata and allowing the homologs to separate, but the sisters stay glued together in the region of their centromeres until anaphase II and therefore do not separate in anaphase I.

In micromanipulation experiments, meiosis I chromosomes transferred to a meiosis II spindle behave as they do in meiosis I, indicating that the specialized behavior of meiosis I chromosomes is determined by the chromosomes themselves rather than by the spindle or other cytoplasmic factors. Various meiosis-specific proteins associated with meiosis I chromosomes account for the special behavior, although they work together with non-meiosis-specific proteins that help mediate both mitosis and meiosis. Meiosis-specific protein complexes, for example, associate with the two kinetochores on each replicated homolog and help ensure that the two sister chromatids attach to microtubules emanating from a single spindle pole. Other proteins (called *shugoshins*) associated with kinetochores help ensure that sister kinetochores do not come apart at anaphase I when the proteolytic enzyme *separase* (discussed in Chapter 17) cleaves the cohesin complexes that tie the arms of sister chromatids together. One way that the shugoshins protect the cohesin complexes at centromeres is by recruiting a specific protein phosphatase to the centromeres; the phosphatase reverses the phosphorylation of the cohesin complexes that is required for separase to cleave them. Thus, whereas the chromatid arms come apart at anaphase I, the centromeres do not. The sisters separate only when separase cleaves the remaining cohesin complexes at the centromeres at anaphase II (see Figure 21–12A), when the shugoshins are gone.

Unlike meiosis I, meiosis II occurs rapidly and closely resembles a mitotic cell division, although it occurs without DNA replication. Prophase II is brief: the nuclear envelope breaks down as the new spindle forms, after which metaphase II, anaphase II, and telophase II usually follow in quick succession. After nuclear envelopes have formed around the four haploid nuclei produced at telophase II, cytokinesis occurs, and meiosis is complete.

Meiosis Frequently Goes Wrong

The sorting of chromosomes that takes place during meiosis is a remarkable feat of intracellular bookkeeping. In humans, each meiosis requires that the starting cell keep track of 92 chromatids (46 chromosomes, each of which has duplicated), distributing one complete set of each type of chromosome to each of the four haploid progeny cells. Not surprisingly, mistakes can occur in allocating the chromosomes during this elaborate process. Mistakes are especially common in human female meiosis, which arrests for years after diplotene: meiosis I is completed only at *ovulation*, and meiosis II only after the egg is fertilized. Indeed, such chromosome segregation errors during egg development are the commonest cause of both spontaneous abortion (miscarriage) and mental retardation in humans.

When homologs fail to separate properly—a phenomenon called **nondisjunction**—the result is that some of the haploid gametes produced lack a particular chromosome, while others have more than one copy of it. (Cells with an abnormal number of chromosomes are said to be *aneuploid*, whereas those with the correct number are said to be *euploid*.) Upon fertilization, aneuploid gametes form abnormal embryos, most of which die. Some survive, however. *Down syndrome* in humans, for example, which is the leading single cause of

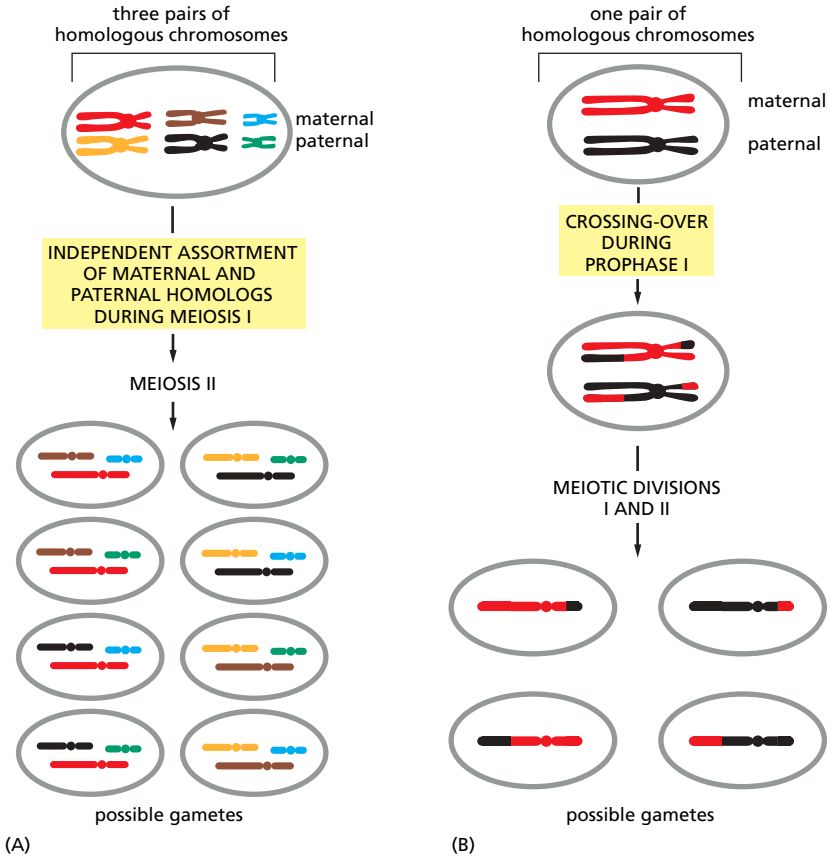


Figure 21–13 Two major contributions to the reassortment of genetic material that occurs in the production of gametes during meiosis. (A) The independent assortment of the maternal and paternal homologs during meiosis produces 2^n different haploid gametes for an organism with n chromosomes. Here $n = 3$, and there are 8 different possible gametes. (B) Crossing-over during prophase I exchanges DNA segments between homologous chromosomes and thereby re-assorts genes on individual chromosomes. Because of the many small differences in DNA sequence that always exist between any two homologs, both mechanisms increase the genetic variability of organisms that reproduce sexually.

mental retardation, is caused by an extra copy of chromosome 21, usually resulting from nondisjunction during meiosis I in the female ovary. Segregation errors during meiosis I increase greatly with advancing maternal age.

Despite its fallibility, almost all eucaryotes use meiosis, intermittently at least, to shuffle their genetic information before passing it on to the next generation. Crossing-over makes a major contribution to this genetic shuffling process, as we now discuss.

Crossing-Over Enhances Genetic Reassortment

Unless they are identical twins, which develop from a single zygote, no two offspring of the same parents are genetically the same. As we discussed earlier, this is because, long before the two gametes fuse at fertilization, two kinds of randomizing genetic reassortment have occurred in meiosis I, during the production of the gametes: the random distribution of maternal and paternal homologs, and crossing-over. The random distribution of maternal and paternal homologs (**Figure 21–13**A) could, in principle, produce 2^n genetically different gametes, where n is the haploid number of chromosomes. In humans, for example, each individual can produce at least $2^{23} = 8.4 \times 10^6$ genetically different gametes. But the actual number of variants is very much greater than this because of **chromosomal crossing-over** (or simply **crossing-over**), which is an outcome of homologous recombination (discussed in Chapter 5), in which DNA segments of homologous chromosomes are exchanged. In meiosis, when the exchange occurs between nonsister chromatids, it mixes the genetic constitution of each of the chromosomes (Figure 21–13B). On average, between two and three crossovers occur between each pair of human homologs (**Figure 21–14**).

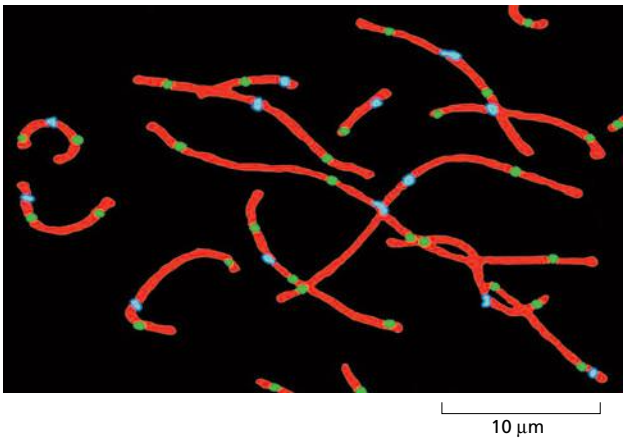


Figure 21–14 Crossovers between homologs in the human testis. In these immunofluorescence micrographs, antibodies have been used to stain the synaptonemal complexes (*red*), the centromeres (*blue*), and the sites of crossing-over (*green*). Note that all of the bivalents have at least 1 crossover and none have more than 3. (Modified from A. Lynn et al., *Science* 296:2222–2225, 2002. With permission from AAAS.)

The molecular details of crossing-over are discussed in Chapter 5 (see Figure 5–64). Briefly, a conserved meiosis-specific protein called *Spo11* initiates crossing-over by creating a double-strand break in the DNA of either a maternal or a paternal chromatid. A very large multienzyme *recombination complex,* containing double-strand DNA repair enzymes, assembles on the break and catalyzes homologous recombination. In most cases, these events do not result in a crossover. In some cases, however, homologous recombination leads to a *crossover,* where DNA segments are exchanged between two nonsister chromatids in a reciprocal fashion. As discussed earlier, after desynapsis, each crossover can be seen in the microscope as a chiasma (see Figure 21–10A). As illustrated in Figure 21–10B, each of the two sister chromatids of a homolog can form one or more crossovers with either of the two sister chromatids of its partner homolog.

Crossing-Over Is Highly Regulated

Crossing-over has two distinct functions in meiosis: it helps hold homologs together so that they are properly segregated to the two daughter cells produced by meiosis I, and it contributes to the genetic diversification of the gametes that are eventually produced. As might be expected, therefore, crossing-over is highly regulated: the number and location of double-strand breaks along each chromosome is controlled, as is the likelihood that a break will be converted into a crossover. Although the double-strand breaks that occur in meiosis I can be located almost anywhere along the chromosome (see Figure 21–14), they are not distributed uniformly: they cluster at "hot spots", where the chromatin is accessible, and occur only rarely in "cold spots", such as the heterochromatin regions around centromeres and telomeres.

At least two kinds of regulation influence the location and number of crossovers that form, neither of which is well understood. Both operate before the synaptonemal complex assembles. One ensures that at least one crossover forms between the members of each homolog pair, as is necessary for normal homolog segregation in meiosis I. In the other, called *crossover interference*, the presence of one crossover event inhibits another from forming close by, perhaps by locally depleting proteins required for converting a double-strand DNA break into a stable crossover.

Meiosis Is Regulated Differently in Male and Female Mammals

The basic mechanisms of meiosis have been conserved in evolution in all sexually reproducing eucaryotes. In all of these organisms, for example, most of meiosis is spent in prophase I, although the details of the timing of different stages vary among organisms (**Figure 21–15**). There are, however, some remarkable differences in the regulation of meiosis in different species and in different sexes of the same species. The difference between the two sexes is very striking in mammals.

In mammalian females, egg precursor cells *(oocytes)* begin meiosis in the fetal ovary but arrest after diplotene, after the synaptonemal complex has disassembled in meiosis I. They complete meiosis I only after the female has become sexually mature and the oocyte is released from the ovary during *ovulation*; moreover, the released oocyte completes meiosis II only if it is fertilized. Thus, there are special stop and start mechanisms during meiosis in female mammals. In humans, some oocytes remain arrested in meiosis I for 40 years or more, which is presumably at least part of the reason why nondisjunction increases dramatically in older women. In mammalian males, by contrast, meiosis only begins in sperm precursor cells *(spermatocytes)* in the testes at puberty and then goes on continuously, without the stop and start mechanisms that operate in female meiosis. It takes about 24 days for a human spermatocyte to complete meiosis.

There is also a big difference in the error rates of meiosis in mammalian females and males, and this is especially striking in humans. About 20% of human eggs are aneuploid, compared with 3–4% of human sperm, and, largely as a result of this, up to 25% of all human fetuses are aneuploid, and most of these result from nondisjunction in oocytes at meiosis I. Mammalian fertilization typically involves the ovulation of a small number of eggs at one end of the female reproductive tract and the entry of millions of sperm at the other. Given the relative scarcity of eggs, one might have expected that egg development would be subject to more stringent quality control than sperm development, but the opposite is the case. If meiosis goes wrong in male cells, a cell-cycle checkpoint mechanism (discussed in Chapter 17) is activated, which arrests meiosis and leads to cell death by apoptosis. Such checkpoint mechanisms apparently do not operate in female meiotic cells: if homolog segregation fails to occur normally, the cells continue through meiosis and produce aneuploid eggs. The male germ line, on the other hand, is thought to be the main source of another type of genetic error. Because many more mitotic cell divisions occur on the way to the production of a sperm, and each round of DNA replication is liable to error, the average number of new mutations contributed by fathers is larger than the number contributed by mothers.

The production of gametes involves more than just meiosis, and the other processes also differ for eggs and sperm. As we will see, by the end of meiosis, a mammalian egg is fully mature, whereas a sperm that has completed meiosis has only just begun its differentiation. Before discussing these gametes, however, we first consider how certain cells in the mammalian embryo initially become specified to develop into germ cells and how these cells then become committed to developing into either sperm or eggs, depending on the sex of the individual.

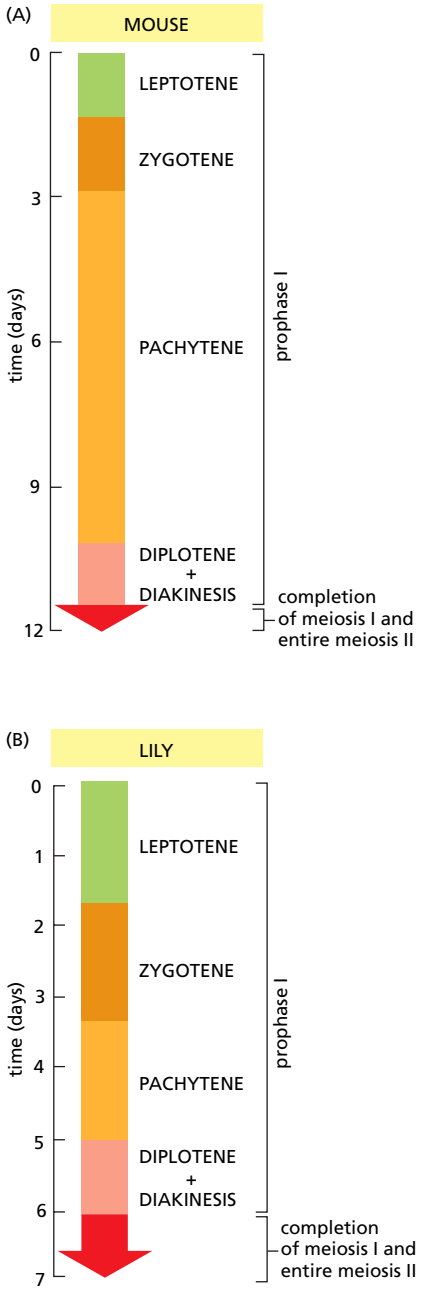


Figure 21–15 Comparison of times required for each of the stages of meiosis. (A) Approximate times for a male mammal (mouse). (B) Approximate times for the male tissue of a plant (lily). Times differ for male and female gametes (sperm and eggs, respectively) of the same species, as well as for the same gametes of different species. Meiosis in a human male, for example, lasts for 24 days, compared with 12 days in the mouse. In human females, it can last 40 years or more, because meiosis I arrests after diplotene. In all species, however, prophase I is always much longer than all the other meiotic stages combined.

Summary

Haploid gametes (eggs, sperm, pollen, and spores) are produced by meiosis, in which two successive cell divisions follow one round of DNA replication to give rise to four haploid cells from a single diploid cell. Meiosis is dominated by a prolonged prophase I, which can occupy 90% or more of the total meiotic period. At the start of prophase I, the chromosomes have replicated and consist of two tightly joined sister chromatids. Homologous chromosomes (homologs) then pair up side-by-side and become progressively more closely juxtaposed as prophase I proceeds. The tightly aligned homologs (bivalents) undergo genetic recombination, forming crossovers that can later be seen as chiasmata, which help hold each pair of homologs together during metaphase I. Both crossing-over and the independent segregation of the maternal and paternal copies of each chromosome during meiosis I have important roles in producing gametes that are genetically different from one another and from both parents. Meiosis-specific, kinetochore-associated proteins help ensure that both sister chromatids in a homolog attach to the same spindle pole; other kinetochore-associated proteins ensure that the homologs remain connected at their centromeres during anaphase I, so that homologs rather than sister chromatids are segregated in meiosis I. After the long meiosis I, meiosis II follows rapidly, without DNA replication, in a process that resembles mitosis, in that sister chromatids are pulled apart at anaphase.

PRIMORDIAL GERM CELLS AND SEX DETERMINATION IN MAMMALS

Sexual reproductive strategies vary enormously among different organisms. In the rest of this chapter, we focus mainly on the strategies used by mammals.

In all vertebrate embryos, certain cells are singled out early in development as progenitors of the gametes. These diploid **primordial germ cells** (**PGCs**) migrate to the developing gonads, which will form the *ovaries* in females and the *testes* in males. After a period of mitotic proliferation in the developing gonads, the PGCs undergo meiosis and differentiate into mature haploid gametes—either eggs or sperm. Later, the fusion of egg and sperm after mating initiates embryogenesis. The subsequent production of new PGCs in this new embryo begins the cycle again (see Figure 21–3A).

In this section, we consider how mammalian PGCs arise, how the sex of a mammal is determined, and how sex determination dictates whether the PGCs develop into sperm or eggs.

Signals from Neighbors Specify PGCs in Mammalian Embryos

In many animals, including many vertebrates, the unfertilized egg contains specific molecules localized in a particular region of the cytoplasm that determine which cells will become the germ cells. When the egg is fertilized and divides repeatedly to produce the cells of the early embryo, the cells that ultimately inherit these *germ cell determinants* become PGCs (**Figure 21–16**). Although the molecular nature and functions of the determinants are largely unknown, proteins of the *Vasa* family are a required component in all of these animals. Vasa proteins are structurally similar to ATP-dependent RNA helicases, but their precise function in germ cell determination remains a mystery.

By contrast, in other animals, including mammals, the egg cytoplasm does not contain localized germ cell determinants. Instead, signals from neighboring cells dictate which cells become PGCs. In mammals, all cells produced by the first few divisions of the fertilized egg are *totipotent*—that is, they have the potential to give rise to any of the cell types of the animal, including the germ cells and cells of extraembryonic tissues such as the placenta. Only later is a small group of cells induced by signals from neighbors to become PGCs. In mice, for instance, about 6 days after fertilization, signals (including bone morphogenic protein 4, BMP4) secreted by cells in tissue lying outside the embryo proper induce about 10 cells in the adjacent underlying embryo to become PGC precursors. These cells divide and mature to become PGCs, turning off the expression of a number of somatic cell genes and turning on the expression of genes involved in maintaining the special character of the germ cells.

Although different mechanisms specify PGCs in different animals, some of the mechanisms that control their proliferation and development have been

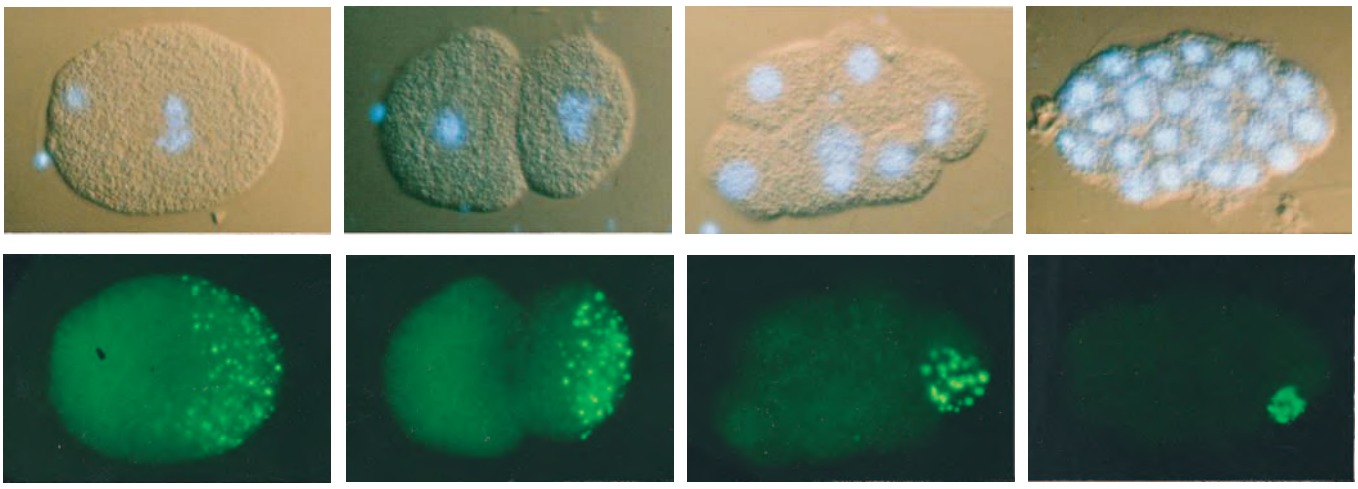


Figure 21–16 Segregation of germ cell determinants in the nematode *C. elegans*. The micrographs in the upper row show the pattern of cell divisions, with the cell nuclei stained *blue*; below, the same cells are stained with an antibody that labels (in *green*) small granules (called *P granules*) that function as germ cell determinants. The P granules are composed of RNA and protein molecules and are distributed randomly throughout the cytoplasm of the unfertilized egg (not shown). As shown in the far left-hand panels, after fertilization, the granules accumulate at one pole of the zygote. The granules are then segregated into one of the two daughter cells at each cell division. The single cell containing the P granules in the embryo shown in the far right-hand panels is the precursor of the germ line. (Courtesy of Susan Strome.)

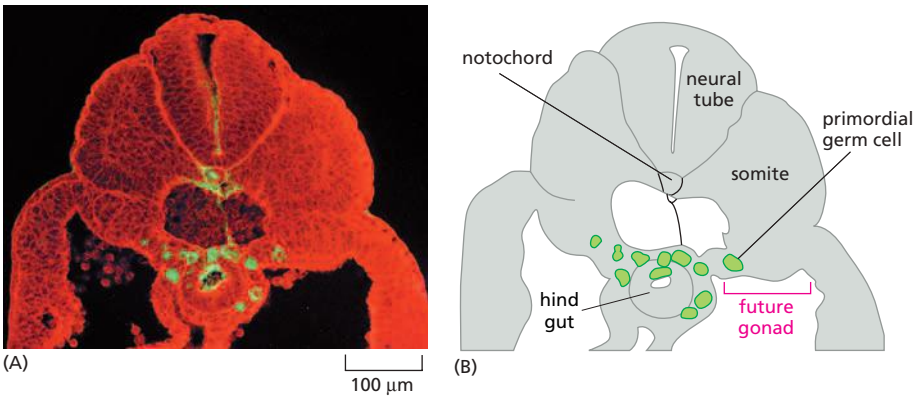


Figure 21–17 Migration of mammalian PGCs. (A) Fluorescence micrograph showing migrating PGCs in a cross section of an early mouse embryo. The PGCs are stained with a monoclonal antibody (in *green*) that specifically labels these cells at this stage of embryogenesis. The remaining cells in the embryo are stained with a lectin (in *red*) that binds to sialic acid, which is found on the surface of all cells. (B) Drawing corresponding to the micrograph shown in (A). (A, courtesy of Robert Anderson and Chris Wylie.)

conserved in evolution from worms to humans. For example, PGC development in all animals that have been studied depends on the suppression of somatic cell fates by gene repression, as well as on the inhibition of translation of specific mRNAs by *Nanos* RNA-binding proteins.

PGCs Migrate into the Developing Gonads

After mammalian PGCs develop, they proliferate and migrate to their final destination in the developing gonads (**Figure 21–17**). As they migrate through the embryo, various extracellular signal proteins produced by adjacent somatic cells signal them to survive, proliferate, and migrate. Among the secreted signal proteins that help attract PGCs into the developing gonad are *chemokines*, which bind to G-protein-coupled receptors (GPCRs) and guide the migration of various cell types, including PGCs and white blood cells (discussed in Chapter 23).

After the PGCs enter the developing gonad, which at this stage is called the *genital ridge,* they go through several more mitotic cell divisions, in the course of which they become specified to follow a pathway that will lead them to develop as either eggs or sperm.

When the PGCs first migrate into the embryonic gonad, however, they are not irreversibly committed to becoming gametes. When removed from the embryo and cultured in the presence of appropriate extracellular signal proteins, they convert into cells that can be maintained in culture indefinitely as a cell line that can produce any of the cell types of the body of the animal, although not the extraembryonic cells that go on to form structures such as the placenta; for this reason, they are said to be *pluripotent*, rather than totipotent. In these respects, these so-called *embryonic germ (EG) cells* resemble *embryonic stem (ES)* cells (discussed in Chapter 23). EG and ES cells are promising sources of various human cell types—both for drug testing and for the treatment of diseases such as heart attacks, strokes, and various neurodegenerative diseases, in which specific cell types die.

What determines whether the PGCs that migrate into the developing mammalian gonad develop into eggs or sperm? Surprisingly, it is not their own sex chromosome constitution but rather whether the genital ridge has begun to develop into an ovary or a testis, respectively. The sex chromosomes in the somatic cells of the genital ridge determine which type of gonad the ridge becomes. Although many genes influence the outcome, a single gene on the Y chromosome has an especially important role.

The *Sry* Gene Directs the Developing Mammalian Gonad to Become a Testis

Aristotle believed that the temperature of the male during sexual intercourse determined the offspring's sex: the higher the temperature, the greater the chance of producing a male. Had he been referring to lizards or alligators rather than to humans, he would have been closer to the truth, as in many egg-laying

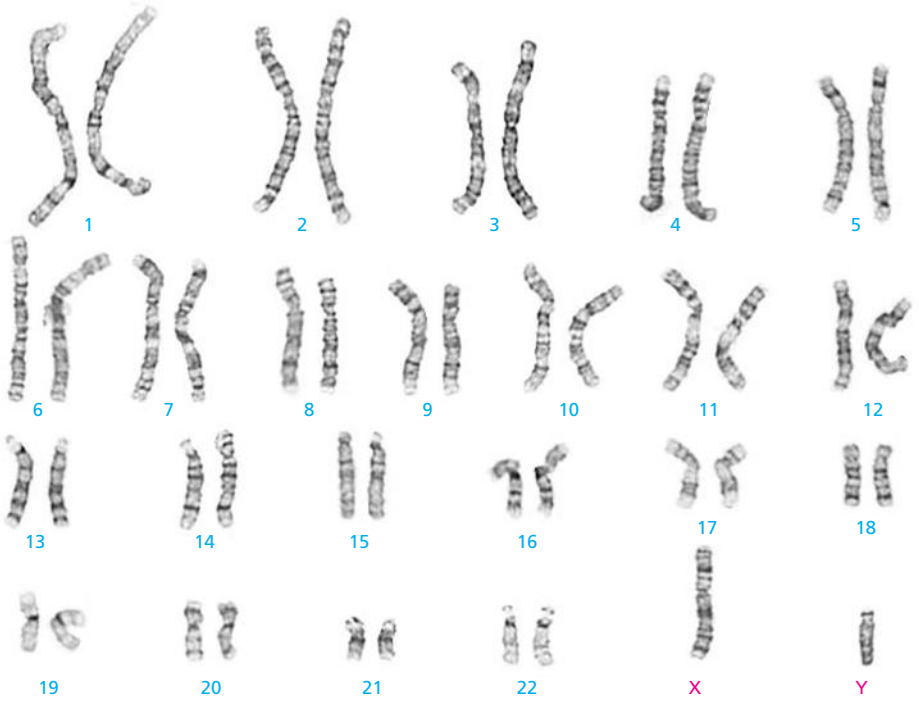


Figure 21–18 Chromosomes of a normal human male. The chromosomes have been labeled with Giemsa stain. See also Figures 4–10 and 4–11. Note the difference in size of the two sex chromosomes. Whereas the X chromosome contains more than 1000 genes, the Y chromosome contains only about 80. (Courtesy of Julie Robertson of the Wisconsin State Laboratory of Hygiene.)

reptiles the temperature of the incubating egg determines the sex of the offspring; in lizards and alligators, males develop at warm temperatures and females develop at cool temperatures. We now know, however, that sex chromosomes, rather than the temperature of the parents or embryo, determine the sex of a mammal.

Female mammals have two X chromosomes in all of their somatic cells, whereas males have one X and one Y. The presence or absence of the Y chromosome, which is the smallest human chromosome (**Figure 21–18**), determines the sex of the individual. Individuals with a Y chromosome develop as males no matter how many X chromosomes they have, whereas individuals without a Y chromosome develop as females, even if they have only one X chromosome. The sperm that fertilizes the egg determines the sex of the resulting zygote: eggs have a single X chromosome, whereas sperm can have either an X or a Y.

The Y chromosome influences the sex of the individual by directing the somatic cells of the genital ridge to develop into a testis instead of an ovary. Mammalian embryos are programmed to develop as females unless prevented from doing so by the testes, which direct the embryo to develop into a male. If the genital ridges are removed before they have started to develop into a testis or an ovary, a mammal develops into a female, regardless of the sex chromosomes it carries. This does not mean that signals are not required for the development of female-specific organs in mammals: the secreted signal protein Wnt4, for example, is required for normal mammalian ovary development.

The crucial gene on the Y chromosome that directs the genital ridge to develop into a testis instead of an ovary is called ***Sry***, for "sex-determining region of Y." Remarkably, when this gene is introduced into the genome of an XX mouse zygote, the transgenic embryo produced develops as a male, even though it lacks all of the other genes on the Y chromosome. Such *sex-reversed* mice, however, cannot produce sperm, because they lack the other genes on the Y chromosome that are required for sperm development. Similarly, XY humans with an inactivating mutation in *Sry* develop as females, even though they are otherwise genetically male.

Sry is expressed in a subpopulation of the somatic cells of the developing gonad, and it causes these cells to differentiate into **Sertoli cells**, which are the main type of supporting cells in the testis (see Figure 21–29). The Sertoli cells then direct sexual development along a male pathway by influencing other cells

in the genital ridge and elsewhere in the embryo in at least four ways:

1. They stimulate the newly arriving PGCs to develop along a pathway that produces sperm. They do so by inhibiting the cells from entering meiosis and developing along the pathway that produces eggs, as we discuss later.
2. They secrete *anti-Müllerian hormone,* which circulates in the blood and suppresses the development of the female reproductive tract by causing the Müllerian duct to regress (this duct otherwise gives rise to the oviduct, uterus, and upper part of the vagina).
3. They stimulate endothelial and smooth muscle cells in adjacent mesenchymal tissue to migrate into the developing gonad. These cells form critical elements within the testis that are required for normal sperm production, which begins when the organism reaches sexual maturity.
4. They help to induce other somatic cells in the developing gonad to become *Leydig cells,* which secrete the male sex hormone *testosterone* into the blood. The secretion of testosterone is responsible for inducing all male secondary sexual characteristics, including the structures of the male reproductive tract, such as the prostate and seminal vesicles, which develop from another duct called the Wolffian duct system. This duct system degenerates in the developing female because it requires testosterone to survive and develop. Testosterone secretion also helps masculinize the early developing brain, influencing sexual identity and sexual orientation, and thereby sexual behavior: female rats that are treated with testosterone around birth, for example, later display malelike sexual behavior.

The *Sry* gene encodes a gene regulatory protein (Sry) that binds to DNA and influences the transcription of other genes involved in Sertoli cell development. One crucial downstream gene encodes another gene regulatory protein related to Sry, which is called Sox9. The *Sox9* gene is not on the Y chromosome, but it is expressed in males in all vertebrates, unlike *Sry*, which is found only in mammals. If *Sox9* is expressed ectopically in the developing gonads of an XX mouse embryo, the embryo develops as a male, even if it lacks the *Sry* gene, suggesting that Sry normally acts by inducing the expression of *Sox9*. The Sox9 protein directly activates the transcription of at least some Sertoli-cell-specific genes, including the gene encoding anti-Müllerian hormone.

In the absence of either *Sry* or *Sox9*, the genital ridge of an XY embryo develops into an ovary instead of a testis. The supporting cells become *follicle cells* instead of Sertoli cells. Other somatic cells become *theca cells* (instead of Leydig cells), which, beginning at puberty, contribute to the production of the female sex hormone *estrogen*. The PGCs develop along a pathway that produces eggs rather than sperm (**Figure 21–19**), and the animal develops as a female.

How do Sertoli cells induce the PGCs that migrate into the developing gonad in males to follow the pathway leading to sperm production rather than to egg production? The mechanism depends on the small signal molecule *retinoic acid* (see Figure 15–13), which, in both sexes, is produced by cells in a transient tubular structure called the mesonephros that lies adjacent to the developing gonad. In the embryonic ovary, the retinoic acid induces the proliferating germ-line cells to enter meiosis and start down the pathway leading to egg production; the cells become arrested after diplotene of prophase I, where they remain until ovulation, beginning when the female reaches sexual maturity. In the embryonic testis, by contrast, Sertoli cells produce an enzyme that degrades retinoic acid, preventing the retinoic acid from the mesonephros from inducing the germ-line cells to enter meiosis and begin egg development. Only much later, when the male becomes sexually mature, do the germ-line cells in the testis begin producting sperm.

Many Aspects of Sexual Reproduction Vary Greatly between Animal Species

Although meiosis is highly conserved in all sexually reproducing eucaryotes, other aspects of sexual reproduction are extremely variable. We have seen that

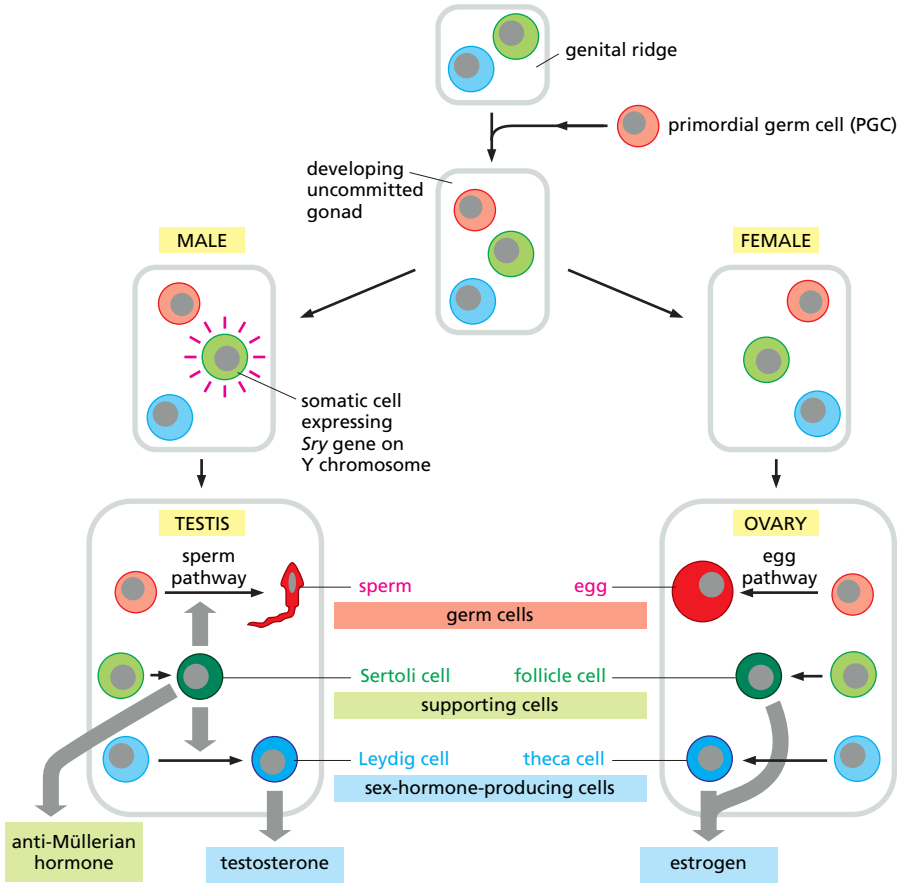


Figure 21–19 Influence of *Sry* on gonad development. The germ-line cells are shaded in *red,* and the somatic cells are shaded in *green* or *blue*. The change from light to darker color indicates that the cell has differentiated. The *Sry* gene acts in a subpopulation of somatic cells in the developing gonad to direct them to differentiate into Sertoli cells instead of follicle cells. The Sertoli cells then prevent the germ-line cells from developing along the egg pathway and help direct them down the sperm pathway of development, beginning at puberty. They also secrete anti-Müllerian hormone, which causes the Müllerian duct to regress, and they help to induce other somatic cells to differentiate into Leydig cells, which secrete testosterone (see Figure 21–29). In the absence of *Sry,* the germ-line cells commit to egg development, and the somatic cells develop into either follicle cells, which support egg development, or theca cells, which produce progesterone; the progesterone is converted to estrogen by the follicle cells. Whereas the testis begins secreting testosterone in the fetus, the ovary does not begin secreting estrogen until puberty.

an animal's sex can depend on either its chromosomes or the environment in which it develops. But even the genetic mechanisms of sex determination vary greatly. In *C. elegans* and *Drosophila*, for example, sex is determined by the ratio of X chromosomes to autosomes, rather than by the presence or absence of a Y chromosome, as in mammals. In *C. elegans*, sex determination depends mainly on transcriptional and translational controls on gene expression, whereas in *Drosophila* it depends on a cascade of regulated RNA splicing events, as discussed in Chapter 7. In *Drosophila*, moreover, the sex-specific character of each cell in the body is programmed individually by its own chromosomes, instead of being controlled mainly by hormones. It remains a mystery why some aspects of sexual reproduction have been conserved in evolution, while others have become so fundamentally different.

Summary

A small number of cells in the early mammalian embryo are signaled by their neighbors to become germ-line cells. The resulting primordial germ cells (PGCs) proliferate and migrate into the developing gonads. Here, the germ-line cells commit to develop into either eggs, if the gonad is becoming an ovary, or sperm, if the gonad is becoming a testis. A developing gonad will develop into an ovary unless its somatic cells contain a Y chromosome, in which case it develops into a testis. The Sry gene on the mammalian Y chromosome is crucial for testis development: it is expressed in a subpopulation of somatic cells in the developing gonad and directs them to differentiate into Sertoli cells, which then produce signal molecules that promote the development of male characteristics, and suppress the development of female characteristics. Mammalian embryos are programmed to follow a female pathway of development unless they are diverted to follow the male pathway by Sertoli cells.

EGGS

In one respect at least, an egg is the most remarkable of animal cells: once activated, it can give rise to a complete new individual within a matter of days or weeks. No other cell in a higher animal has this capacity. Activation is usually the consequence of *fertilization*—fusion of a sperm with the egg—but eggs can also be activated artificially by various nonspecific chemical or physical treatments. Indeed, some organisms, including a few vertebrates such as some lizards, normally reproduce from eggs that become activated in the absence of sperm—that is, **parthenogenetically**. Mammals are the only animals that cannot be produced parthenogenetically; because of *genomic imprinting* (discussed in Chapter 7), they require both maternal and paternal genetic contributions.

Although an egg can give rise to every cell type in the adult organism, it is itself a highly specialized cell, uniquely equipped for the single function of generating a new individual. The cytoplasm of an egg can even reprogram a somatic cell nucleus so that the nucleus can direct the development of a new individual, although the egg components responsible are mostly unknown. That is how the famous sheep Dolly was produced. The nucleus of an unfertilized sheep egg was removed with a glass pipette and replaced with the nucleus of an adult somatic cell. An electric shock was used to activate the egg, and the resulting embryo was implanted in the uterus of a surrogate mother. The resulting adult sheep had the genome of the donor somatic cell nucleus and was therefore a *clone* of the donor sheep.

The same approach, called *reproductive cloning*, has been used to produce clones of various mammals, including mice, rats, cats, dogs, goats, pigs, cattle, and horses (see Figure 21–38). In all cases, the efficiency is low: most of the clones die before birth and less than 5% of them develop to adulthood, probably because the transplanted somatic nucleus is not completely reprogrammed and so expresses many genes inappropriately.

In this section, we briefly consider some of the specialized features of an egg, before discussing how it undergoes its final preparations for fertilization.



Figure 21–20 The actual sizes of three eggs. The human egg is 0.1 mm in diameter.

An Egg Is Highly Specialized for Independent Development

The eggs of most animals are giant single cells. They contain stockpiles of all the materials needed for initial development of the embryo until the stage at which the new individual can begin feeding. Before the feeding stage, the giant cell cleaves into many smaller cells, but no net growth occurs. The mammalian embryo is an exception. It can start to grow early by taking up nutrients from the mother via the placenta. Thus, a mammalian egg, although still a large cell, is not as large as the egg of a frog or bird, for example. Eggs are typically spherical or ovoid, with a diameter of about 0.1 mm in humans and sea urchins (whose feeding larvae are tiny), 1–2 mm in frogs and fishes, and many centimeters in birds and reptiles (**Figure 21–20**). A typical somatic cell, by contrast, has a diameter of only about 10–30 μm (**Figure 21–21**).

The egg cytoplasm usually contains nutritional reserves in the form of **yolk**, which is rich in lipids, proteins, and polysaccharides and is often contained within discrete structures called *yolk granules*. In some species, a membrane encloses each yolk granule. In eggs that develop into large animals outside the mother's body, yolk can account for more than 95% of the volume of the cell. In mammals, whose embryos are largely nourished by their mothers via the placenta, there is little, if any, yolk.

The **egg coat** is another peculiarity of eggs. It is a specialized form of extracellular matrix consisting largely of glycoproteins—some secreted by the egg and some by surrounding cells. In many species, the major coat is a layer immediately surrounding the egg plasma membrane; in nonmammalian eggs, such as those of sea urchins or chickens, it is called the *vitelline layer*, whereas in mammalian eggs it is called the *zona pellucida* (**Figure 21–22**). This layer protects the egg from mechanical damage, and in many eggs it also acts as a species-specific barrier to sperm, admitting only those of the same or closely related species.

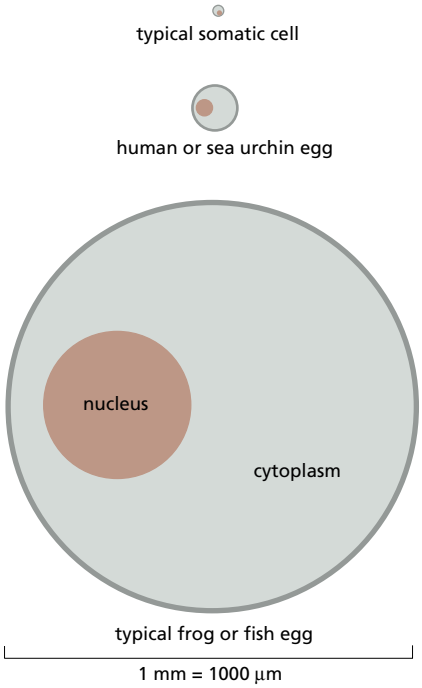


Figure 21–21 The relative sizes of various eggs, compared with that of a typical somatic cell.

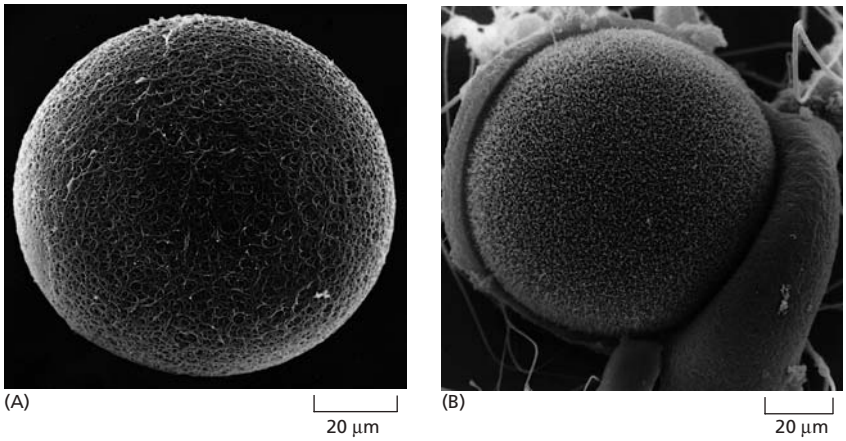


Figure 21–22 The zona pellucida. (A) Scanning electron micrograph of a hamster egg, showing the zona pellucida. (B) A scanning electron micrograph of a similar egg in which the zona (to which many sperm are attached) has been peeled back to reveal the underlying plasma membrane, which contains numerous microvilli. The zona is made entirely by the developing oocyte. (From D.M. Phillips, *J. Ultrastruct. Res.* 72:1–12, 1980. With permission from Elsevier.)

Many eggs (including those of mammals) contain specialized secretory vesicles just under the plasma membrane in the outer region, or *cortex*, of the egg cytoplasm. When a sperm activates the egg, these **cortical granules** release their contents by exocytosis; the contents of the granules alter the egg coat so as to help prevent more than one sperm from fusing with the egg.

Cortical granules are usually distributed evenly throughout the egg cortex. In many organisms, however, some egg cytoplasmic components have a strikingly asymmetrical distribution. Some of these localized components later serve as germ cell determinants (see Figure 21–16) or help establish the polarity of the embryo, as discussed in Chapter 22.

Eggs Develop in Stages

A developing egg, or **oocyte**, differentiates into a mature egg (or *ovum*) through a progressive series of changes. The timing of these changes is coordinated with the steps of meiosis, in which the germ cells go through their two final, highly specialized, divisions. As discussed earlier, oocytes arrest for a prolonged period in meiosis I while they grow in size and progressively differentiate; in many cases, after completing meiosis I, they arrest again in metaphase II while awaiting fertilization (although they can await fertilization at various other points, depending on the species).

While the details of oocyte development (**oogenesis**) vary from species to species, the general stages are similar, as outlined in **Figure 21–23**. Primordial germ cells migrate to the forming gonad to become *oogonia,* which proliferate by mitosis for a period before beginning meiosis I, at which point they are called *primary oocytes*; this usually occurs before birth in mammals. As discussed earlier, prior to the onset of meiosis I, the DNA replicates so that each chromosome consists of two sister chromatids; at the start of prophase I, the duplicated homologous chromosomes pair along their long axes, and crossing-over occurs between nonsister chromatids of these paired homologs (see Figure 21–10). After these events, the cell remains arrested after diplotene of prophase I for a period lasting from a few days to many years, depending on the species. During this long arrest period (or, in some cases, at the onset of sexual maturity), the primary oocytes synthesize a coat and cortical granules. The large oocytes of nonmammalian species also accumulate ribosomes, yolk, glycogen, lipid, and the mRNAs that will later direct the synthesis of proteins required for early embryonic growth and development. In many of these oocytes, we can observe the intensive biosynthetic activities in the structure of the chromosomes, which decondense and form lateral loops, taking on a characteristic "lampbrush" appearance, signifying that the genes in the loops are being busily transcribed (see Figures 4–54 and 4–55).

The next phase of oocyte development, *oocyte maturation*, usually does not occur until sexual maturity, when hormones stimulate the oocyte. Under these hormonal influences, the cell resumes its progress through meiosis I. The chromosomes recondense, the nuclear envelope breaks down, the meiotic

spindle assembles, and the replicated homologous chromosomes segregate at anaphase I into two sets, each containing half the original number of chromosomes. To end meiosis I, the cytoplasm divides asymmetrically to produce two cells that differ greatly in size: one is a small *polar body*, and the other is a large **secondary oocyte**, the precursor of the egg. At this stage, each chromosome is still composed of two sister chromatids held together at their centromeres. The sister chromatids do not separate until anaphase II, after which the cytoplasm of the large secondary oocyte again divides asymmetrically to produce the mature **egg** (or **ovum**) and a second small polar body, each with a haploid set of single chromosomes (see Figure 21–23). Because of these asymmetrical divisions of

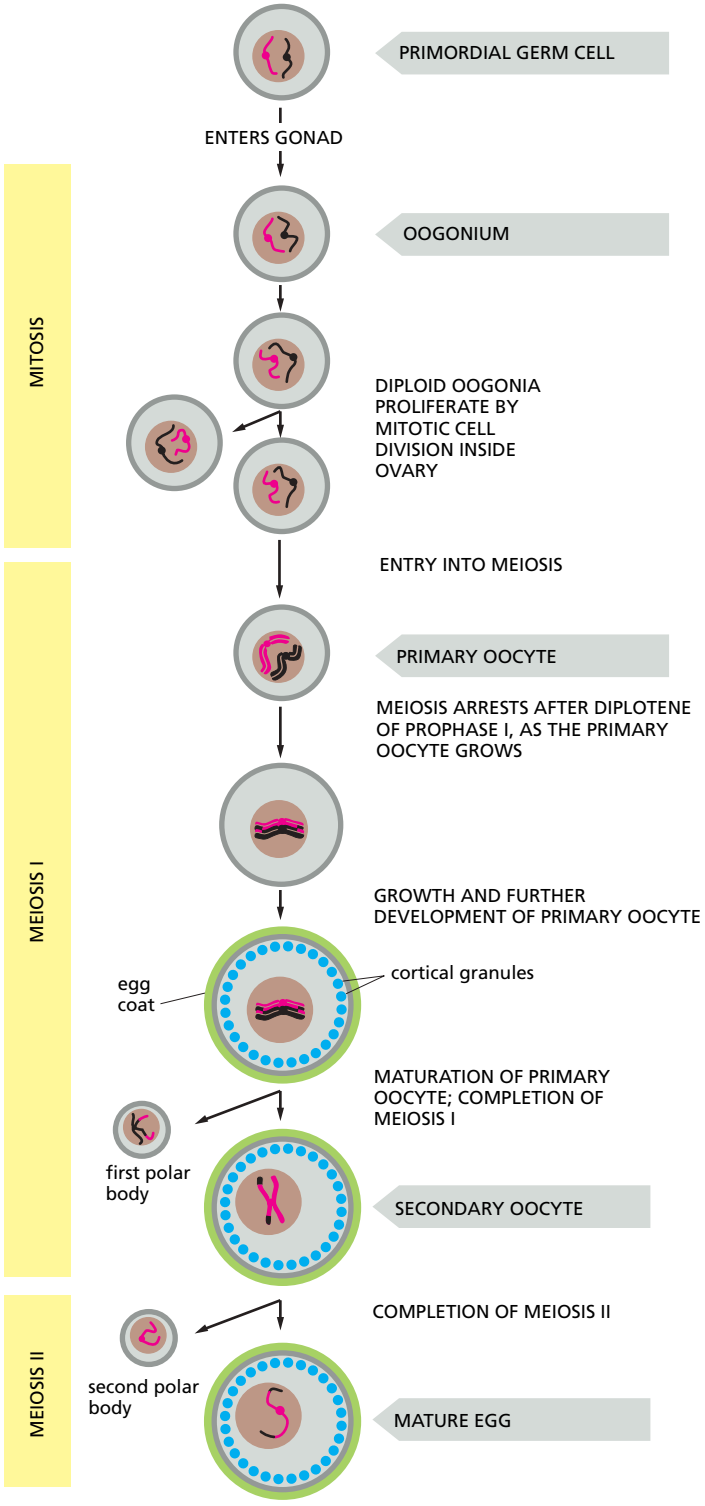


Figure 21–23 The stages of oogenesis. Oogonia develop from primordial germ cells (PGCs) that migrate into the developing gonad early in embryogenesis. For clarity, only one pair of homologous chromosomes is shown. After several mitotic divisions, oogonia begin meiosis and are now called primary oocytes. In mammals, primary oocytes are formed very early (between 3 and 8 months of gestation in the human embryo) and remain arrested after diplotene of prophase I until the female becomes sexually mature. At this point, a small number of primary oocytes periodically mature under the influence of hormones, completing meiosis I to produce secondary oocytes, which eventually undergo meiosis II to produce mature eggs (ova). The stage at which the egg or oocyte is released from the ovary and is fertilized varies from species to species. In most vertebrates, oocyte maturation is arrested at metaphase II, and the secondary oocyte completes meiosis II only after fertilization. All of the polar bodies eventually degenerate. In most animals, the developing oocyte is surrounded by specialized accessory cells that help to isolate and nourish it (not shown).

their cytoplasm, oocytes maintain their large size despite undergoing the two meiotic divisions. Both of the polar bodies are small, and they eventually degenerate.

In most vertebrates, oocyte maturation proceeds to metaphase of meiosis II, at which point they become arrested. At **ovulation**, the arrested secondary oocyte is released from the ovary, ready to be fertilized. If fertilization occurs, the block is lifted and the cell completes meiosis, becoming a mature egg. Because it is fertilized, it is also called a zygote.

Oocytes Use Special Mechanisms to Grow to Their Large Size

A somatic cell with a diameter of 10–20 μm typically takes about 24 hours to double its mass in preparation for cell division. At this rate of biosynthesis, such a cell would take a very long time to reach the thousand-fold greater mass of a mammalian egg with a diameter of 100 μm. It would take even longer to reach the million-fold greater mass of an insect egg with a diameter of 1000 μm. Yet, some insects live only a few days and manage to produce eggs with diameters even greater than 1000 μm. Eggs must have special mechanisms for achieving their large size.

One simple strategy for rapid growth is to have extra gene copies in the cell. Most of the growth of an oocyte occurs after DNA replication, during the prolonged arrest after diplotene in prophase I, when the diploid chromosome set is in duplicate (see Figure 21–23). In this way, it has twice as much DNA available for RNA synthesis as does an average somatic cell in the G_1 phase of the cell cycle. The oocytes of some species go even further to accumulate extra DNA: they produce many extra copies of certain genes. As discussed in Chapter 6, the somatic cells of most organisms contain 100 to 500 copies of the ribosomal RNA genes so as to produce enough ribosomes for protein synthesis. Eggs need even greater numbers of ribosomes to support the increased rate of protein synthesis required during early embryogenesis, and in the oocytes of many animals the ribosomal RNA genes are specifically amplified; some amphibian eggs, for example, contain 1 or 2 million copies of these genes.

Oocytes may also depend partly on the synthetic activities of other cells for their growth. Yolk, for example, is usually synthesized outside the ovary and imported into the oocyte. In birds, amphibians, and insects, yolk proteins are made by liver cells (or their equivalents), which secrete these proteins into the blood. Within the ovaries, oocytes use receptor-mediated endocytosis to take up the yolk proteins from the extracellular fluid (see Figure 13–46). Nutritive help can also come from neighboring accessory cells in the ovary. These can be of two types. In some invertebrates, some of the progeny of the oogonia become **nurse cells** instead of becoming oocytes. Cytoplasmic bridges connect these cells to the oocyte, allowing macromolecules to pass directly from the nurse cells into the oocyte cytoplasm (**Figure 21–24**). For the insect oocyte, the nurse cells manufacture many of the products—ribosomes, mRNA, protein, and so on—that vertebrate oocytes have to make for themselves.

The other accessory cells in the ovary that help to nourish developing oocytes are ordinary somatic cells called **follicle cells**, which surround each developing oocyte in both invertebrates and vertebrates. They are arranged as an epithelial layer around the oocyte (**Figure 21–25**, and see Figure 21–24), and they are connected to each other and to the oocyte by *gap junctions*, which permit the exchange of small molecules but not macromolecules (discussed in Chapter 19). Although follicle cells are unable to provide the oocyte with preformed macromolecules through these junctions, they can supply the smaller precursor molecules from which macromolecules are made. The critical importance of gap-junction communication has been elegantly demonstrated in the mouse ovary, where the gap-junction proteins *(connexins)* involved in connecting follicle cells to each other are different from those connecting follicle cells to the oocyte. If the genes that encode either of these proteins are disrupted in mice, both the follicle cells and oocytes fail to develop normally and the female mice are sterile. In many species, follicle cells secrete macromolecules that

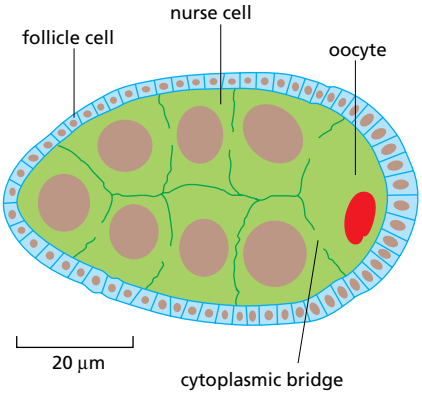
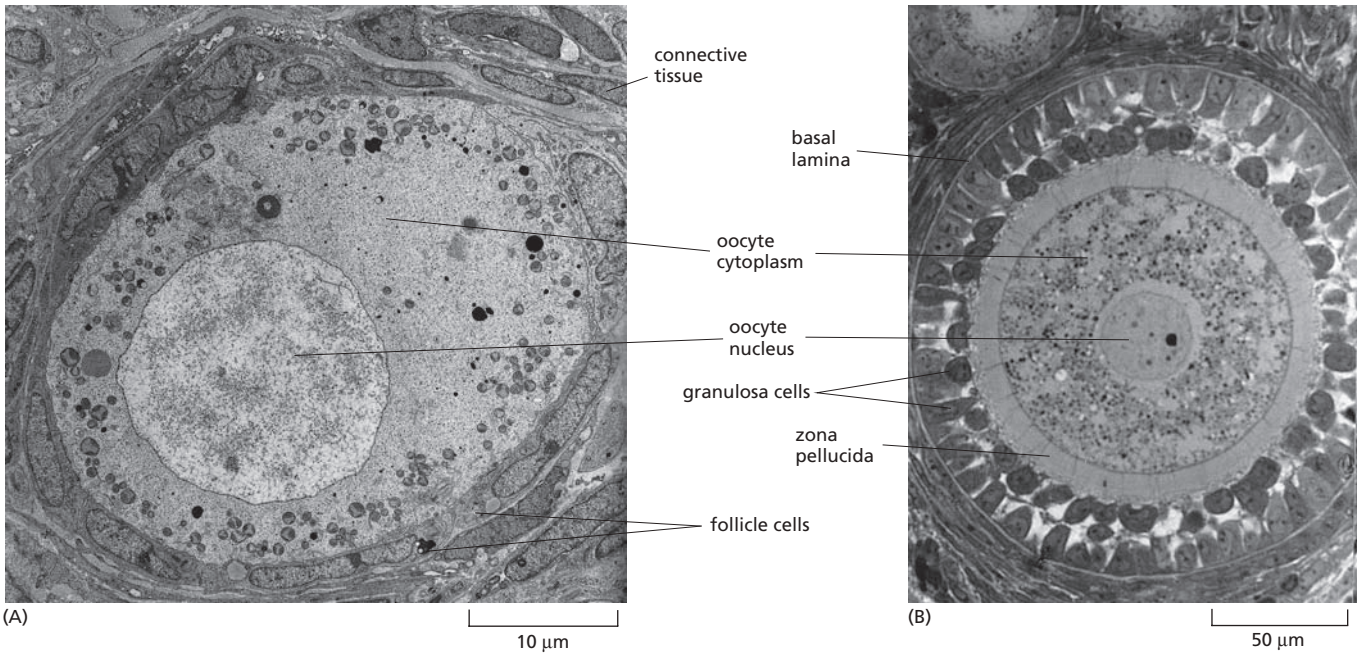


Figure 21–24 Nurse cells and follicle cells associated with a *Drosophila* oocyte. The nurse cells and the oocyte arise from a common oogonium, which gives rise to one oocyte and 15 nurse cells (only 7 of which are seen in this plane of section). These cells remain joined by cytoplasmic bridges, which result from incomplete cell division. Eventually, the nurse cells dump their cytoplasmic contents into the developing oocyte and then kill themselves. The follicle cells develop independently from mesodermal cells.



either contribute to the egg coat, are taken up by receptor-mediated endocytosis into the growing oocyte, or act on egg cell-surface receptors to control the spatial patterning and axial asymmetries of the egg (discussed in Chapter 22).

The communication between oocytes and their follicle cells occurs in both directions. The timing of developmental processes in the two sets of cells has to be coordinated, and it seems that this depends on signals from the oocyte to the follicle cells. Experiments in which young oocytes are combined with older follicle cells, or vice versa, show that an intrinsic developmental program in the oocyte normally controls the rate of follicle cell development.

Figure 21–25 Electron micrographs of developing primary oocytes in the rabbit ovary. (A) An early stage of primary oocyte development. Neither a zona pellucida nor cortical granules have developed, and a single layer of flattened follicle cells surrounds the oocyte. (B) A more mature primary oocyte, which is shown at a sixfold lower magnification because it is much larger than the oocyte in (A). This oocyte has acquired a thick zona pellucida and is surrounded by several layers of follicle cells (now called granulosa cells) and a basal lamina, which isolate the oocyte from the other cells in the ovary. The granulosa cells are connected to one another and to the oocyte by gap junctions. (From The Cellular Basis of Mammalian Reproduction [J. Van Blerkom and P. Motta eds.]. Baltimore–Munich: Urban & Schwarzenberg, 1979.)

Most Human Oocytes Die Without Maturing

Figure 21–26 outlines the stages in human oocyte development in the ovary. A single layer of follicle cells surrounds most of the primary oocytes in newborn girls. Such an oocyte, together with its surrounding follicle cells, is called a *primordial follicle* (see Figure 21–25A). Periodically, beginning sometime before birth, a small proportion of primordial follicles begin to grow to become *developing follicles*, in which multiple layers of follicle cells (now called *granulosa cells*)

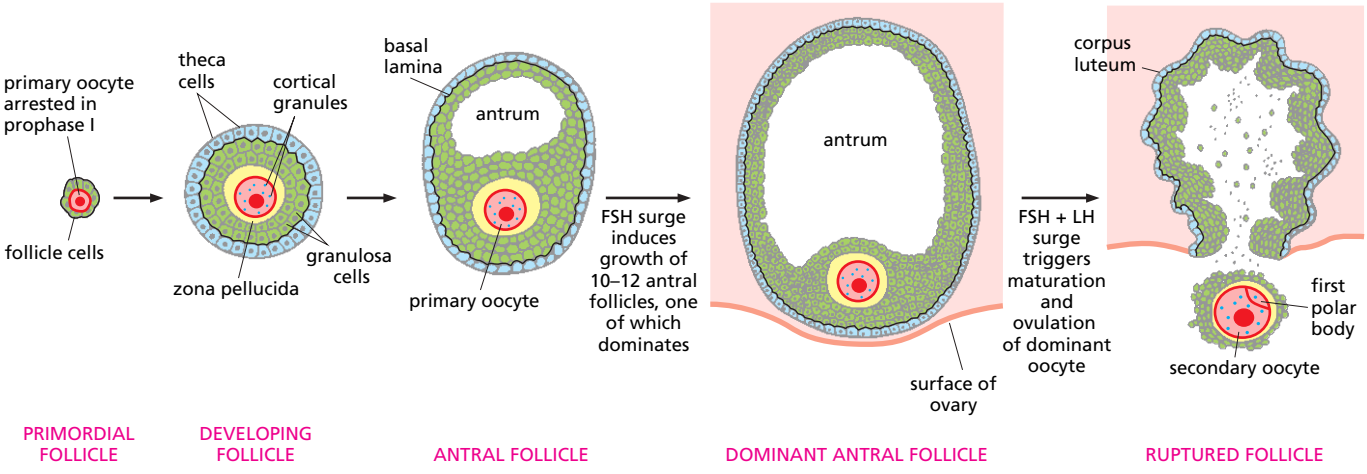


Figure 21–26 The stages in human oocyte development. Note that for most of its development the oocyte is surrounded by granulosa cells *(green),* which are separated from an outer layer of *theca cells (blue)* by an intervening basal lamina *(black).* After ovulation, the emptied follicle transforms into an endocrine structure, the *corpus luteum*, which secretes progesterone to help prepare the uterus for pregnancy. If fertilization does not occur, the corpus luteum regresses, and the lining of the uterus is sloughed off during menstruation.

surround the growing oocyte (see Figure 21–25B). It is not known what causes certain primordial follicles to begin growing. Some of these developing follicles go on to acquire a fluid-filled cavity, or *antrum,* to become *antral follicles.*

After puberty, about once each month, the pituitary secretes a surge of *follicle stimulating hormone (FSH)*, which accelerates the growth of about 10–12 antral follicles. One of these antral follicles becomes dominant, and, toward the middle of the menstral cycle, a surge in FSH and *luteinizing hormone (LH)* triggers *ovulation*: the dominant primary oocyte completes meiosis I, and the resulting secondary oocyte arrests at metaphase II; the follicle rapidly enlarges and ruptures at the surface of the ovary, releasing the secondary oocyte, still surrounded by a shell of granulosa cells embedded in a hyaluronan-rich gel-like matrix. The released oocyte is triggered to complete meiosis II only if a sperm fertilizes it within a day or so.

It remains a mystery why only a small proportion of the many antral follicles present in the ovaries at the time of the FSH surge each month are stimulated to accelerate their growth, and why only one of these matures and releases its oocyte, while the rest degenerate. Once the selected follicle has matured beyond a certain point, some feedback mechanism must operate to ensure that no other follicles complete maturation and ovulate during that cycle. Whatever the mechanism, the result is that, during the 40 or so years of a woman's reproductive life, only 400 to 500 oocytes will be released. All of the other million or so primary oocytes present at birth die without maturing. It is still a puzzle why so many oocytes are formed only to die in the ovaries.

Summary

Oocytes develop in stages from primordial germ cells (PGCs) that migrate into the developing gonad, where they become oogonia. After a period of mitotic proliferation, the oogonia begin meiosis I and are now called primary oocytes. Primary oocytes remain arrested after diplotene of prophase I for days to years, depending on the species. During this arrest period, they grow, synthesize a coat, and accumulate ribosomes, mRNAs, and proteins, often enlisting the help of other cells, including surrounding follicle cells. Bidirectional signaling between oocytes and their follicle cells is required for normal oocyte growth and development. In the process of hormonally induced oocyte maturation, primary oocytes complete meiosis I to form a small polar body and a large secondary oocyte, which proceeds to metaphase of meiosis II. In most vertebrates, the secondary oocyte arrests at metaphase II until stimulated by fertilization to complete meiosis and begin embryonic development.

SPERM

In most species, there are two radically different types of gametes. The egg is among the largest cells in an organism, whereas the **sperm** (**spermatozoon**, plural **spermatozoa**) is often the smallest. The egg and the sperm are optimized in opposite ways for the propagation of the genes they carry. The egg is nonmotile and aids the survival of the maternal genes by providing large stocks of raw materials for embryo growth and development, together with an effective protective wrapping. The sperm, by contrast, is optimized to propagate the paternal genes by exploiting this maternal investment: it is usually highly motile and streamlined for speed and efficiency in the task of fertilization. Competition between sperm is fierce, and the vast majority fail in their mission: of the billions of sperm released during a human male's reproductive life, only a few ever manage to fertilize an egg.

Sperm Are Highly Adapted for Delivering Their DNA to an Egg

Typical sperm are "stripped-down" cells, equipped with a strong flagellum to propel them through an aqueous medium but unencumbered by cytoplasmic

organelles such as ribosomes, endoplasmic reticulum, or Golgi apparatus, which are unnecessary for the task of delivering the DNA to the egg. Sperm, however, contain many mitochondria strategically placed where they can most efficiently power the flagellum. Sperm usually consist of two morphologically and functionally distinct regions enclosed by a single plasma membrane: the *tail,* which propels the sperm to the egg and helps it to burrow through the egg coat, and the *head,* which contains a highly condensed haploid nucleus (**Figure 21–27**). The DNA in the nucleus is extremely tightly packed, to minimize its volume for transport, and transcription is shut down. The chromosomes of many sperm lack the histones of somatic cells and are packaged instead with simple, highly positively charged proteins called *protamines*, as well as with sperm-specific histones.

In the head of most animal sperm, closely apposed to the anterior end of the nuclear envelope, is a specialized secretory vesicle called the **acrosomal vesicle**. This vesicle contains hydrolytic enzymes that are thought to help the sperm penetrate the egg's outer coat. When a sperm contacts the egg coat, the contents of the vesicle are released by exocytosis in the *acrosome reaction*. This reaction is required for the sperm to bind to the coat, burrow through it, and fuse with the egg.

The motile tail of a sperm is a long flagellum, whose central *axoneme* emanates from a basal body situated just behind the nucleus. As described in Chapter 16, the axoneme consists of two central singlet microtubules surrounded by nine evenly spaced microtubule doublets. The flagellum of some sperm (including those of mammals) differs from other flagella in that the 9 + 2 pattern of microtubules is surrounded by nine outer *dense fibers* (**Figure 21–28**). The dense fibers are stiff and noncontractile, and they are thought to restrict the flexibility of the flagellum and protect it from shear forces; defects in these fibers lead to abnormal sperm morphology and to infertility. The active bending of the flagellum is caused by the sliding of adjacent microtubule doublets past one another, driven by dynein motor proteins, which use the energy of ATP hydrolysis to slide the microtubules. The ATP is generated by a large number of highly specialized mitochondria that are concentrated in the anterior part of the sperm tail (called the *midpiece*), where the ATP is needed.

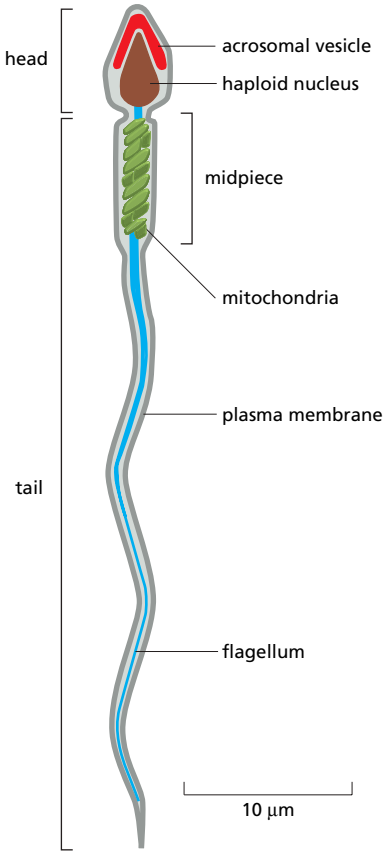


Figure 21–27 A human sperm. It is shown in longitudinal section.

Sperm Are Produced Continuously in the Mammalian Testis

In contrast to oocytes, which begin meiosis before birth and remain arrested after diplotene of prophase I until a human female reaches puberty, meiosis and sperm production (**spermatogenesis**) do not begin in the testes of human males until puberty. They then go on continuously in the epithelial lining of very long, tightly coiled tubes, called *seminiferous tubules*. Immature germ cells, called *spermatogonia* (singular, *spermatogonium*), are located around the outer edge of these tubes next to the basal lamina (**Figure 21–29**A). Most of these cells divide a limited number of times by mitosis before they stop proliferating and begin meiosis I, at which point they are now called *primary spermatocytes;* the primary spermatocytes give rise to *secondary spermatocytes,* which differentiate into *spermatids* and ultimately sperm (Figure 21–29B). A small proportion of the spermatogonia serve as stem cells, which divide slowly by mitosis throughout life, producing daughter cells that either remain stem cells or commit to maturation.

The stages of spermatogenesis and their relationship to meiosis are illustrated in **Figure 21–30**. During prophase I, the paired homologous chromosomes participate in crossing-over. The primary spermatocytes then complete

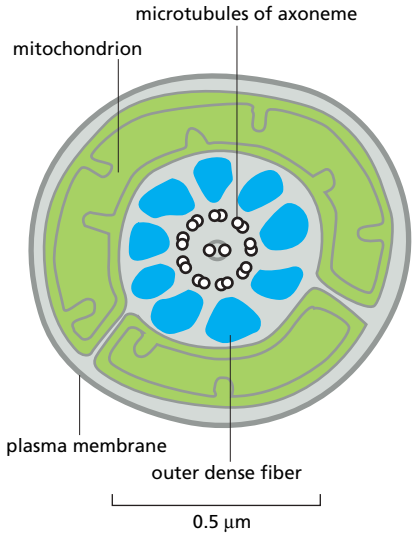
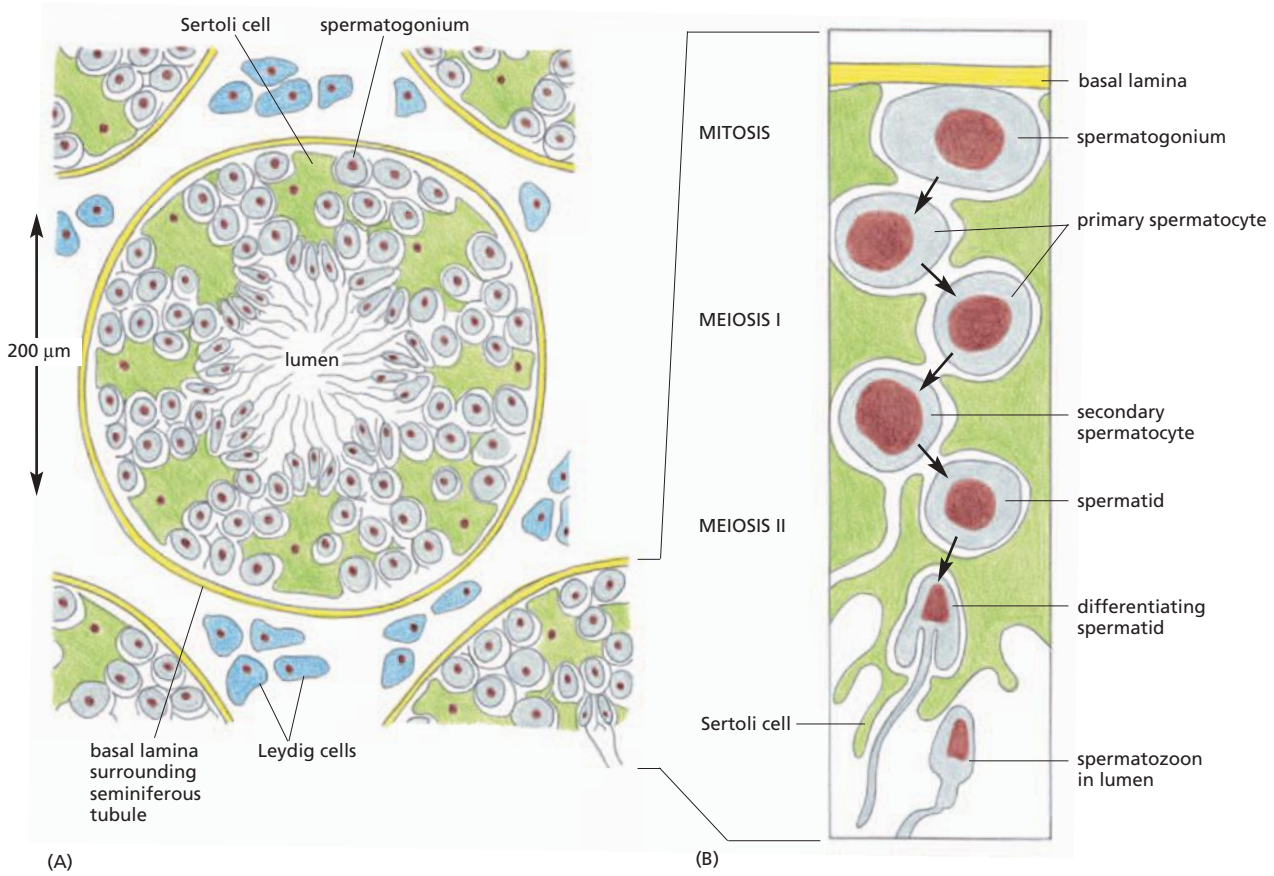


Figure 21–28 Drawing of the midpiece of a mammalian sperm as seen in cross section in an electron microscope. The core of the flagellum is composed of an axoneme surrounded by nine dense fibers. The axoneme consists of two singlet microtubules surrounded by nine microtubule doublets. The mitochondria (shown in *green*) are well placed for providing the ATP required for flagellar movement; they are distributed in an unusual spiral arrangement around the axoneme (see Figure 21–27).



meiosis I to produce two secondary spermatocytes, each containing 22 duplicated autosomal chromosomes and either a duplicated X or a duplicated Y chromosome. The two secondary spermatocytes derived from each primary spermatocyte proceed through meiosis II to produce four spermatids, each with a haploid number of single chromosomes. The haploid spermatids then undergo dramatic morphological changes as they differentiate into sperm, which escape into the lumen of the seminiferous tubule. The sperm subsequently pass into the *epididymis*, a coiled tube overlying the testis, where they are stored and undergo further maturation. The stored sperm are still not ready to fertilize an egg, however; as we discuss later, they undergo further maturation in the female genital tract—a process called *capacitation*.

Sperm Develop as a Syncytium

An intriguing feature of spermatogenesis is that once a spermatogonium begins to mature, its progeny no longer complete cytoplasmic division (cytokinesis) during mitosis and subsequent meiosis. Consequently, large clones of differentiating daughter cells that have descended from one maturing spermatogonium remain connected by cytoplasmic bridges, forming a syncytium (**Figure 21–31**). The cytoplasmic bridges persist until the very end of sperm differentiation, when individual sperm are released into the lumen of the seminiferous tubule. As a result, mature sperm are produced in synchronous batches in any given area of a seminiferous tubule. What is the function of this syncytial arrangement?

We saw earlier that oocytes grow and differentiate while containing the diploid set of chromosomes in duplicate. Sperm, by contrast, do not grow, and they undergo most of their differentiation after their nuclei have completed meiosis to become haploid. The presence of cytoplasmic bridges between them, however, means that each developing haploid sperm shares a common cytoplasm with its neighbors. In this way, it can be supplied with all the gene

Figure 21–29 Highly simplified drawings of a cross section of a seminiferous tubule in a mammalian testis. (A) All of the stages of spermatogenesis shown take place while the developing germline cells are in intimate association with Sertoli cells. Sertoli cells direct sexual differentiation along a male pathway. They are large cells, extending from the basal lamina to the lumen of the seminiferous tubule; they are required for the survival of the spermatogonia and are analogous to follicle cells in the ovary (see Figure 21–19). Spermatogenesis also depends on testosterone secreted by Leydig cells, located between the seminiferous tubules. (B) Spermatogonia divide by mitosis at the periphery of the seminiferous tubule. Some of these cells enter meiosis I to become primary spermatocytes; they then complete meiosis I to become secondary spermatocytes. The secondary spermatocytes then complete meiosis II to become spermatids, which differentiate into spermatozoa (sperm) and are released into the lumen of the tubule (see Figure 21–30). In man, it takes a spermatogonium about 24 days from the onset of meiosis to emergence as a spermatid and another 5 weeks for the spermatid to develop into a sperm.

products of a complete diploid genome. Developing sperm that carry a Y chromosome, for example, can be supplied with essential proteins encoded by genes on the X chromosome. Thus, the diploid genome directs sperm differentiation just as it directs egg differentiation.

Some of the genes that regulate spermatogenesis have been conserved in evolution from flies to humans. The *Daz* genes, for example, encode RNA-binding proteins and are located in a cluster on the human Y chromosome. The cluster is found to be deleted in a substantial proportion of infertile men, many of whom cannot make sperm. A *Drosophila* gene that is homologous to the human

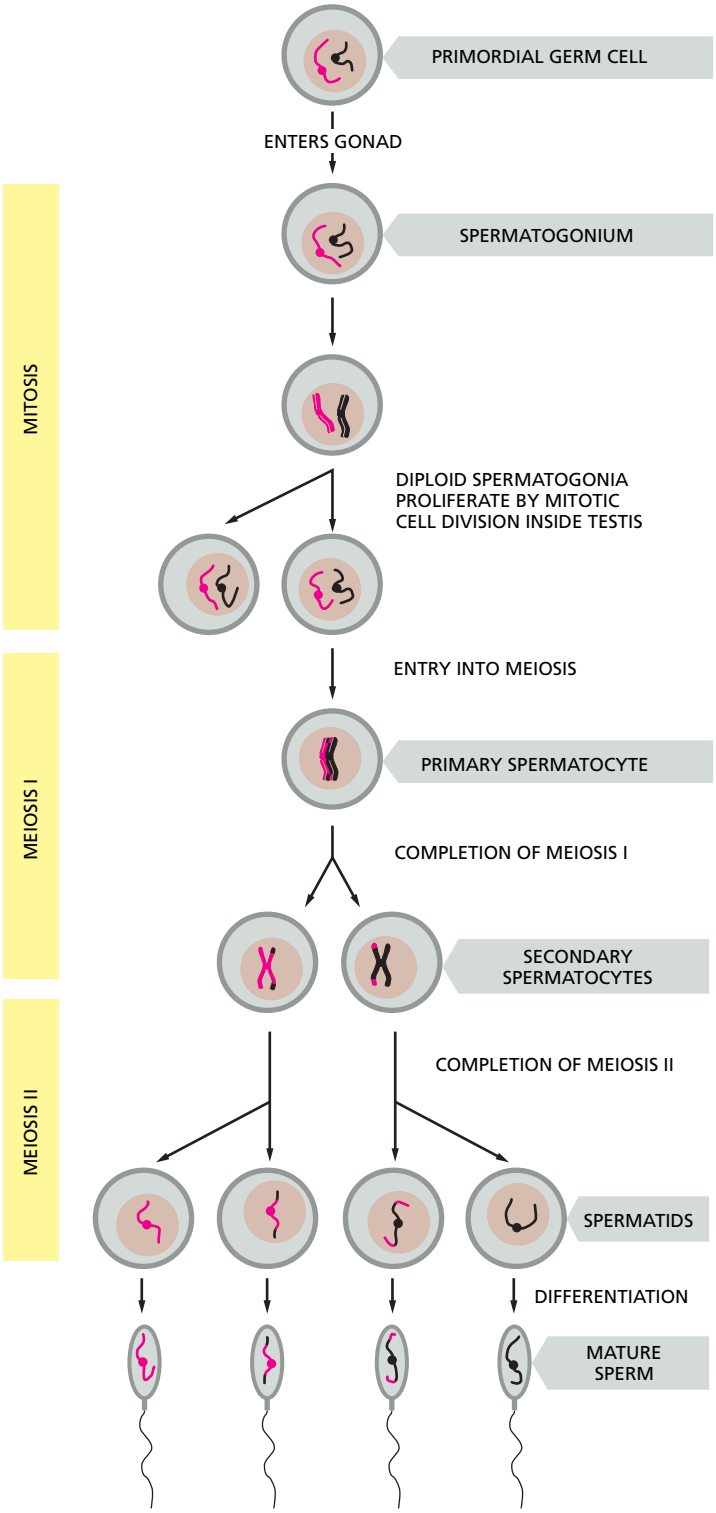


Figure 21–30 The stages of spermatogenesis. Spermatogonia develop from primordial germ cells (PGCs) that migrate into the developing gonad early in embryogenesis. When the animal becomes sexually mature, the spermatogonia begin to proliferate rapidly by mitosis. Some retain the capacity to divide indefinitely (as stem-cell spermatogonia). Others (maturing spermatogonia) undergo a limited number of mitotic division cycles before beginning meiosis to become spermatocytes, which eventually become haploid spermatids and then sperm. Spermatogenesis differs from oogenesis (see Figure 21–23) in several ways. (1) New cells enter meiosis continually from the time of puberty. (2) Each cell that begins meiosis gives rise to four mature gametes rather than one. (3) Mature sperm form by an elaborate process of cell differentiation that begins after meiosis is complete. (4) About twice as many cell divisions occur in the production of a sperm as in the production of an egg; in a mouse, for example, it is estimated that on average about 56 divisions occur from zygote to mature sperm, and about 27 divisions occur from zygote to mature egg.

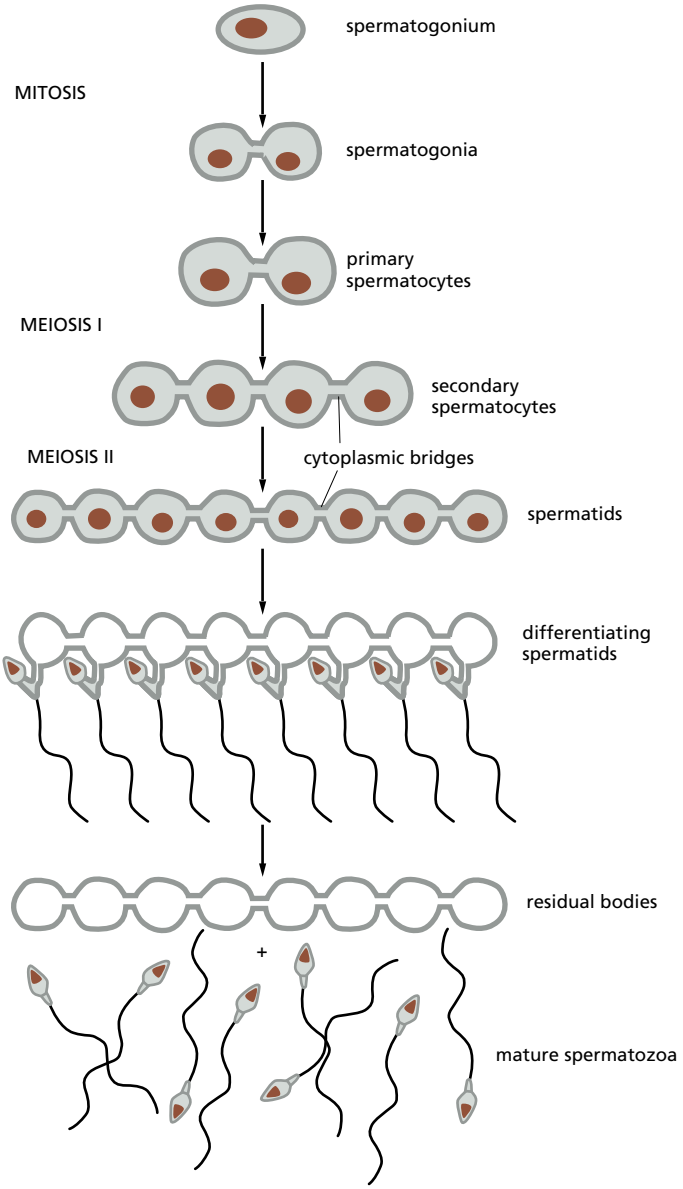


Figure 21–31 Cytoplasmic bridges in developing sperm cells and their precursors. The progeny of a single maturing spermatogonium remain connected to one another by cytoplasmic bridges throughout their differentiation into mature sperm. For the sake of simplicity, only two connected maturing spermatogonia are shown entering meiosis, eventually to form eight connected haploid spermatids. In fact, the number of connected cells that go through meiosis and differentiate synchronously is very much larger than shown here. Note that in the process of differentiating, most of the spermatid cytoplasm is discarded as *residual bodies*, which are phagocytosed by Sertoli cells.

Daz genes is similarly essential for spermatogenesis in the fly: Daz-deficient male flies are infertile because they cannot make sperm, but, remarkably, they can be cured by a human *Daz* transgene. RNA-binding proteins are especially important in spermatogenesis, because many of the genes expressed in the sperm lineage are regulated at the level of RNA translation.

Summary

A sperm is usually a small, compact cell, highly specialized for the task of fertilizing an egg. Whereas in women a large pool of oocytes is produced before birth, in men spermatogonia only begin to enter meiosis to produce spermatocytes (and sperm) after sexual maturation, and they continue to do so from then on. Each diploid primary spermatocyte gives rise to four mature haploid sperm. The process of sperm differentiation occurs after meiosis is complete, requiring five weeks in humans. Because the maturing spermatogonia and spermatocytes fail to complete cytokinesis, however, the progeny of a single maturing spermatogonium develop as a large syncytium. Thus, the products encoded by both parental chromosomes direct sperm differentiation, even though each sperm nucleus is haploid.

FERTILIZATION

Once released, egg and sperm alike are destined to die within minutes or hours unless they find each other and fuse in the process of **fertilization**. Through fertilization, the egg and sperm are saved: the egg is activated to begin its developmental program, and the haploid nuclei of the two gametes come together to form the diploid genome of a new organism. Fertilization was originally studied most intensively in marine invertebrates such as sea urchins and starfish, where fertilization occurs in seawater after the release of huge numbers of both sperm and eggs. Such external fertilization is easier to study than the internal fertilization of mammals, which normally occurs in the female reproductive tract after mating. In the late 1950s, however, it became possible to fertilize mammalian eggs *in vitro*, opening the way to an analysis of the cellular and molecular events in mammalian fertilization.

In this section, we focus on mammalian fertilization. We begin by considering the maturation of sperm that occurs during their passage through the female genital tract. We then discuss the binding of sperm to the egg coat (zona pellucida), which induces the *acrosome reaction*, required for the sperm to burrow through the zona and fuse with the egg. We next consider the binding of the sperm to the egg plasma membrane and its subsequent fusion with this membrane. After discussing how the fusion of a sperm activates the egg and how the haploid nuclei of the two gametes come together in the zygote to complete fertilization, we briefly consider the burgeoning field of assisted reproductive technology, which has revolutionized the treatment of human infertility and opened up new ways of manipulating the reproductive process.

Ejaculated Sperm Become Capacitated in the Female Genital Tract

Of the 300,000,000 or so human sperm ejaculated during coitus, only about 200 reach the site of fertilization in the *oviduct*. Once a sperm finds an egg, it must first migrate through the layers of granulosa cells that surround the egg and then bind to and cross the zona pellucida. Finally, it must bind to and fuse with the egg plasma membrane.

Ejaculated mammalian sperm are initially not competent to accomplish any of these tasks. They must first be modified by conditions in the female reproductive tract. Because it is required for sperm to acquire the capacity to fertilize an egg, the process is called **capacitation**. Capacitation takes about 5–6 hours in humans and is completed only when the sperm arrive in the oviduct. The sperm undergo extensive biochemical and functional changes, including changes in glycoproteins, lipids, and ion channels in the sperm plasma membrane and a large change in the resting potential of this membrane (the membrane potential moves to a more negative value so that the membrane becomes hyperpolarized). Capacitation is also associated with an increase in cytosolic pH, tyrosine phosphorylation of various sperm proteins, and the unmasking of cell-surface receptors that help bind the sperm to the zona pellucida. Capacitation alters two crucial aspects of sperm behavior: it greatly increases the motility of the flagellum, and it makes the sperm capable of undergoing the acrosome reaction.

Capacitation can occur *in vitro* in the appropriate culture medium and is usually a required part of *in vitro* fertilization. Three critical components are needed in the medium, all of which are normally in high concentration in the female genital tract—albumin, Ca^{2+}, and HCO_3^-. The albumin protein helps extract cholesterol from the plasma membrane, increasing the ability of this membrane to fuse with the acrosome membrane during the acrosome reaction. The Ca^{2+} and HCO_3^- enter the sperm and directly activate a soluble adenylyl cyclase enzyme in the cytosol to produce cyclic AMP (discussed in Chapter 15), which helps to initiate many of the changes associated with capacitation.

Capacitated Sperm Bind to the Zona Pellucida and Undergo an Acrosome Reaction

During ovulation, mammalian eggs are released from the ovary into the peritoneal cavity next to the entrance to the oviduct, into which they are rapidly swept. They are covered with several layers of granulosa cells embedded in an extracellular matrix that is rich in hyaluronic acid (discussed in Chapter 19). The granulosa cells may help the egg get picked up into the oviduct, and they may also secrete unidentified chemical signals that attract sperm to the egg.

On encountering an egg, a capacitated sperm first must penetrate the layers of granulosa cells, making use of a hyaluronidase enzyme on the surface of the sperm. It can then bind to the **zona pellucida** (see Figure 21–22). The zona usually acts as a barrier to fertilization across species, and removing it often eliminates this barrier. Human sperm, for example, will fertilize hamster eggs from which the zona has been removed with specific enzymes; not surprisingly, such hybrid zygotes fail to develop. Zona-free hamster eggs are sometimes used in infertility clinics to assess the fertilizing capacity of human sperm *in vitro* (**Figure 21–32**).

The zona pellucida of most mammalian eggs is composed mainly of three glycoproteins, all of which are produced exclusively by the growing oocyte. Two of them, ZP2 and ZP3, assemble into long filaments, while the other, ZP1, cross-links the filaments into a three-dimensional network. The ZP3 protein is crucial: female mice with a disrupted *Zp3* gene produce eggs that lack a zona and are infertile. *O*-linked oligosaccharides on ZP3 seem to be at least partly responsible for the species-specific binding of sperm to the zona. The binding of sperm to the zona is complex, however, and involves both ZP3-dependent and ZP3-independent mechanisms and a variety of proteins on the sperm surface.

The zona induces sperm to undergo the **acrosome reaction**, in which the contents of the acrosome are released by exocytosis (**Figure 21–33**). The acrosome reaction is required for normal fertilization, as it exposes various hydrolytic enzymes that are believed to help the sperm tunnel through the zona pellucida, and it alters the sperm surface so that it can bind to and fuse with the plasma membrane of the egg, as we discuss below. *In vitro*, purified ZP3 can trigger the acrosome reaction, possibly by activating a lectin-like receptor on the sperm surface, thought to be a transmembrane form of the enzyme galactosyltransferase. Receptor activation leads to an increase in Ca^{2+} in the sperm cytosol, which initiates the exocytosis.

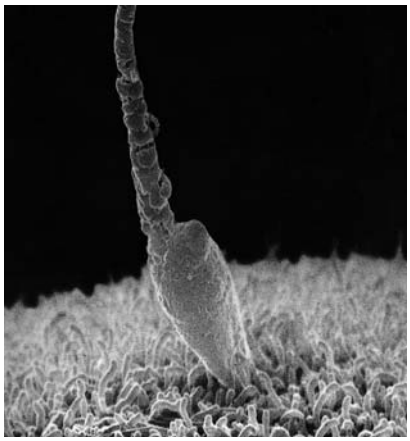


Figure 21–32 Scanning electron micrograph of a human sperm contacting a hamster egg. The zona pellucida of the egg has been removed, exposing the plasma membrane, which contains numerous microvilli. The ability of an individual's sperm to penetrate hamster eggs is used as an assay of male fertility; penetration of more than 10–25% of the eggs is considered to be normal. (Courtesy of David M. Phillips.)

The Mechanism of Sperm–Egg Fusion Is Still Unknown

After a sperm has undergone the acrosome reaction and penetrated the zona pellucida, it binds to the egg plasma membrane overlying the tips of microvilli on the egg surface (see Figure 21–32). The sperm binds initially by its tip and then by its side (see Figure 21–33). Neighboring microvilli on the egg surface rapidly elongate and cluster around the sperm to ensure that it is held firmly so that it can fuse with the egg. After fusion, all of the sperm contents are drawn into the egg, as the microvilli are resorbed.

The molecular mechanisms responsible for sperm–egg binding and fusion are largely unknown, although, after a number of false starts, two membrane proteins have been shown to be required for the fusion. One is a sperm-specific transmembrane protein of the immunoglobulin superfamily called *Izumo* (after a Japanese shrine dedicated to marriage). It becomes exposed on the surface of mouse and human sperm during the acrosome reaction. Anti-Izumo antibodies block the fusion, and Izumo-deficient mouse sperm fail to fuse with normal eggs, but it is still unknown how Izumo promotes sperm–egg fusion. The only protein on the egg surface demonstrated to be required for fusion with a sperm is the *CD9* protein, which is a member of the *tetraspanin* family, so-called because these proteins have four membrane-spanning segments. Normal sperm fail to fuse with CD9-deficient mouse eggs, indicating that sperm–egg fusion depends on CD9, but it is not known how. CD9 does not act alone on the egg surface to promote fusion: normal sperm also fail to fuse with eggs treated

with an enzyme that removes proteins attached to the plasma membrane by a glycosylphosphatidylinositol (GPI) anchor (discussed in Chapter 10), indicating that one or more GPI-linked proteins is also required for fusion, although the relevant protein or proteins have yet to be identified.

Sperm Fusion Activates the Egg by Increasing Ca^{2+} in the Cytosol

Fusion with a sperm activates the egg, causing the cortical granules to release their contents by exocytosis, a process called the *cortical reaction*. Meiosis, which was arrested in metaphase II, resumes, producing a second polar body, and a zygote, which begins to develop.

An increase in Ca^{2+} in the cytosol of the fertilized egg triggers all of these events. <AGGA> If the Ca^{2+} concentration in the cytosol of an unfertilized egg is increased artificially—either directly by an injection of Ca^{2+} or indirectly by the use of a Ca^{2+}-carrying ionophore (discussed in Chapter 11)—the eggs of all animals so far tested, including mammals, are activated. Conversely, preventing the increase in Ca^{2+} by injecting the Ca^{2+} chelator EGTA inhibits activation of the egg in response to fertilization.

When the sperm fuses with the egg plasma membrane in the normal way, it causes a local increase in cytosolic Ca^{2+}, which spreads through the cell in a wave (see Figure 15–40). The wave propagates by positive feedback: the increase in cytosolic Ca^{2+} causes Ca^{2+} channels to open, allowing still more Ca^{2+} to enter the cytosol. The initial wave of Ca^{2+} release is usually followed within a few minutes by Ca^{2+} oscillations (discussed in Chapter 15), which persist for several hours.

The fused sperm triggers the Ca^{2+} wave and oscillations by bringing a factor into the egg cytosol. An injection of an intact sperm, a sperm head, or a sperm extract into an egg does the same thing. All of these treatments increase the concentration of inositol 1,4,5-trisphosphate (IP_3), which releases Ca^{2+} from the endoplasmic reticulum and initiates the Ca^{2+} wave and oscillations (discussed in Chapter 15). A strong candidate for the critical factor that mammalian sperm introduce into the egg is a sperm-specific form of phospholipase C (PLCζ), which directly cleaves phosphoinositol 4,5-bisphosphate ($PI(4,5)P_2$) to produce IP_3 (and diacylglycerol) (see Figure 15–39).

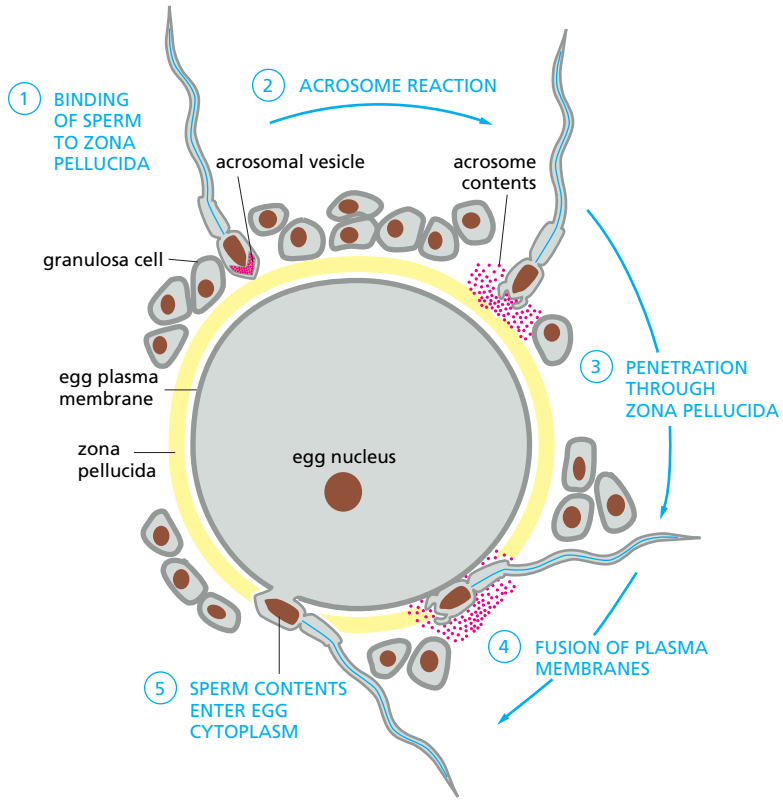


Figure 21–33 The acrosome reaction that occurs when a mammalian sperm fertilizes an egg. In mice, the zona pellucida is about 6 µm thick, and sperm tunnel through it at a rate of about 1 µm/min.

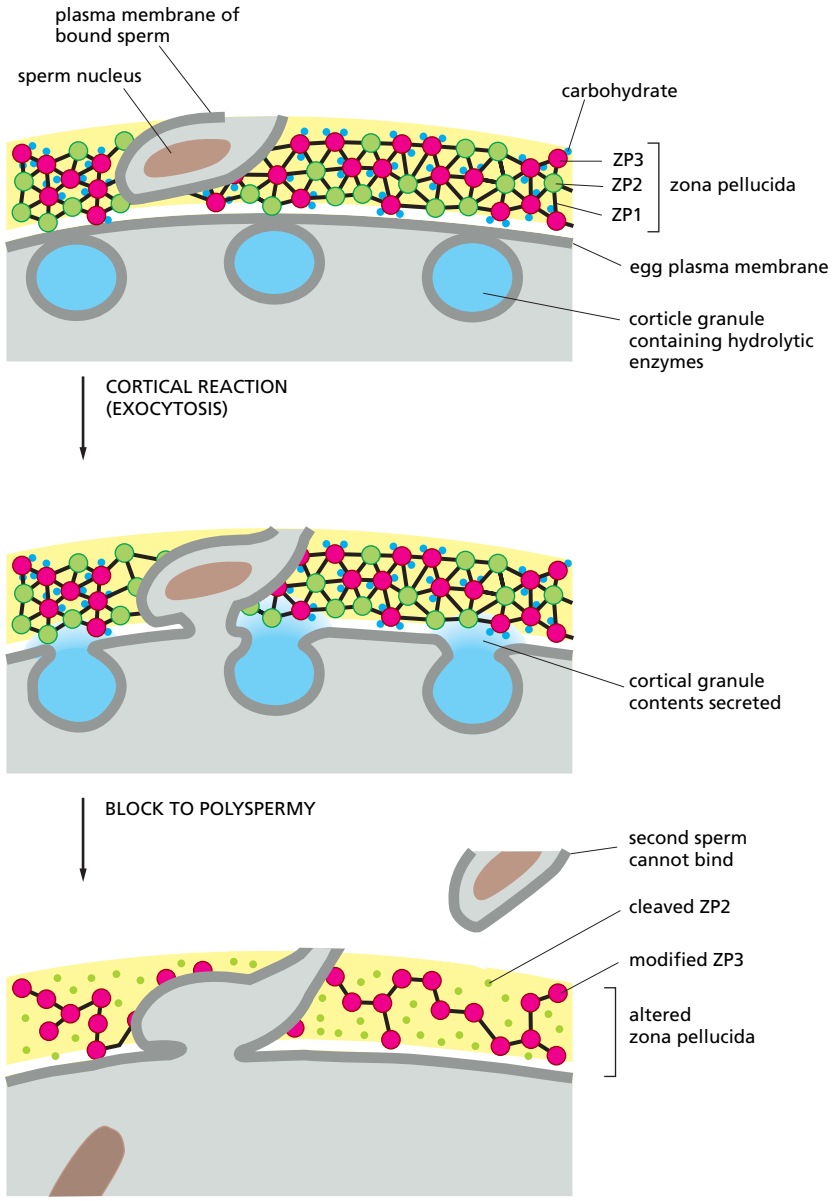


Figure 21–34 How the cortical reaction in a mouse egg is thought to prevent additional sperm from entering the egg. <TGAC> The released contents of the cortical granules inactivate ZP3 so it can no longer bind to the sperm plasma membrane. They also partly cleave ZP2, hardening the zona pellucida so that sperm cannot penetrate it. Together, these changes provide a block to polyspermy.

The Cortical Reaction Helps Ensure That Only One Sperm Fertilizes the Egg

Although many sperm can bind to an egg, normally only one fuses with the egg plasma membrane and injects its cytosol, nucleus, and other organelles into the egg cytoplasm. If more than one sperm fuses—a condition called *polyspermy*—extra or multipolar mitotic spindles are formed, resulting in faulty segregation of chromosomes during the first mitotic cell divisions; aneuploid cells are produced, and development usually stops.

Two mechanisms operate to ensure that only one sperm fertilizes the egg. First, a change in the egg plasma membrane caused by the fusion of the first sperm prevents other sperm from fusing. In sea urchin eggs, the change is a rapid depolarization of the egg membrane; in mammalian eggs, the mechanism is not known. The second block to polyspermy is provided by the egg **cortical reaction**, which releases various enzymes that change the structure of the zona pellucida so that sperm cannot bind to or penetrate it. Among the changes that occur in the mammalian zona is the inactivation of ZP3 so that it can no longer bind sperm or induce an acrosome reaction; in addition, ZP2 is cleaved, which somehow helps to make the zona impenetrable (**Figure 21–34**).

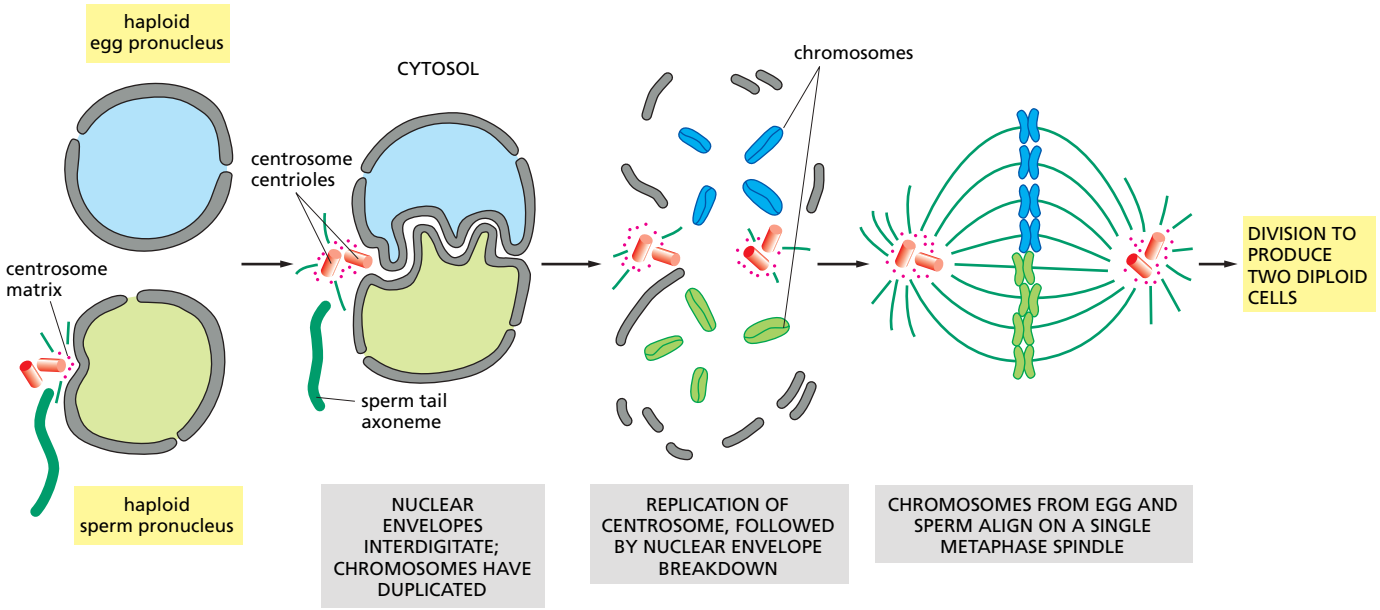


Figure 21–35 The coming together of the sperm and egg pronuclei after mammalian fertilization. The pronuclei migrate toward the center of the egg. When they come together, their nuclear envelopes interdigitate. The centrosome duplicates, the nuclear envelopes break down, and the chromosomes of both gametes are eventually integrated into a single mitotic spindle, which organizes the first cleavage division of the zygote. (Adapted from drawings and electron micrographs provided by Daniel Szöllösi.)

The Sperm Provides Centrioles as Well as Its Genome to the Zygote

Once fertilized, the egg is called a **zygote**. Fertilization is not complete, however, until the two haploid nuclei (called *pronuclei*)–one from the egg, the other from the sperm—have come together and combined their chromosomes into a single diploid nucleus. In fertilized mammalian eggs, the two pronuclei do not fuse directly as they do in many other species. They approach each other but remain distinct until after the membrane of each pronucleus has broken down in preparation for the zygote's first mitotic division (**Figure 21–35**).

In most animals, including humans, the sperm contributes more than its genome to the zygote. It also donates its centrioles—structures that are lacking in unfertilized human eggs. The sperm centrioles enter the egg along with the sperm nucleus and tail, and a centrosome forms around them. In humans, the centrosome duplicates, and the two resulting centrosomes then help organize the assembly of the first mitotic spindle in the zygote (**Figure 21–36**, and see Figure 21–35). This explains why polyspermy, in which several sperm contribute their centrioles to the egg, causes extra or multipolar mitotic spindles to form.

IVF and ICSI Have Revolutionized the Treatment of Human Infertility

About 10% of human couples have reduced fertility, such that the female partner fails to become pregnant after 12–18 months of unprotected sex. In roughly half of these cases it is the male that is the problem, and in half it is the female. Although there are numerous reasons for reduced fertility in both males and females, in the great majority of cases some form of assisted reproductive technology can solve the problem.

The first major breakthrough in the treatment of infertility occurred in 1978, with the birth of Louise Brown, the first child produced by ***in vitro* fertilization** (**IVF**). Before this success, there were heated debates about the ethics and safety of IVF—remarkably similar to the current ethical debates about the production and use of human embryonic stem (ES) cells. IVF is now a routine procedure and has produced more than a million children. To begin the process, the woman is usually pretreated with hormones to stimulate the simultaneous maturation of multiple oocytes. Just before their release by ovulation, the eggs are harvested from the ovary (using a long needle inserted through the vagina) and are fertilized in a culture dish with sperm from the man. After a few days in culture, 2 or

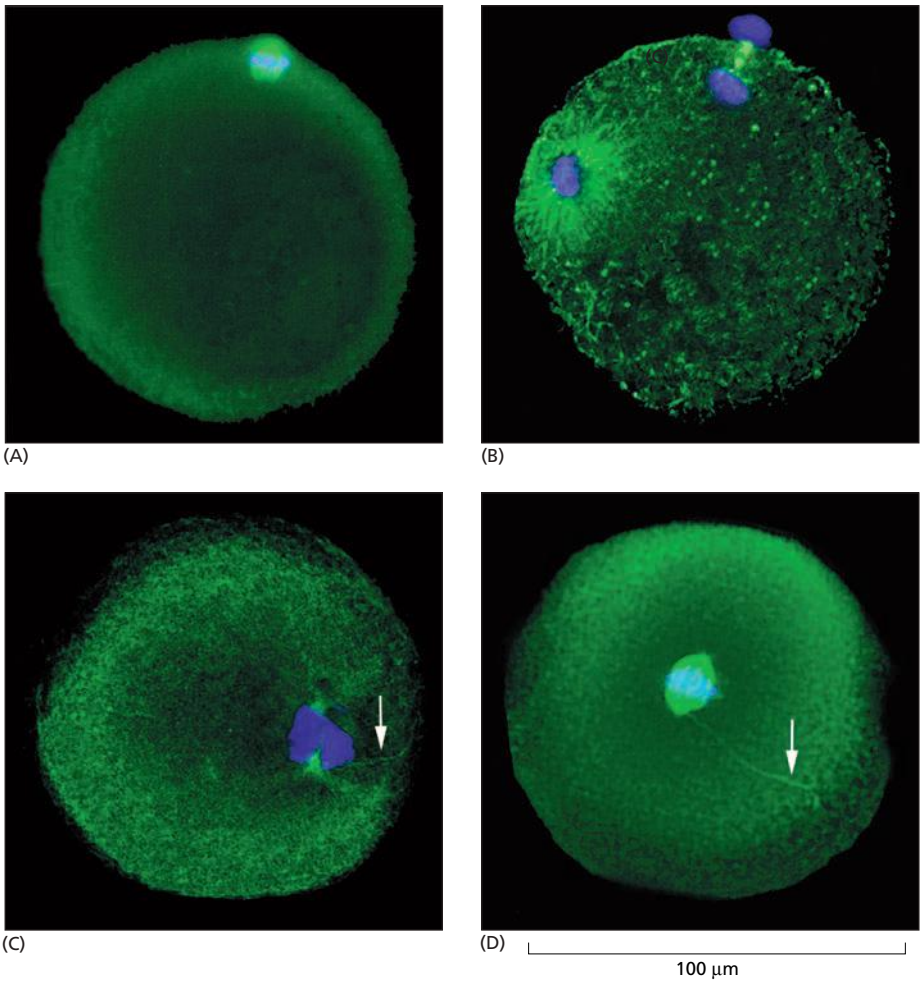


Figure 21–36 Immunofluorescence micrographs of human sperm and egg pronuclei coming together after *in vitro* fertilization. Spindle microtubules are stained in *green* with anti-tubulin antibodies, and DNA is labeled in *blue* with a DNA stain. (A) A meiotic spindle in a mature unfertilized secondary oocyte. (B) A fertilized egg extruding its second polar body about 5 hours after fusion with a sperm. The sperm head *(left)* has nucleated an array of microtubules. The egg and sperm pronuclei are still far apart. (C) The two pronuclei have come together. (D) By 16 hours after fusion with a sperm, the centrosome that entered the egg with the sperm has duplicated, and the daughter centrosomes have organized a bipolar mitotic spindle. The chromosomes of both pronuclei are aligned at the metaphase plate of the spindle. As indicated by the arrows in (C) and (D), the sperm tail is still associated with one of the centrosomes. (From C. Simerly et al., *Nat. Med.* 1:47–53, 1995. With permission from Macmillan Publishers Ltd.)

3 of the best-looking early embryos are transferred with a catheter into the woman's uterus; the remaining embryos are usually kept frozen in liquid nitrogen, for further implantations, if necessary. The main complication of IVF is multiple pregnancies, which occur in over 30% of cases, compared with about 2% in unassisted pregnancies.

The IVF procedure just described has enabled many previously infertile women to produce normal children. It does not, however, solve the problem for infertile men who usually produce too few or abnormal sperm. A second breakthrough, which occurred in 1992, provided the solution for most such men. In this modification of IVF, called **intracytoplasmic sperm injection** (**ICSI**), an egg is fertilized by injecting a single sperm into it (**Figure 21–37**). This strategy eliminates the need for large numbers of motile sperm and bypasses many of the hurdles that a sperm normally has to clear to fertilize an egg, including capacitation, swimming to the egg, undergoing an acrosomal reaction, burrowing through the zona pellucida, and fusing with the egg plasma membrane. ICSI has a success rate of better than 50% and has produced more than 100,000 children.

In addition to revolutionizing the treatment of infertility, IVF has opened up many new possibilities for manipulating the reproductive process. It has, for example, made it possible for parents carrying a defective gene to avoid passing the gene on to their children, by screening IVF embryos for the gene before implanting them into the uterus.

As discussed earlier, *in vitro* techniques for handling mammalian eggs have made it possible to produce clones of many types of mammals by transferring the nucleus of a somatic cell from the animal to be cloned into an unfertilized egg that has had its own nucleus removed or destroyed. It is not an easy procedure; the success rate is low, and it is still uncertain whether a human could be cloned in the same way. Moreover, there are serious ethical arguments about

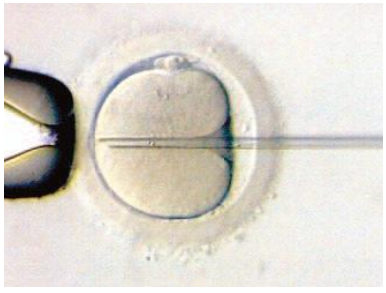


Figure 21–37 Intracytoplasmic sperm injection (ICSI). Light micrograph of a human secondary oocyte being held with a suction pipette (on the left) and injected with a single human sperm through a glass needle. The zona pellucida surrounds the egg and the polar body. (Courtesy of Reproductive Biology Associates, Atlanta, Georgia.)

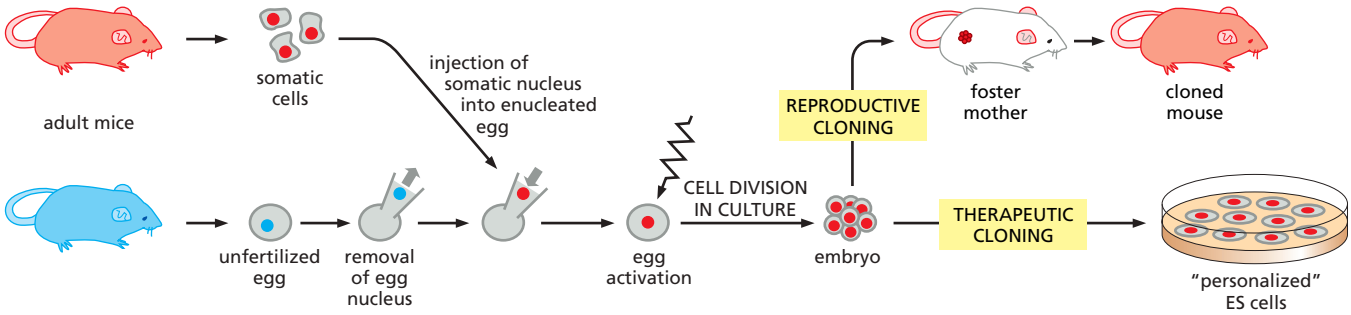


Figure 21–38 The difference between reproductive cloning and the preparation of "personalized" embryonic stem cells. In both cases, one produces a reconstructed embryo by removing (or destroying) the nucleus from an unfertilized egg and replacing it with the nucleus of a somatic cell from the animal to be cloned. The reconstructed egg is activated to develop by an electric shock. In *reproductive cloning*, the embryo that develops in culture is transplanted into the uterus of a foster mother and develops into a cloned animal. In the preparation of personalized embryonic stem (ES) cells (sometimes called *therapeutic cloning*), by contrast, the embryo is used to produce ES cells in culture, which can then be used to produce various specialized cell types for the treatment of the individual that provided the somatic nucleus; because the specialized cells produced by these ES cells are genetically the same as the donor of the somatic nucleus, they will not be immunologically rejected.

whether one should ever attempt to clone a human. There is general agreement, however, that it should not be attempted with existing technology, as the likelihood of producing an abnormal child is high; indeed, many countries and American states have made the attempt illegal.

Such *reproductive cloning*, however, should not be confused with *therapeutic cloning*, in which the early embryo produced *in vitro* from a such a reconstituted zygote is not implanted into a uterus to produce a new individual, but is instead used to make ES cells that are genetically the same as the donor of the somatic nucleus (**Figure 21–38**). Various types of specialized cells produced from such "personalized" ES cells could then be used to treat the donor, avoiding the problem of immunological rejection associated with using cells derived from genetically dissimilar ES cells. Clearly, societies will have to make some difficult decisions about how far they are willing to go in exploiting these new technologies to manipulate the reproductive process for the potential benefit of individuals. Alternatively, it might be possible in the future to produce personalized ES-like cells in ways that bypass these ethical dilemmas: in recent experiments, for example, genetic engineering was used to express in cultured mouse fibroblasts a number of gene regulatory proteins normally expressed in ES cells; when four such transgenes were expressed simultaneously, the fibroblasts behaved much like ES cells.

Fertilization marks the beginning of one of the most remarkable phenomena in all of biology—the process of embryogenesis, in which the zygote develops into a new individual. This is the subject of the next chapter.

Summary

Mammalian fertilization normally begins when a sperm, which has undergone capacitation in the female genital tract, binds to the zona pellucida surrounding an egg in the oviduct. This induces the sperm to undergo an acrosome reaction, releasing the contents of the acrosomal vesicle, which are thought to help the sperm to digest its way through the zona. The acrosome reaction is also required for the sperm to bind to and fuse with the egg plasma membrane. The fusion of the sperm with the egg induces a Ca^{2+} wave and oscillations in the egg cytosol, which activate the egg. The activation includes the egg cortical reaction, in which cortical granules release their contents, which alter the zona pellucida so that other sperm cannot bind to or penetrate it. The Ca^{2+} signal also triggers the development of the zygote, which begins after the two haploid pronuclei have come together and their chromosomes have aligned on a single mitotic spindle, which mediates the first mitotic division of the zygote. Many previously infertile couples can now reproduce thanks to IVF and ICSI.

REFERENCES

General

Austin CR and Short RV eds (1984) Reproduction in Mammals Vol I Germ Cells and Fertilization 2nd ed. Cambridge, UK: Cambridge University Press.

Gilbert SF (2006) Developmental Biology 8th ed pp 175–210, 529–554, 593–626. Sunderland, MA: Sinauer Associates, Inc.

Overview of Sexual Reproduction

Cavalier-Smith T (2002) Origins of the machinery of recombination and sex. *Heredity* 88:125–41.

Charlesworth B (2006) The evolutionary biology of sex. *Curr Biol* 16:R693–R695.

Hoekstra RF (2005) Evolutionary biology: why sex is good. *Nature* 434:571–573.

Maynard Smith J (1978) Evolution of Sex. Cambridge, UK: Cambridge University Press.

Meiosis

Blat Y, Protacio RU, Hunter N and Kleckner N (2002) Physical and functional interactions among basic chromosome organizational features govern early steps of meiotic chiasma formation. *Cell* 111:791–802.

Börner GV, Kleckner N and Hunter N (2004) Crossover/noncrossover differentiation, synaptonemal complex formation and regulatory surveillance at the leptotene/zygotene transition of meiosis. *Cell* 117:29–45.

De Massy B (2003) Distribution of meiotic recombination sites. *Trends Genet* 19:514–522.

Gerton JL, and Hawley RS (2005) Homologous chromosome interactions in meiosis: diversity amidst conservation. *Nature Rev Genet* 6:477–487.

Hall H, Hunt P and Hassold T (2006) Meiosis and sex chromosome aneuploidy: how meiotic errors cause aneuploidy; how aneuploidy causes meiotic errors. *Curr Opin Genet Dev* 16:323–329.

Hauf S and Watanabe Y (2004) Kinetochore orientation in mitosis and meiosis. *Cell* 119:317–327.

Hunt PA and Hassold TJ (2002) Sex matters in meiosis. *Science* 296:2181–2183.

Jordan P (2006) Initiation of homologous chromosome pairing during meiosis. *Biochem Soc Trans* 34:545–549.

Nasmyth K (2001) Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. *Annu Rev Genet* 35:673–745.

Page SL and Hawley RS (2004) The genetics and molecular biology of the synaptonemal complex. *Annu Rev Cell Dev Biol* 20:525–558.

Petronczki M, Siomos MF and Nasmyth K (2003) Un menage a quatre: the molecular biology of chromosome segregation in meiosis. *Cell* 112:423–440.

Primordial Germ Cells and Sex Determination in Mammals

Bowles J, Knight D, Smith C et al (2006) Retinoid signaling determines germ cell fate in mice. *Science* 312:596–600.

Extavour CG and Akam M (2003) Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. *Development* 130:5869–5884.

Goodfellow PN and Lovell-Badge R (1993) SRY and sex determination in mammals. *Annu Rev Genet* 27:71–92.

Knower KC, Kelly S and Harley VR (2003) Turning on the male—SRY, SOX9 and sex determination in mammals. *Cytogenet Genome Res* 101:185–198.

McLaren A (2003) Primordial germ cells in the mouse. *Dev Biol* 262:1–15.

Morrish BC and Sinclair AH (2002) Vertebrate sex determination: many means to an end. *Reproduction* 124:447–457.

Surani MA, Ancelin K, Hajkova P et al (2004) Mechanism of mouse germ cell specification: a genetic program regulating epigenetic reprogramming. *Cold Spring Harb Symp Quant Biol* 69:1–9.

Tilmann C and Capel B (2002) Cellular and molecular pathways regulating mammalian sex determination. *Recent Prog Horm Res* 57:1–18.

Eggs

Gosden RG (2002) Oogenesis as a foundation for embryogenesis. *Mol Cell Endocrinol* 186149–186153.

Matzuk MM, Burns KH, Viveiros MM and Eppig JJ (2002) Intercellular communication in the mammalian ovary: oocytes carry the conversation. *Science* 296:2178–2180.

Tanghe S, Van Soom A, Nauwynck H et al (2002) Minireview: Functions of the cumulus oophorus during oocyte maturation, ovulation, and fertilization. *Mol Reprod Dev* 61:414–424.

Wassarman PM, Jovine L and Litscher ES (2004) Mouse zona pellucida genes and glycoproteins. *Cytogenet Genome Res* 105:228–234.

Sperm

Clermont Y (1972) Kinetics of spermatogenesis in mammals: seminiferous epithelium cycle and spermatogonial renewal. *Physiol Rev* 52:198–236.

Moreno RD and Alvarado CP (2006) The mammalian acrosome as a secretory lysosome: new and old evidence. *Mol Reprod Dev* 73:1430–1434.

Ogawa T, Ohmura M and Ohbo K (2005) The niche for spermatogonial stem cells in the mammalian testis. *Int J Hematol* 82:381–388.

Toshimori K (2003) Biology of spermatozoa maturation: an overview with an introduction to this issue. *Microsc Res Tech* 61:1–6.

Fertilization

De Jonge C (2005) Biological basis for human capacitation. *Hum Reprod Update* 11:205–214.

Hafez ES, Goff L and Hafez B (2004) Mammalian fertilization, IVF, ICSI:physiological/molecular parameters, clinical application. *Arch Androl* 50:69–88.

Jaenisch R (2004) Human Cloning—The Science and Ethics of Nuclear Transplantation. *N Eng J Med* 351:2787–2791.

Shur BD, Rodeheffer C and Ensslin MA (2004) Mammalian fertilization. *Curr Biol* 14:R691–692.

Stein KK, Primakoff P and Myles D (2004) Sperm-egg fusion: events at the plasma membrane. *J Cell Sci* 117:6269–6274.

Takahashi K and Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126:663–676.

Tsaadon A, Eliyahu E, Shtraizent N and Shalgi R (2006) When a sperm meets an egg: block to polyspermy. *Mol Cell Endocrinol* 252:107–114.

Wassarman PM (2005) Contribution of mouse egg zona pellucida glycoproteins to gamete recognition during fertilization. *J Cell Physiol* 204:388–391.

Whitaker M (2006) Calcium at fertilization and in early development. *Physiol Rev* 86:25–88.

Development of Multicellular Organisms

22

An animal or plant starts its life as a single cell—a fertilized egg. During development, this cell divides repeatedly to produce many different cells in a final pattern of spectacular complexity and precision. Ultimately, the genome determines the pattern, and the puzzle of developmental biology is to understand how it does so.

The genome is normally identical in every cell; the cells differ not because they contain different genetic information, but because they express different sets of genes. This selective gene expression controls the four essential processes by which the embryo is constructed: (1) *cell proliferation*, producing many cells from one, (2) *cell specialization,* creating cells with different characteristics at different positions, (3) *cell interactions,* coordinating the behavior of one cell with that of its neighbors, and (4) *cell movement,* rearranging the cells to form structured tissues and organs (**Figure 22–1**).

In a developing embryo, all these processes are happening at once, in a kaleidoscopic variety of different ways in different parts of the organism. To understand the basic strategies of development, we have to narrow our focus. In particular, we must understand the course of events from the standpoint of the individual cell and the way the genome acts within it. There is no commanding officer standing above the fray to direct the troops; each of the millions of cells in the embryo has to make its own decisions, according to its own copy of the genetic instructions and its own particular circumstances.

The complexity of animals and plants depends on a remarkable feature of the genetic control system. Cells have a memory: the genes a cell expresses and the way it behaves depend on the cell's past as well as its present environment. The cells of your body—the muscle cells, the neurons, the skin cells, the gut cells, and so on—maintain their specialized characters not because they continually receive the same instructions from their surroundings, but because they retain a record of signals their ancestors received in early embryonic development. The molecular mechanisms of cell memory have been introduced in Chapter 7. In this chapter we shall encounter its consequences.

In This Chapter

UNIVERSAL MECHANISMS OF ANIMAL DEVELOPMENT 1305

CAENORHABDITIS ELEGANS: DEVELOPMENT FROM THE PERSPECTIVE OF THE INDIVIDUAL CELL 1321

DROSOPHILA AND THE MOLECULAR GENETICS OF PATTERN FORMATION: GENESIS OF THE BODY PLAN 1328

HOMEOTIC SELECTOR GENES AND THE PATTERNING OF THE ANTEROPOSTERIOR AXIS 1341

ORGANOGENESIS AND THE PATTERNING OF APPENDAGES 1347

CELL MOVEMENTS AND THE SHAPING OF THE VERTEBRATE BODY 1363

THE MOUSE 1378

NEURAL DEVELOPMENT 1383

PLANT DEVELOPMENT 1398

UNIVERSAL MECHANISMS OF ANIMAL DEVELOPMENT

There are about ten million species of animals, and they are fantastically varied. One would no more expect the worm, the flea, the eagle and the giant squid all to be generated by the same developmental mechanisms, than one would suppose that the same methods were used to make a shoe and an airplane. Some similar abstract principles might be involved, perhaps, but surely not the same specific molecules?

One of the most astonishing revelations of the past 10 or 20 years has been that our initial suspicions are wrong. In fact, much of the basic machinery of development is essentially the same, not just in all vertebrates but in all the major phyla of invertebrates too. Recognizably similar, evolutionarily related molecules define our specialized cell types, mark the differences between body regions, and help create the body's pattern. Homologous proteins are often

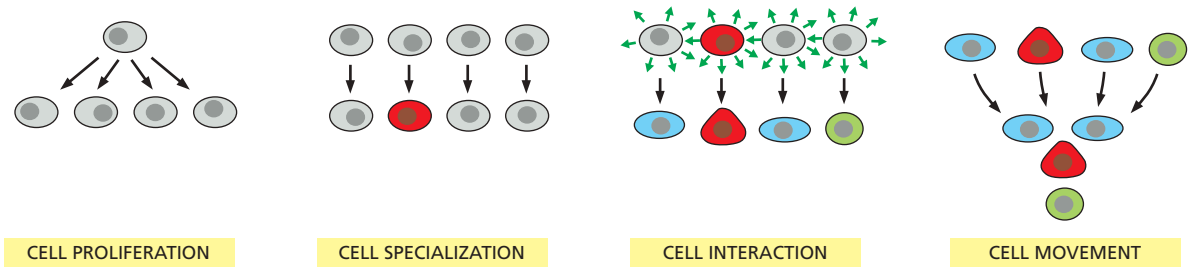


Figure 22–1 The four essential processes by which a multicellular organism is made: cell proliferation, cell specialization, cell interaction, and cell movement.

functionally interchangeable between very different species. A mouse protein produced artificially in a fly can often perform the same function as the fly's own version of that protein, and vice versa, successfully controlling the development of an eye, for example, or the architecture of the brain (**Figure 22–2**). Thanks to this underlying unity of mechanism, as we shall see, developmental biologists are now well on their way toward a coherent understanding of animal development.

Plants are a separate kingdom: they have evolved their multicellular organization independently of animals. For their development too, a unified account can be given, but it is different from that for animals. Animals will be our main concern in this chapter, but we shall return to plants briefly at the end.

We begin by reviewing some of the basic general principles of animal development and by introducing the seven animal species that developmental biologists have adopted as their chief model organisms.

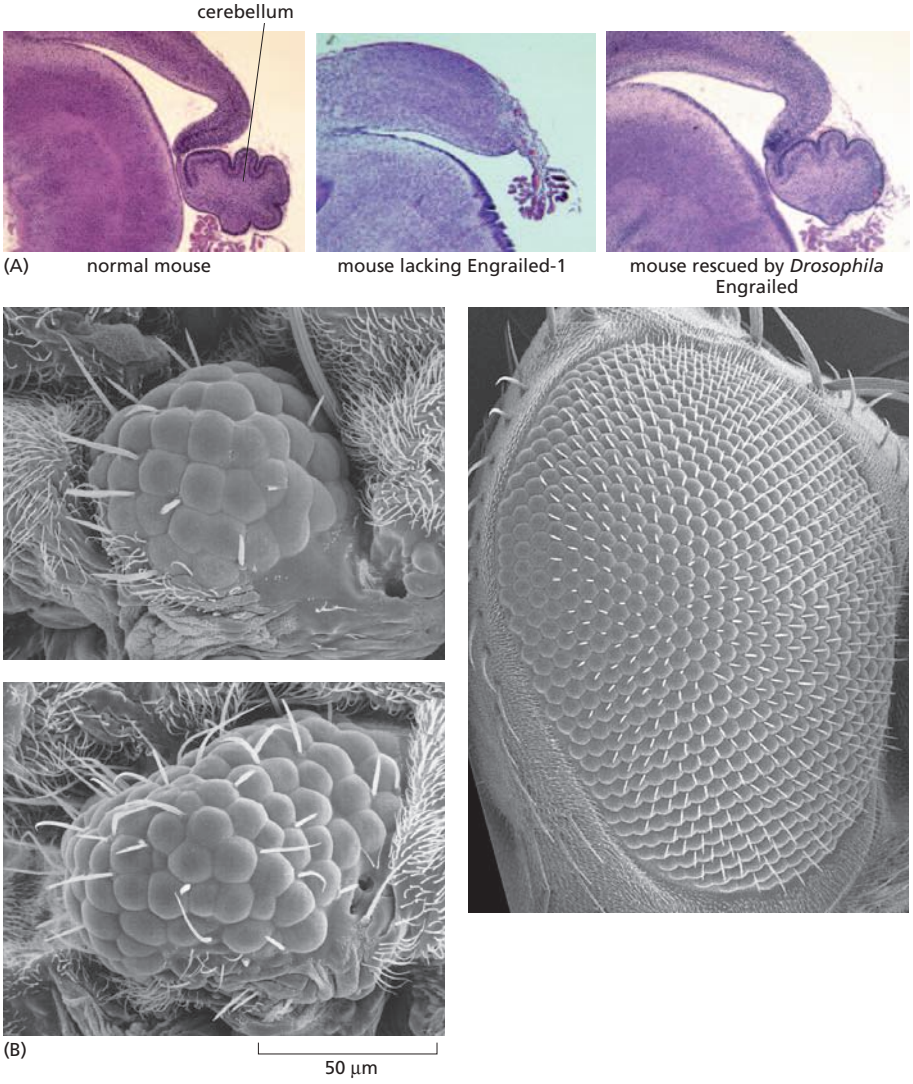


Figure 22–2 Homologous proteins functioning interchangeably in the development of mice and flies. (A) A fly protein used in a mouse. The DNA sequence from *Drosophila* coding for the Engrailed protein (a gene regulatory protein) can be substituted for the corresponding sequence coding for the Engrailed-1 protein of the mouse. Loss of Engrailed-1 in the mouse causes a defect in its brain (the cerebellum fails to develop); the *Drosophila* protein acts as an efficient substitute, rescuing the transgenic mouse from this deformity. (B) A mollusk protein used in a fly. The Eyeless protein controls eye development in *Drosophila,* and when misexpressed can cause an eye to develop in an abnormal site, such as a leg. The homologous protein, Pax6, from a mouse, a squid, or practically any animal possessing eyes, when similarly misexpressed in a transgenic fly, has the same effect. The scanning electron micrographs show a patch of eye tissue on the leg of a fly resulting from misexpression of *Drosophila Eyeless* (top) and of squid *Pax6* (bottom). The right panel shows, at lower magnification, the entire eye of a normal *Drosophila,* for comparison. (A, from M.C. Hanks et al., *Development* 125:4521–4530, 1998. With permission from The Company of Biologists; B, from S.I. Tomarev et al., *Proc. Natl Acad. Sci. U.S.A.* 94:2421–2426, 1997. With permission from National Academy of Sciences and courtesy of Kevin Moses.)

Animals Share Some Basic Anatomical Features

The similarities between animal species in the genes that control development reflect the evolution of animals from a common ancestor in which these genes were already present. Although we do not know what it looked like, the common ancestor of worms, mollusks, insects, vertebrates, and other complex animals must have had many differentiated cell types that would be recognizable to us: epidermal cells, for example, forming a protective outer layer; gut cells to absorb nutrients from ingested food; muscle cells to move; neurons and sensory cells to control the movements. The body must have been organized with a sheet of skin covering the exterior, a mouth for feeding and a gut tube to contain and process the food—with muscles, nerves and other tissues arranged in the space between the external sheet of skin and the internal gut tube.

These features are common to almost all animals, and they correspond to a common basic anatomical scheme of development. The egg cell—a giant storehouse of materials—divides, or **cleaves**, to form many smaller cells. <ATTT> These cohere to create an epithelial sheet facing the external medium. Much of this sheet remains external, constituting the **ectoderm**—the precursor of the epidermis and of the nervous system. A part of the sheet becomes tucked into the interior to form **endoderm**—the precursor of the gut and its appendages, such as lung and liver. Another group of cells move into the space between ectoderm and endoderm, and form the **mesoderm**—the precursor of muscles, connective tissues, and various other components. This transformation of a simple ball or hollow sphere of cells into a structure with a gut is called **gastrulation** (from the Greek word for a belly), and in one form or another it is an almost universal feature of animal development. **Figure 22–3** illustrates the process as it is seen in the sea urchin.

Evolution has diversified upon the molecular and anatomical fundamentals that we describe in this chapter to produce the wonderful variety of present-day species. But the underlying conservation of genes and mechanisms means that studying the development of one animal very often leads to general insights into

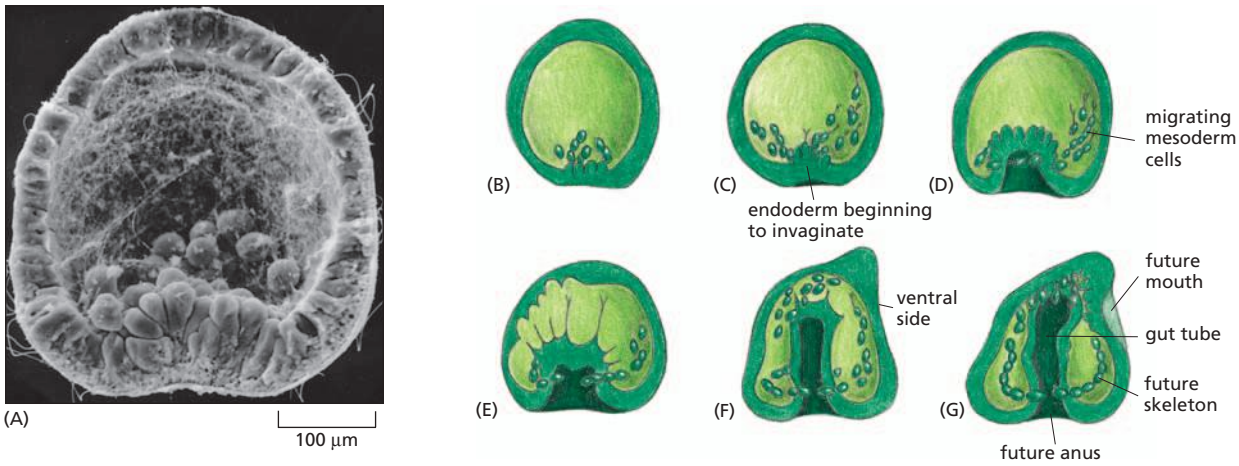
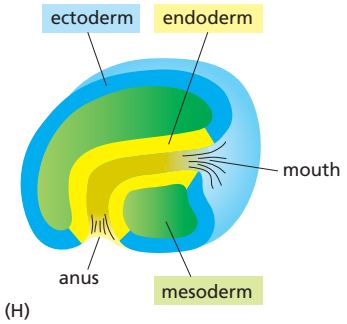


Figure 22–3 Sea urchin gastrulation. A fertilized egg divides to produce a blastula—a hollow sphere of epithelial cells surrounding a cavity. Then, in the process of gastrulation, some of the cells tuck into the interior to form the gut and other internal tissues. (A) Scanning electron micrograph showing the initial intucking of the epithelium. (B) Drawing showing how a group of cells break loose from the epithelium to become mesoderm. (C) These cells then crawl over the inner face of the wall of the blastula. (D) Meanwhile, epithelium is continuing to tuck inward to become endoderm. (E and F) The invaginating endoderm extends into a long gut tube. (G) The end of the gut tube makes contact with the wall of the blastula at the site of the future mouth opening. Here the ectoderm and endoderm will fuse and a hole will form. (H) The basic animal body plan, with a sheet of ectoderm on the outside, a tube of endoderm on the inside, and mesoderm sandwiched between them. (A, from R.D. Burke et al., *Dev. Biol.* 146:542–557, 1991. With permission from Academic Press; B-G, after L. Wolpert and T. Gustafson, *Endeavour* 26:85–90, 1967. With permission from Elsevier.)



the development of many other types of animals. As a result, developmental biologists today, like cell biologists, have the luxury of addressing fundamental questions in whatever species offers the easiest path to an answer.

Multicellular Animals Are Enriched in Proteins Mediating Cell Interactions and Gene Regulation

Genome sequencing reveals the extent of molecular similarities between species. The nematode worm *Caenorhabditis elegans*, the fly *Drosophila melanogaster*, and the vertebrate *Homo sapiens* are the first three animals for which a complete genome sequence was obtained. In the family tree of animal evolution, they are very distant from one another: the lineage leading to the vertebrates is thought to have diverged from that leading to the nematodes, insects and mollusks more than 600 million years ago. Nevertheless, when the 20,000 genes of *C. elegans,* the 14,000 genes of *Drosophila*, and the 25,000 genes of the human are systematically compared with one another, it is found that about 50% of the genes in each of these species have clearly recognizable homologs in one or both of the other two species. In other words, recognizable versions of at least 50% of all human genes were already present in the common ancestor of worms, flies, and humans.

Of course, not everything is conserved: there are some genes with key roles in vertebrate development that have no homologs in the genome of *C. elegans* or *Drosophila*, and vice versa. However, a large proportion of the 50% of genes that lack identifiable homologs in other phyla may do so simply because their functions are of minor importance. Although these nonconserved genes are transcribed and well represented in cDNA libraries, studies of DNA and amino acid sequence variability in and between natural populations indicate that these genes are unusually free to mutate without seriously harming fitness; when they are artificially inactivated, the consequences are not so often severe as for genes with homologs in distantly related species. Because they are free to evolve so rapidly, a few tens of millions of years may be enough to obliterate any family resemblance or to permit loss from the genome.

The genomes of different classes of animals differ also because, as discussed in Chapter 1, there are substantial variations in the extent of gene duplication: the amount of gene duplication in the evolution of the vertebrates has been particularly large, with the result that a mammal or a fish often has several homologs corresponding to a single gene in a worm or a fly.

Despite such differences, to a first approximation we can say that all these animals have a similar set of proteins at their disposal for their key functions. In other words, they construct their bodies using roughly the same molecular kit of parts.

What genes, then, are needed to produce a multicellular animal, beyond those necessary for a solitary cell? Comparison of animal genomes with that of budding yeast—a unicellular eucaryote—suggests that two classes of proteins are especially important for multicellular organization. The first class is that of the transmembrane molecules used for cell adhesion and cell signaling. As many as 2000 *C. elegans* genes encode cell surface receptors, cell adhesion proteins, and ion channels that are either not present in yeast or present in much smaller numbers. The second class is that of gene regulatory proteins: these DNA-binding proteins are much more numerous in the *C. elegans* genome than in yeast. For example, the basic helix–loop–helix family has 41 members in *C. elegans*, 84 in *Drosophila*, 131 in humans and only 7 in yeast, and other families of regulators of gene expression are also dramatically overrepresented in animals as compared to yeast. Not surprisingly, these two classes of proteins are central to developmental biology: as we shall see, the development of multicellular animals is dominated by cell–cell interactions and by differential gene expression.

As discussed in Chapter 7, micro-RNAs also play a significant part in controlling gene expression during development, but they seem to be of secondary importance by comparison with proteins. Thus a mutant zebrafish embryo that completely lacks the Dicer enzyme, which is required for production of functional miRNAs, will still begin its development almost normally, creating

specialized cell types and a more-or-less correctly organized body plan, before abnormalities become severe.

Regulatory DNA Defines the Program of Development

A worm, a fly, a mollusc and a mammal share many of the same essential cell types, and they do all have a mouth, a gut, a nervous system and a skin; but beyond a few such basic features they seem radically different in their body structure. If the genome determines the structure of the body and these animals all have such a similar collection of genes, how can they be so different?

The proteins encoded in the genome can be viewed as the components of a construction kit. Many things can be built with this kit, just as a child's construction kit can be used to make trucks, houses, bridges, cranes, and so on by assembling the components in different combinations. Some components necessarily go together—nuts with bolts, wheels with tires and axles—but the large-scale organization of the final object is not defined by these substructures. Rather, it is defined by the instructions that accompany the components and prescribe how they are to be assembled.

To a large extent, the instructions needed to produce a multicellular animal are contained in the noncoding, regulatory DNA that is associated with each gene. As discussed in Chapter 4, each gene in a multicellular organism is associated with thousands or tens of thousands of nucleotides of noncoding DNA. This DNA may contain, scattered within it, dozens of separate regulatory elements or *enhancers*—short DNA segments that serve as binding sites for specific complexes of gene regulatory proteins. Roughly speaking, as explained in Chapter 7, the presence of a given regulatory module of this sort leads to expression of the gene whenever the complex of proteins recognizing that segment of DNA is appropriately assembled in the cell (in some cases, an inhibition or a more complicated effect on gene expression is produced instead). If we could decipher the full set of regulatory modules associated with a gene, we would understand all the different molecular conditions under which the product of that gene is to be made. This regulatory DNA can therefore be said to define the sequential program of development: the rules for stepping from one state to the next, as the cells proliferate and read their positions in the embryo by reference to their surroundings, switching on new sets of genes according to the activities of the proteins that they currently contain (**Figure 22–4**). Variations in the proteins themselves do, of course, also contribute to the differences between species. But even if the set of proteins encoded in the genome remained completely unchanged, the variation in the regulatory DNA would be enough to generate radically different tissues and body structures.

When we compare animal species with similar body plans—different vertebrates such as a fish, a bird and a mammal, for example—we find that corresponding genes usually have similar sets of regulatory modules: the DNA sequences of many of the individual modules have been well conserved and are recognizably homologous in the different animals. The same is true if we compare different species of nematode worm, or different species of insect. But when we compare vertebrate regulatory regions with those of worm or fly, it is

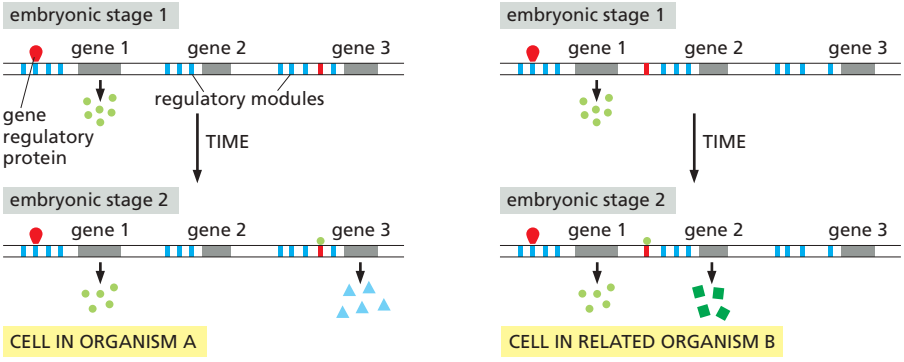


Figure 22–4 How regulatory DNA defines the succession of gene expression patterns in development. The genomes of organisms A and B code for the same set of proteins but have different regulatory DNA. The two cells in the cartoon start in the same state, expressing the same proteins at stage 1, but step to quite different states at stage 2 because of their different arrangements of regulatory modules.

hard to see any such resemblance. The protein-coding sequences are unmistakably similar, but the corresponding regulatory DNA sequences appear very different. This is the expected result if different body plans are produced mainly by changing the program embodied in the regulatory DNA, while retaining most of the same kit of proteins.

Manipulation of the Embryo Reveals the Interactions Between Its Cells

Confronted with an adult animal, in all its complexity, how does one begin to analyze the process that brought it into being? The first essential step is to describe the anatomical changes—the patterns of cell division, growth, and movement—that convert the egg into the mature organism. This is the job of *descriptive embryology*, and it is harder than one might think. To explain development in terms of cell behavior, we need to be able to track the individual cells through all their divisions, transformations, and migrations in the embryo. The foundations of descriptive embryology were laid in the nineteenth century, but the fine-grained task of *cell lineage tracing* continues to tax the ingenuity of developmental biologists (**Figure 22–5**)

Given a description, how can one go on to discover the causal mechanisms? Traditionally, *experimental embryologists* have tried to understand development in terms of the ways in which cells and tissues interact to generate the multicellular structure. *Developmental geneticists*, meanwhile, have tried to analyze development in terms of the actions of genes. These two approaches are complementary, and they have converged to produce our present understanding.

In experimental embryology, cells and tissues from developing animals are removed, rearranged, transplanted, or grown in isolation, in order to discover how they influence one another. The results are often startling: an early embryo cut in half, for example, may yield two complete and perfectly formed animals, or a small piece of tissue transplanted to a new site may reorganize the whole structure of the developing body (**Figure 22–6**). Observations of this type can be

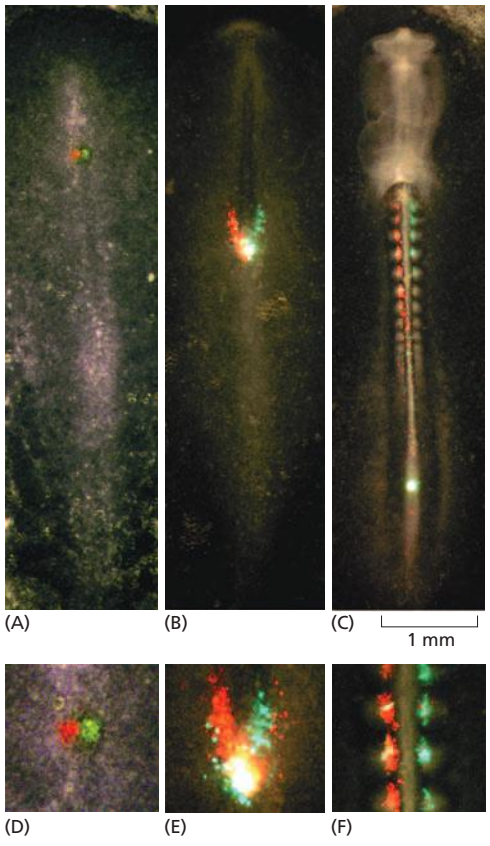


Figure 22–5 Cell lineage tracing in the early chick embryo. The pictures in the top row are at low magnification and show the whole embryo; the pictures below are details, showing the distribution of labeled cells. The tracing experiment reveals complex and dramatic cell rearrangements. (A,D) Two tiny dots of fluorescent dye, one *red,* the other *green,* have been used to stain small groups of cells in an embryo at 20 hours of incubation. Though the embryo still appears as an almost featureless sheet of tissue, there is already some specialization. The dots have been placed on each side of a structure called the node. (B,E) Six hours later, some of the labeled cells have remained at the node (which has moved backwards), giving a bright spot of fluorescence there, while other cells have begun to move forwards relative to the node. (C,F) After a further 8 hours, the body plan is clearly visible, with a head at the front end *(top),* a central axis, and rows of embryonic body segments, called somites, on either side of this. The node has regressed still further tailwards; some of the originally labeled cells have stayed in the node, forming a bright spot of fluorescence, while others have migrated to positions far anterior to this and become parts of somites. (Courtesy of Raquel Mendes and Leonor Saúde.)

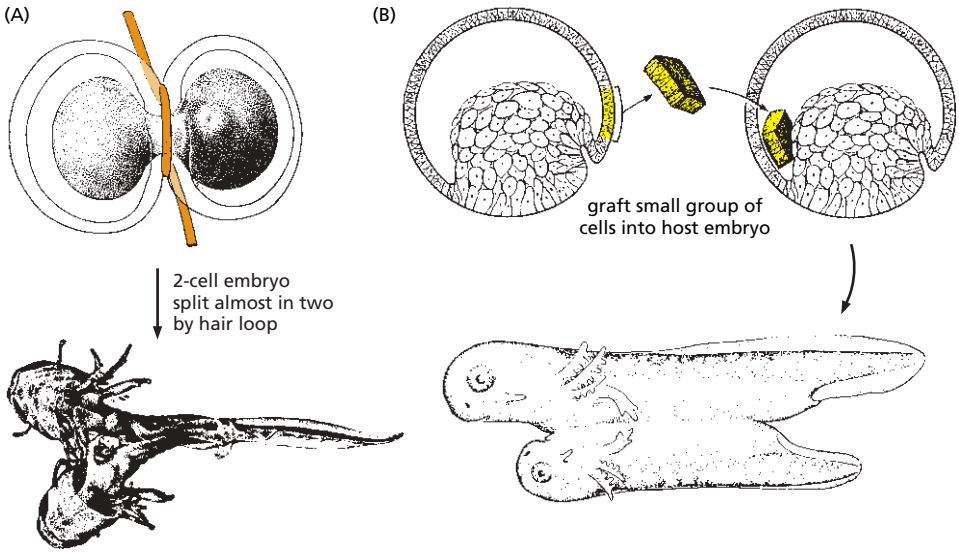


Figure 22–6 Some striking results obtained by experimental embryology. <ATTG> In (A), an early amphibian embryo is split almost into two parts with a hair loop. In (B), an amphibian embryo at a somewhat later stage receives a graft of a small cluster of cells from another embryo at that stage. The two quite different operations both cause a single embryo to develop into a pair of conjoined (Siamese) twins. It is also possible in experiment (A) to split the early embryo into two completely separate halves; two entire separate well-formed tadpoles are then produced. (A, after H. Spemann, Embryonic Development and Induction. New Haven: Yale University Press, 1938; B, after J. Holtfreter and V. Hamburger, in Analysis of Development [B.H. Willier, P.A. Weiss and V. Hamburger, eds.], pp. 230–296. Philadelphia: Saunders, 1955.)

extended and refined to decipher the underlying cell–cell interactions and rules of cell behavior. The experiments are easiest to perform in large embryos that are readily accessible for microsurgery. Thus, the most widely used species have been birds—especially the chick—and amphibians—particularly the African frog *Xenopus laevis*.

Studies of Mutant Animals Identify the Genes That Control Developmental Processes

Developmental genetics begins with the isolation of mutant animals whose development is abnormal. This typically involves a *genetic screen*, as described in Chapter 8. Parent animals are treated with a chemical mutagen or ionizing radiation to induce mutations in their germ cells, and large numbers of their progeny are examined. Those rare mutant individuals that show some interesting developmental abnormality—altered development of the eye, for example—are picked out for further study. In this way, it is possible to discover genes that are required specifically for the normal development of any chosen feature. By cloning and sequencing a gene found in this way, it is possible to identify its protein product, to investigate how it works, and to begin an analysis of the regulatory DNA that controls its expression.

The genetic approach is easiest in small animals with short generation times that can be grown in the laboratory. The first animal to be studied in this way was the fruit fly *Drosophila melanogaster,* which will be discussed at length below. But the same approach has been successful in the nematode worm, *Caenorhabditis elegans,* the zebrafish, *Danio rerio,* and the mouse, *Mus musculus*. Although humans are not intentionally mutagenized, they get screened for abnormalities in enormous numbers through the medical care system. Many mutations have arisen in humans that cause abnormalities compatible with life, and analyses of the affected individuals and of their cells have provided important insights into developmental processes.

A Cell Makes Developmental Decisions Long Before It Shows a Visible Change

By simply watching closely, or with the help of tracer dyes and other cell-marking techniques, one can discover what the fate of a given cell in an embryo will be if that embryo is left to develop normally. The cell may be fated to die, for example, or to become a neuron, to form part of an organ such as the foot, or to give progeny cells scattered all over the body. To know the **cell fate**, in this sense, however, is to know next to nothing about the cell's intrinsic character. At one

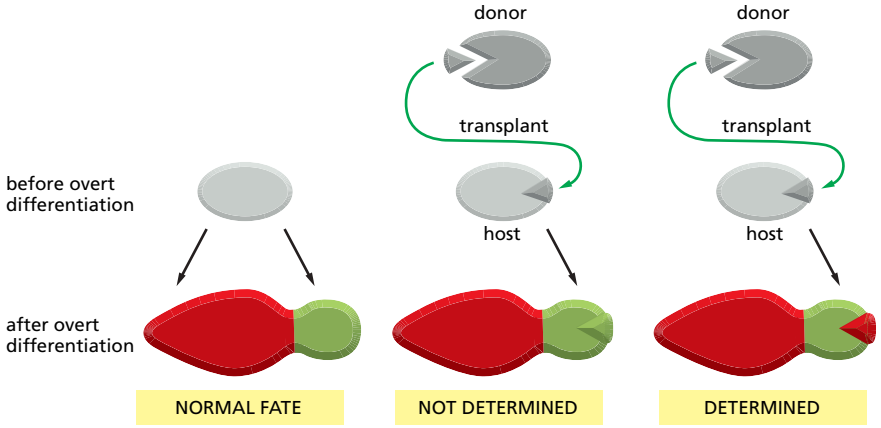


Figure 22–7 The standard test for cell determination.

extreme, the cell that is fated to become, say, a neuron may be already specialized in a way that guarantees that it will become a neuron no matter how its surroundings are disturbed; such a cell is said to be **determined** for its fate. At the opposite extreme, the cell may be biochemically identical to other cells destined for other fates, the only difference between them being the accident of position, which exposes the cells to different future influences.

A cell's state of determination can be tested by transplanting it to altered environments (**Figure 22–7**). One of the key conclusions of experimental embryology has been that, thanks to cell memory, a cell can become determined long before it shows any obvious outward sign of differentiation.

Between the extremes of the fully determined and the completely undetermined cell, there is a whole spectrum of possibilities. A cell may, for example, be already somewhat specialized for its normal fate, with a strong tendency to develop in that direction, but still able to change and undergo a different fate if it is put in a sufficiently coercive environment. (Some developmental biologists would describe such a cell as *specified* or *committed*, but not yet determined.) Or the cell may be determined, say, as a brain cell, but not yet determined as to whether it is to be a neuronal or a glial component of the brain. And often, it seems, adjacent cells of the same type interact and depend on mutual support to maintain their specialized character, so that they will behave as determined if kept together in a cluster, but not if taken singly and isolated from their usual companions.

Cells Have Remembered Positional Values That Reflect Their Location in the Body

In many systems, long before cells become committed to differentiating as a specific cell type, they become *regionally determined*: that is, they switch on and maintain expression of genes that can best be regarded as markers of position or region in the body. This position-specific character of a cell is called its **positional value**, and it shows its effects in the way the cell behaves in subsequent steps of pattern formation.

The development of the chick leg and wing provides a striking example. The leg and the wing of the adult both consist of muscle, bone, skin, and so on—almost exactly the same range of differentiated tissues. The difference between the two limbs lies not in the types of tissues, but in the way in which those tissues are arranged in space. So how does the difference come about?

In the chick embryo the leg and the wing originate at about the same time in the form of small tongue-shaped buds projecting from the flank. The cells in the two pairs of limb buds appear similar and uniformly undifferentiated at first. But a simple experiment shows that this appearance of similarity is deceptive. A small block of undifferentiated tissue at the base of the leg bud, from the region that would normally give rise to part of the thigh, can be cut out and grafted into the tip of the wing bud. Remarkably, the graft forms not the appropriate part of the wing tip, nor a misplaced piece of thigh tissue, but a toe (**Figure 22–8**). This

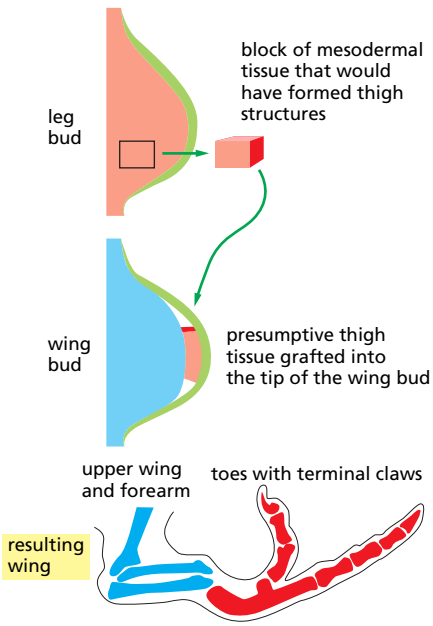


Figure 22–8 Prospective thigh tissue grafted into the tip of a chick wing bud forms toes. (After J.W. Saunders et al., *Dev. Biol.* 1:281–301, 1959. With permission from Academic Press.)



Figure 22–9 Chick embryos at 6 days of incubation, showing the limb buds stained by *in situ* hybridization with probes to detect expression of the *Tbx4, Tbx5,* and *Pitx1* genes, all coding for related gene regulatory proteins. The cells expressing *Tbx5* will form a wing; those expressing *Tbx4* and *Pitx1* will form a leg. *Pitx1*, when artificially misexpressed in the wing bud, causes the limb to develop with leg-like characteristics. (Courtesy of Malcolm Logan.)

experiment shows that the early leg-bud cells are already determined as leg but are not yet irrevocably committed to form a particular part of the leg: they can still respond to cues in the wing bud so that they form structures appropriate to the tip of the limb rather than the base. The signaling system that controls the differences between the parts of the limb is apparently the same for leg and wing. The difference between the two limbs results from a difference in the internal states of their cells at the outset of limb development.

The difference of positional value between vertebrate forelimb cells and hindlimb cells corresponds to expression of different sets of genes, coding for gene regulatory proteins that are thought to make the cells in the two limb buds behave differently (**Figure 22–9**).Later in this chapter we shall explain how the next, more detailed level of patterning is set up inside an individual limb bud.

Inductive Signals Can Create Orderly Differences Between Initially Identical Cells

At each stage in its development, a cell in an embryo is presented with a limited set of options according to the state it has attained: the cell travels along a developmental pathway that branches repeatedly. At each branch in the pathway it has to make a choice, and its sequence of choices determines its final destiny. In this way, a complicated array of different cell types is produced.

To understand development, we need to know how each choice between options is controlled, and how those options depend on the choices made previously. To reduce the question to its simplest form: how do two cells with the same genome, but separated in space, come to be different?

The most straightforward way to make cells different is by exposing them to different environments, and the most important environmental cues acting on cells in an embryo are signals from neighboring cells. Thus, in what is probably the commonest mode of pattern formation, a group of cells start out all having the same developmental potential, and a signal from cells outside the group then drives one or more of the members of the group into a different developmental pathway, leading to a changed character. This process is called an **inductive interaction**. Generally, the signal is limited in time and space so that only a subset of the competent cells—those closest to the source of the signal—take on the induced character (**Figure 22–10**).

Some inductive signals are short-range—notably those transmitted via cell–cell contacts; others are long-range, mediated by molecules that can diffuse through the extracellular medium. The group of initially similar cells competent to respond to the signal is sometimes called an *equivalence group* or a *morphogenetic field*. It can consist of as few as two cells or as many as thousands, and any number of the total can be induced depending on the amount and distribution of the signal.

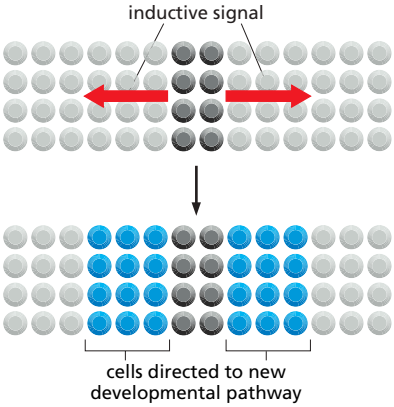


Figure 22–10 Inductive signaling.

Sister Cells Can Be Born Different by an Asymmetric Cell Division

Cell diversification does not always have to depend on extracellular signals: in some cases, sister cells are born different as a result of an **asymmetric cell division**, in which some significant set of molecules is divided unequally between

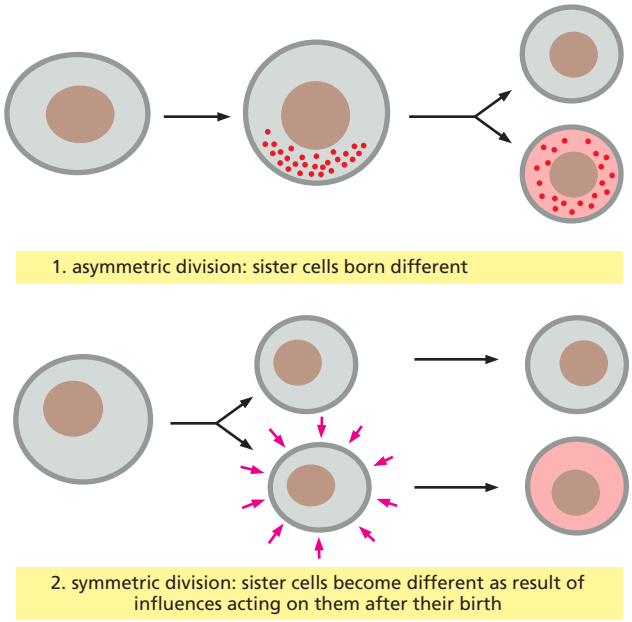


Figure 22–11 Two ways of making sister cells different.

the two of them at the time of division. This asymmetrically segregated molecule (or set of molecules) then acts as a *determinant* for one of the cell fates by directly or indirectly altering the pattern of gene expression within the daughter cell that receives it (**Figure 22–11**).

Asymmetric divisions often occur at the beginning of development, when the fertilized egg divides to give daughter cells with different fates, but they are also encountered at some later stages—in the genesis of nerve cells, for example.

Positive Feedback Can Create Asymmetry Where There Was None Before

Inductive signaling and asymmetric cell division represent two distinct strategies for creating differences between cells. Both of them, however, presuppose some prior asymmetry in the system: the source of inductive signal must be localized so that some cells receive the signal strongly and others do not; or the mother cell must already have an internal asymmetry before she divides. Very often, the history of the system ensures that some such asymmetry will be present. But what if it is not, or if the initial asymmetry is only very slight?

The answer lies in **positive feedback**: through positive feedback, a system that starts off homogeneous and symmetrical can pattern itself spontaneously, even where there is no organized external signal at all. And where, as very often happens, the environment or the starting conditions impose some weak but definite initial asymmetry, positive feedback provides the means to magnify the effect and create a full-blown pattern.

To illustrate the idea, consider a pair of adjacent cells that start off in a similar state and can exchange signals to influence one another's behavior (**Figure 22–12**). The more that either cell produces of some product X, the more it signals to its neighbor to inhibit production of X by the neighbor. This type of cell–cell interaction is called *lateral inhibition*, and it gives rise to a positive feedback loop that tends to amplify any initial difference between the two cells. Such a difference may arise from a bias imposed by some external or prior factor, or it may simply originate from spontaneous random fluctuations, or "noise"—an inevitable feature of the genetic control circuitry in cells, as discussed in Chapter 7. In either case, lateral inhibition means that if cell #1 makes a little more of X, it will thereby cause cell #2 to make less; and because cell #2 makes less X, it delivers less inhibition to cell #1 and so allows the amount of X in cell #1 to rise higher still; and so on, until a steady state is reached where cell #1 contains a lot of X and cell #2 contains very little.

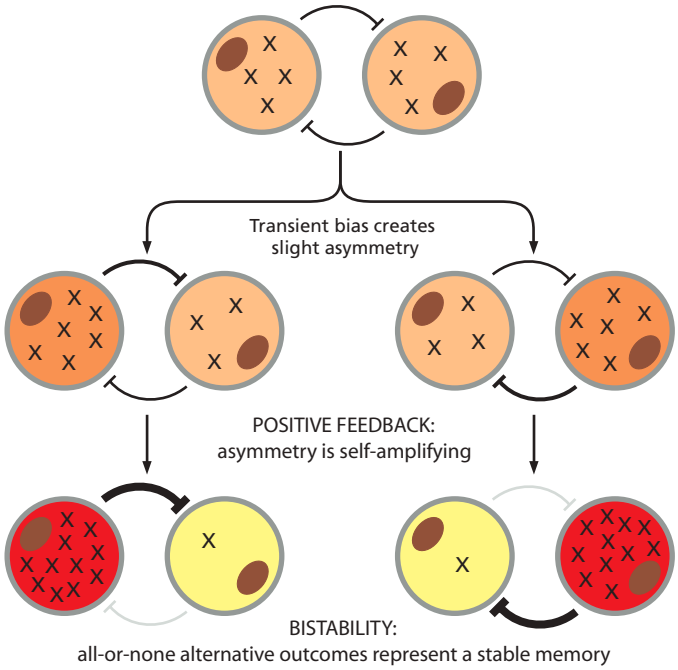


Figure 22–12 Genesis of asymmetry through positive feedback. In this example, two cells interact, each producing a substance X that acts on the other cell to inhibit its production of X, an effect known as lateral inhibition. An increase of X in one of the cells leads to a positive feedback that tends to increase X in that cell still further, while decreasing X in its neighbor. This can create a runaway instability, making the two cells become radically different. Ultimately the system comes to rest in one or the other of two opposite stable states. The final choice of state represents a form of memory: the small influence that initially directed the choice is no longer required to maintain it.

Mathematical analysis shows that this phenomenon depends on the strength of the lateral inhibition effect: if it is too weak, fluctuations will fade and have no lasting effect; but if it is strong enough and steep enough, they will be self-amplifying in a runaway fashion, breaking the initial symmetry between the two cells. Lateral inhibition, often mediated by exchange of signals at cell–cell contacts via the Notch signaling pathway (as discussed in Chapter 15), is a common mechanism for cell diversification in animal tissues, driving neighboring cells to specialize in different ways.

Positive Feedback Generates Patterns, Creates All-or-none Outcomes, and Provides Memory

Somewhat similar positive feedback processes can operate over larger arrays of cells to create many types of spatial patterns. For example, a substance A (a short-range activator) may stimulate its own production in the cells that contain it and their immediate neighbors, while also causing them to produce a signal H (a long-range inhibitor) that diffuses widely and inhibits production of A in the cells at larger distances. If the cells all start out on an equal footing, but one group of cells gains a slight advantage by making a little more A than the rest, the asymmetry can be self-amplifying. Short-range activation combined with long-range inhibition in this way may account for the formation of clusters of cells within an initially homogeneous tissue that become specialized as localized *signaling centers*.

At the opposite end of the size spectrum, positive feedback can also be the means by which an individual cell becomes spontaneously polarized and internally asymmetrical, through systems of intracellular signals that make a weak initial asymmetry self-amplifying.

Through all these and many other variations on the theme of positive feedback, certain general principles apply. In each of the above examples, the positive feedback leads to *broken symmetry*, and the symmetry-breaking is an *all-or-none* phenomenon. If the feedback is below a certain threshold strength, the cells remain essentially the same; if the feedback is above the threshold, they become sharply different. Above this threshold, the system is *bistable* or *multistable*—it lurches toward one or other of two or more sharply different outcomes, according to which of the cells (or which of the ends of the single cell) gains the initial advantage.

The choice between the alternative outcomes can be dictated by an external signal that gives one of the cells a small initial advantage. But once the positive

feedback has done its work, this external signal becomes irrelevant. The broken symmetry, once established, is very hard to reverse: positive feedback makes the chosen asymmetric state self-sustaining, even after the biasing signal has disappeared. In this way, positive feedback provides the system with a *memory* of past signals.

All these effects of positive feedback—symmetry-breaking, all-or-none outcomes, bistability, and memory—go hand in hand and are encountered again and again in developing organisms. They are fundamental to the production of sharply delineated, stable patterns of cells in different states.

A Small Set of Signaling Pathways, Used Repeatedly, Controls Developmental Patterning

What, then, are the molecules that act as signals to coordinate spatial patterning in an embryo, either to create asymmetry *de novo,* or as inducers from established signaling centers to control the diversification of neighboring cells? In principle, any kind of extracellular molecule could serve. In practice, most of the known inductive events in animal development are governed by just a handful of highly conserved families of signal proteins, which are used over and over again in different contexts. The discovery of this limited vocabulary that cells use for developmental communications has emerged over the past 10 or 20 years as one of the great simplifying discoveries of developmental biology. In **Table 22–1**, we briefly review six major families of signal proteins that serve repeatedly as inducers in animal development. Details of the intracellular mechanisms through which these molecules act are given in Chapter 15.

The ultimate result of most inductive events is a change in DNA transcription in the responding cell: some genes are turned on and others are turned off. Different signaling molecules activate different kinds of gene regulatory proteins. Moreover, the effect of activating a given gene regulatory protein will depend on which other gene regulatory proteins are also present in the cell, since these generally function in combinations. As a result, different types of cells will generally respond differently to the same signal, and the same cells will often respond differently to the same signal given at a different time. The response will depend both on the other gene regulatory proteins that are present before the signal arrives—reflecting the cell's memory of signals received previously—and on the other signals that the cell is receiving concurrently.

Morphogens Are Long-Range Inducers That Exert Graded Effects

Signal molecules often seem to govern a simple yes–no choice: one outcome when their concentration is high, another when it is low. Positive feedback can

Table 22–1 Some Signal Proteins That Are Used Over and Over Again as Inducers in Animal Development

SIGNALING PATHWAY	LIGAND FAMIILY	RECEPTOR FAMILY	EXTRACELLULAR INHIBITORS/MODULATORS
Receptor tyrosine kinase (RTK)	EGF	EGF receptors	Argos
	FGF (Branchless)	FGF receptors (Breathless)	
	Ephrins	Eph receptors	
TGFβ superfamily	TGFβ	TGFβ receptors	chordin (Sog), noggin
	BMP (Dpp)	BMP receptors	
	Nodal		
Wnt	Wnt (Wingless)	Frizzled	Dickkopf, Cerberus
Hedgehog	Hedgehog	Patched, Smoothened	
Notch	Delta	Notch	Fringe

Only a few representatives of each class of proteins are listed—mainly those mentioned in this chapter. Names peculiar to *Drosophila* are shown in parentheses. Many of the listed components have several homologs distinguished by numbers (FGF1, FGF2, etc.) or by forenames (Sonic hedgehog, Lunatic fringe). Other signaling pathways, including the JAK/STAT, nuclear hormone receptor, and G-protein-coupled receptor pathways, also play important parts in some developmental processes.

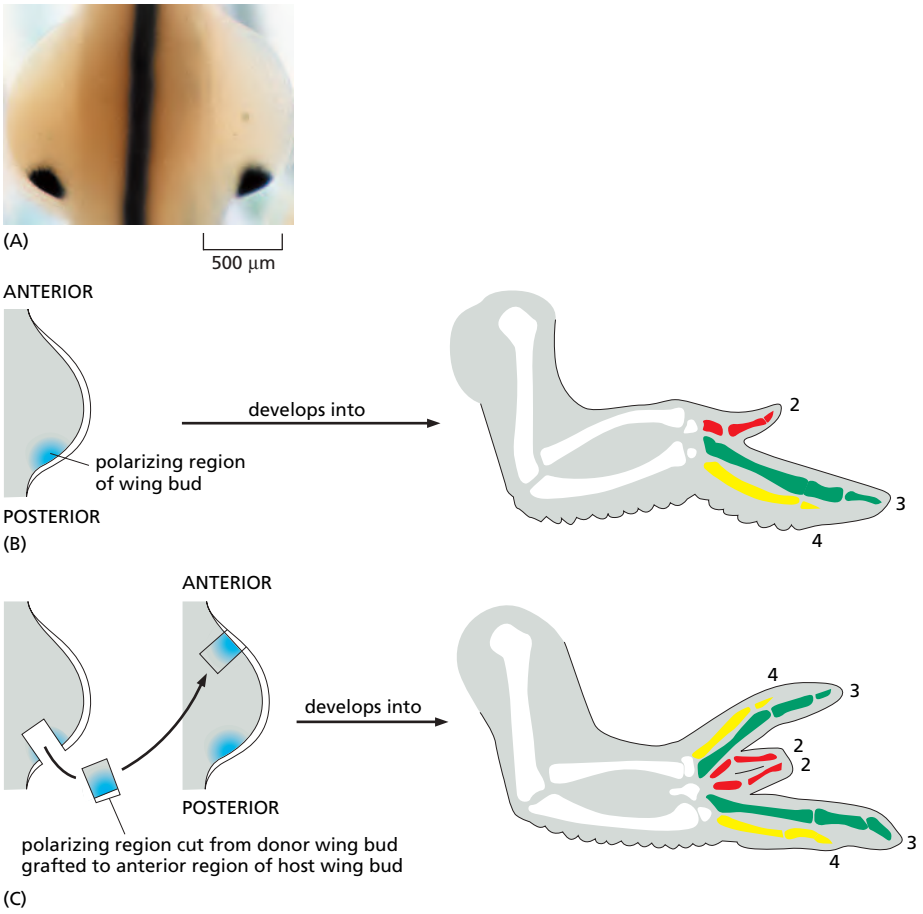


Figure 22–13 Sonic hedgehog as a morphogen in chick limb development. (A) Expression of the Sonic hedgehog gene in a 4-day chick embryo, shown by *in situ* hybridization (dorsal view of the trunk at the level of the wing buds). The gene is expressed in the midline of the body and at the posterior border (the polarizing region) of each of the two wing buds. Sonic hedgehog protein spreads out from these sources. (B) Normal wing development. (C) A graft of tissue from the polarizing region causes a mirror-image duplication of the pattern of the host wing. The type of digit that develops is thought to be dictated by the local concentration of Sonic hedgehog protein; different types of digit (labeled 2, 3, and 4) therefore form according to their distance from a source of Sonic hedgehog. (A, courtesy of Randall S. Johnson and Robert D. Riddle.)

make the cellular responses all-or-none, so that one result is obtained when the signal is below a certain critical strength, and another result when it is above that strength. In many cases, however, responses are more finely graded: a high concentration may, for example, direct target cells into one developmental pathway, an intermediate concentration into another, and a low concentration into yet another. An important case is that in which the signal molecule diffuses out from a localized signaling center, creating a signal concentration gradient. Cells at different distances from the source are driven to behave in a variety of different ways, according to the signal concentration that they experience.

A signal molecule that imposes a pattern on a whole field of cells in this way is called a **morphogen**. Vertebrate limbs provide a striking example: a group of cells at one side of the embryonic limb bud become specialized as a signaling center and secrete Sonic hedgehog protein—a member of the Hedgehog family of signal molecules. This protein spreads out from its source, forming a *morphogen gradient* that controls the characters of the cells along the thumb-to-little-finger axis of the limb bud. If an additional group of signaling cells is grafted into the opposite side of the bud, a mirror duplication of the pattern of digits is produced (**Figure 22–13**).

Extracellular Inhibitors of Signal Molecules Shape the Response to the Inducer

Especially for molecules that can act at a distance, it is important to limit the action of the signal, as well as to produce it. Most developmental signal proteins have extracellular antagonists that can inhibit their function. These antagonists are generally proteins that bind to the signal or its receptor, preventing a productive interaction from taking place.

A surprisingly large number of developmental decisions are actually regulated by inhibitors rather than by the primary signal molecule. The nervous system in a

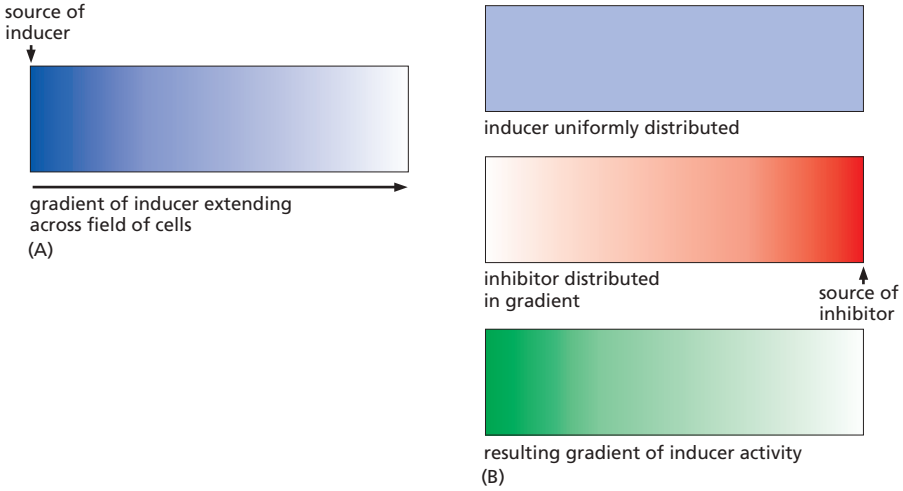


Figure 22–14 Two ways to create a morphogen gradient. (A) By localized production of an inducer—a morphogen—that diffuses away from its source. (B) By localized production of an inhibitor that diffuses away from its source and blocks the action of a uniformly distributed inducer.

frog embryo arises from a field of cells that is competent to form either neural or epidermal tissue. An inducing tissue releases the protein chordin, which favors the formation of neural tissue. Chordin does not have its own receptor. Instead it is an inhibitor of signal proteins of the BMP/TGFβ family, which induce epidermal development and are present throughout the neuroepithelial region where neurons and epidermis form. The induction of neural tissue is thus due to an inhibitory gradient of an antagonistic signal (**Figure 22–14**).

Developmental Signals Can Spread Through Tissue in Several Different Ways

Many developmental signals are thought to spread through tissues by simple diffusion through the spaces between cells. If some specialized group of cells produces a signal molecule at a steady rate, and this morphogen is then degraded as it diffuses away from this source, a smooth gradient will be set up, with its maximum at the source. The speed of diffusion and the half-life of the morphogen will together determine the steepness of the gradient (**Figure 22–15**).

This simple mechanism can be modified in many ways to adjust the shape and steepness of the gradient. Receptors on the surfaces of cells along the way may trap the diffusing morphogen and cause it to be endocytosed and degraded, shortening its effective halflife. Or it may bind to molecules in the extracellular matrix, reducing its effective diffusion rate. In some cases, it seems

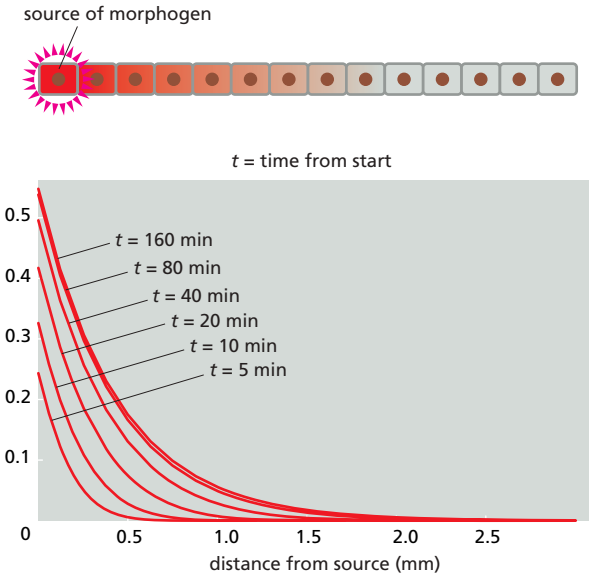


Figure 22–15 Setting up a signal gradient by diffusion. The graphs show successive stages in the build-up of the concentration of a signal molecule that is produced at a steady rate at the origin, with production starting at time 0. The molecule undergoes degradation as it diffuses away from the source, creating a concentration gradient with its peak at the source. The graphs are calculated on the assumption that diffusion is occurring along one axis in space, that the molecule has a half-life $t_{1/2}$ of 20 minutes, and that it diffuses with a diffusion constant $D = 0.4\ mm^2\ hr^{-1}$, typical of a small (30 kilodalton) protein molecule in water. Note that the gradient is already close to its steady-state form within an hour, and that the concentration at steady state (large times) falls off exponentially with distance.

that a morphogen is taken up into cells by endocytosis and then disgorged, only to be taken up and then disgorged by other cells in turn, so that the signal spreads through a largely intracellular route.

Yet another mechanism for signal distribution depends on long thin filopodia or *cytonemes* that extend over several cell diameters from cells in some epithelial tissues. A cell may send out cytonemes to make contact with distant cells, either to deliver or to receive signals from them. In this way, for example, a cell can deliver lateral inhibition via the Notch pathway to an extended set of neighbors.

Programs That Are Intrinsic to a Cell Often Define the Time-Course of its Development

Signals such as those we have just discussed play a large part in controlling the timing of events in development, but it would be wrong to imagine that every developmental change needs an inductive signal to trigger it. Many of the mechanisms that alter cell character are intrinsic to the cell and require no cue from the cell's surroundings: the cell will step through its developmental program even when kept in a constant environment. There are numerous cases where one might suspect that something of this sort is occurring to control the duration of a developmental process. For example, in a mouse, the neural progenitor cells in the cerebral cortex of the brain carry on dividing and generating neurons for just 11 cell cycles, and in a monkey for approximately 28 cycles, after which they stop. Different kinds of neurons are generated at different stages in this program, suggesting that as the progenitor cell ages, it changes the specifications that it supplies to the differentiating progeny cells.

It is difficult to prove in the context of the intact embryo that such a course of events is strictly the result of a cell-autonomous timekeeping process, since the cell environment is changing. Experiments on cells in culture, however, give clear-cut evidence. For example, glial progenitor cells isolated from the optic nerve of a 7-day postnatal rat and cultured under constant conditions in an appropriate medium will carry on proliferating for a strictly limited time (corresponding to a maximum of about eight cell division cycles) and then differentiate into oligodendrocytes (the glial cells that form myelin sheaths around axons in the brain), obeying a timetable similar to the one that they would have followed if they had been left in place in the embryo.

The molecular mechanisms underlying such slow changes in the internal states of cells, played out over days, weeks, months or even years, are still unknown. One possibility is that they reflect progressive changes in the state of the chromatin (discussed in Chapter 4).

The mechanisms that control the timing of more rapid processes, though still poorly understood, are not quite such a mystery. Later, we shall discuss an example—the gene expression oscillator, known as the *segmentation clock*, that governs formation of the somites in vertebrate embryos—the rudiments of the series of vertebrae, ribs, and associated muscles.

Initial Patterns Are Established in Small Fields of Cells and Refined by Sequential Induction as the Embryo Grows

The signals that organize the spatial pattern of an embryo generally act over short distances and govern relatively simple choices. A morphogen, for example, typically acts over a distance of less than 1 mm—an effective range for diffusion (see Figure 22–15)—and directs choices between no more than a handful of developmental options for the cells on which it acts. But the organs that eventually develop are much larger and more complex than this.

The cell proliferation that follows the initial specification accounts for the size increase, while the refinement of the initial pattern is explained by a series of local inductions that embroider successive levels of detail on an initially simple sketch. As soon as two sorts of cells are present, one of them can produce a

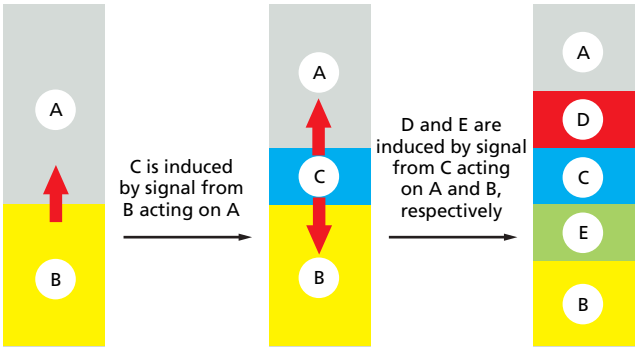


Figure 22–16 Patterning by sequential induction. A series of inductive interactions can generate many types of cells, starting from only a few.

factor that induces a subset of the neighboring cells to specialize in a third way. The third cell type can in turn signal back to the other two cell types nearby, generating a fourth and a fifth cell type, and so on (**Figure 22–16**).

This strategy for generating a progressively more complicated pattern is called **sequential induction**. It is chiefly through sequential inductions that the body plan of a developing animal, after being first roughed out in miniature, becomes elaborated with finer and finer details as development proceeds.

In the sections that follow, we focus on a small selection of model organisms to see how the principles that we have outlined in this first section operate in practice. We begin with the nematode worm, *Caenorhabditis elegans*.

Summary

The obvious changes of cell behavior that we see as a multicellular organism develops are the outward signs of a complex molecular computation, dependent on cell memory, that is taking place inside the cells as they receive and process signals from their neighbors and emit signals in return. The final pattern of differentiated cell types is thus the outcome of a more hidden program of cell specialization—a program played out in the changing patterns of expression of gene regulatory proteins, giving one cell different potentialities from another long before terminal differentiation begins. Developmental biologists seek to decipher the hidden program and to relate it, through genetic and microsurgical experiments, to the signals the cells exchange as they proliferate, interact, and move.

Animals as different as worms, flies, and humans use remarkably similar sets of proteins to control their development, so that what we discover in one organism very often gives insight into the others. A handful of evolutionarily conserved cell–cell signaling pathways are used repeatedly, in different organisms and at different times, to regulate the creation of an organized multicellular pattern. Differences of body plan seem to arise to a large extent from differences in the regulatory DNA associated with each gene. This DNA has a central role in defining the sequential program of development, calling genes into action at specific times and places according to the pattern of gene expression that was present in each cell at the previous developmental stage.

Differences between cells in an embryo arise in various ways. Positive feedback can lead to broken symmetry, creating a radical and permanent difference between cells that are initially almost identical. Sister cells can be born different as a result of an asymmetric cell division. Or a group of initially similar cells may receive different exposures to inductive signals from cells outside the group; long-range inducers with graded effects, called morphogens, can organize a complex pattern. Through cell memory, such transient signals can have a lasting effect on the internal state of a cell, causing it, for example, to become determined for a specific fate. In these ways, sequences of simple signals acting at different times and places in growing cell arrays give rise to the intricate and varied multicellular organisms that fill the world around us.

CAENORHABDITIS ELEGANS: DEVELOPMENT FROM THE PERSPECTIVE OF THE INDIVIDUAL CELL

The nematode worm *Caenorhabditis elegans* is a small, relatively simple, and precisely structured organism. The anatomy of its development has been described in extraordinary detail, and one can map out the exact lineage of every cell in the body. Its complete genome sequence is also known, and large numbers of mutant phenotypes have been analyzed to determine gene functions. If there is any multicellular animal whose development we should be able to understand in terms of genetic control, this is it.

DNA sequence comparisons indicate that, while the lineages leading to nematodes, insects, and vertebrates diverged from one another at about the same time, the rate of evolutionary change in the nematode lineage has been substantially greater: its genes, its body structure, and its developmental strategies are more divergent from our own than are those of *Drosophila.* Nevertheless, at a molecular level many of its developmental mechanisms are similar to those of insects or vertebrates, and governed by homologous systems of genes. If one wants to know how an eye, a limb, or a heart develops, one must look elsewhere: *C. elegans* lacks these organs. But at a more fundamental level, it is highly instructive: it poses the basic general questions of animal development in a relatively simple form, and it lets us answer them in terms of gene functions and the behavior of individual, identified cells.

Caenorhabditis elegans Is Anatomically Simple

As an adult, *C. elegans* consists of only about 1000 somatic cells and 1000–2000 germ cells (exactly 959 somatic cell nuclei plus about 2000 germ cells in one sex; exactly 1031 somatic cell nuclei plus about 1000 germ cells in the other) (**Figure 22–17**). The anatomy has been reconstructed, cell by cell, by electron microscopy of serial sections. The body plan of the worm is simple: it has a roughly bilaterally symmetrical, elongate body composed of the same basic tissues as in other animals (nerve, muscle, gut, skin), organized with mouth and brain at the anterior end and anus at the posterior. The outer body wall is composed of two layers: the protective epidermis, or "skin," and the underlying muscular layer. A tube of endodermal cells forms the intestine. A second tube, located between the intestine and the body wall, constitutes the gonad; its wall is composed of somatic cells, with the germ cells inside it.

C. elegans has two sexes—a hermaphrodite and a male. The hermaphrodite can be viewed most simply as a female that produces a limited number of sperm: she can reproduce either by self-fertilization, using her own sperm, or by cross-fertilization after transfer of male sperm by mating. Self-fertilization allows a single heterozygous worm to produce homozygous progeny. This is an important feature that helps to make *C. elegans* an exceptionally convenient organism for genetic studies.

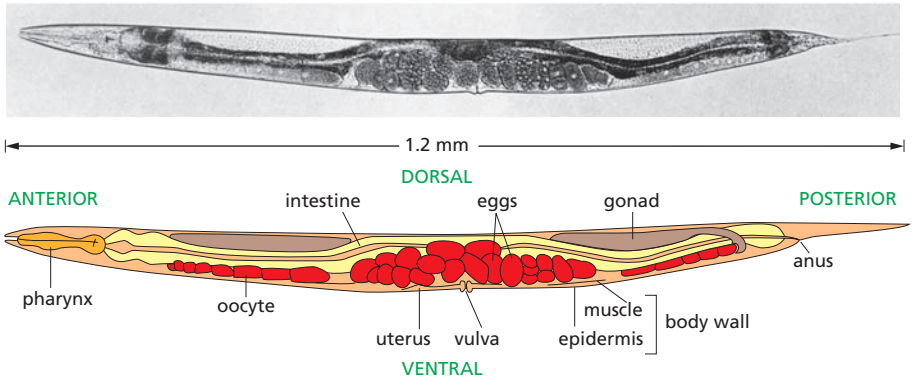
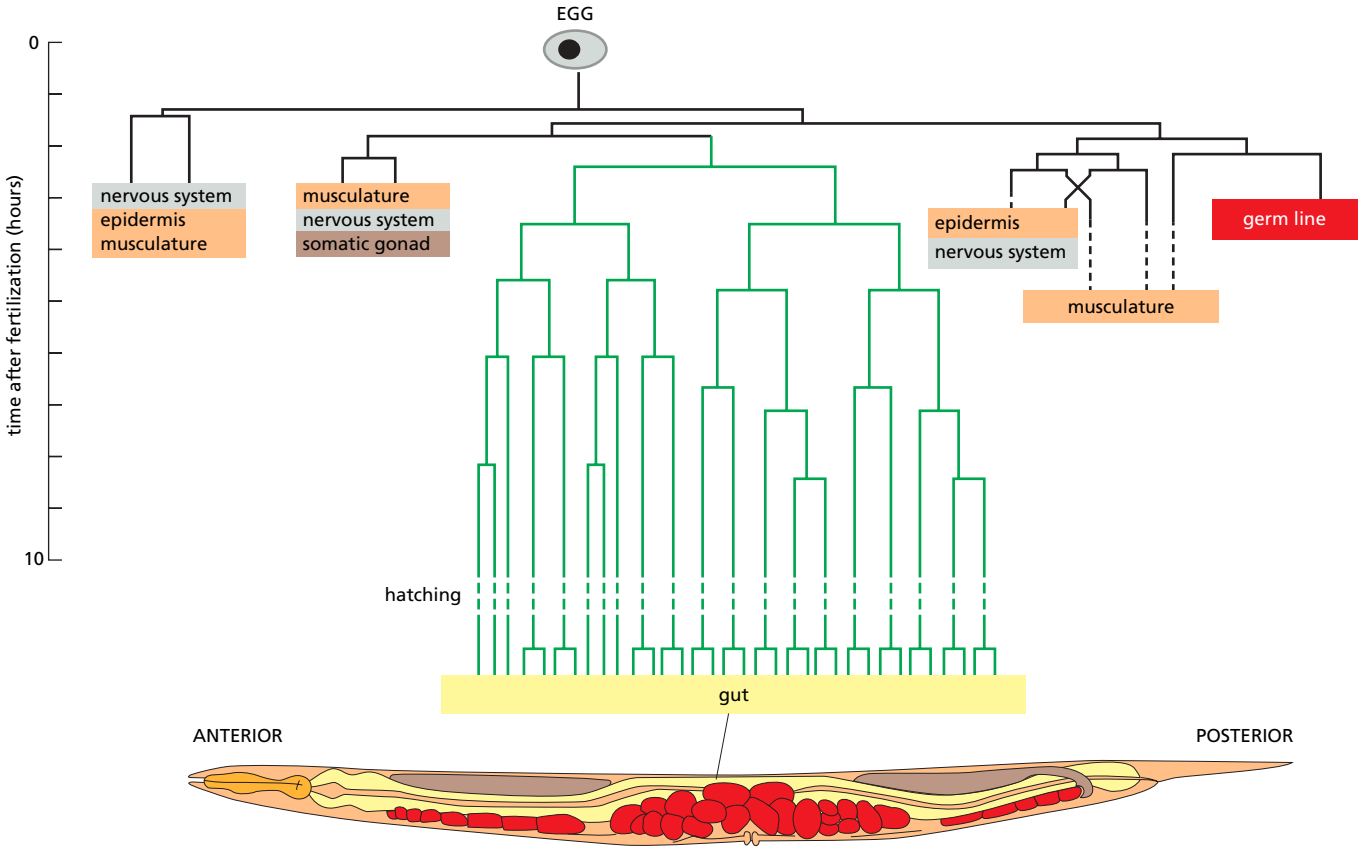


Figure 22–17 ***Caenorhabditis elegans.*** A side view of an adult hermaphrodite is shown. (From J.E. Sulston and H.R. Horvitz, *Dev. Biol.* 56:110–156, 1977. With permission from Academic Press.)



Cell Fates in the Developing Nematode Are Almost Perfectly Predictable

C. elegans begins life as a single cell, the fertilized egg, which gives rise, through repeated cell divisions, to 558 cells that form a small worm inside the egg shell. After hatching, further divisions result in the growth and sexual maturation of the worm as it passes through four successive larval stages separated by molts. After the final molt to the adult stage, the hermaphrodite worm begins to produce its own eggs. The entire developmental sequence, from egg to egg, takes only about three days.

The lineage of all of the cells from the single-cell egg to the multicellular adult was mapped out by direct observation of the developing animal. In the nematode, a given precursor cell follows the same pattern of cell divisions in every individual, and with very few exceptions the fate of each descendant cell can be predicted from its position in the lineage tree (**Figure 22–18**).

This degree of stereotyped precision is not seen in the development of larger animals. At first sight, it might seem to suggest that each cell lineage in the nematode embryo is rigidly and independently programmed to follow a set pattern of cell division and cell specialization, making the worm a woefully unrepresentative model organism for development. We shall see that this is far from true: as in other animals, development depends on cell–cell interactions as well as on processes internal to the individual cells. The outcome in the nematode is almost perfectly predictable simply because the pattern of cell–cell interactions is highly reproducible and is accurately correlated with the sequence of cell divisions.

In the developing worm, as in other animals, most cells do not become restricted to generate progeny cells of a single differentiated type until quite late in development, and cells of a particular type, such as muscle, usually derive from several spatially dispersed precursors that also give rise to other types of cells. The exceptions, in the worm, are the gut and the gonad, each of which forms from a single dedicated *founder cell*, born at the 8-cell stage of development for the gut-cell lineage and at the 16-cell stage for the germ-cell lineage, or

Figure 22–18 The lineage tree for the cells that form the gut (the intestine) of *C. elegans*. Note that although the intestinal cells form a single clone (as do the germ-line cells), the cells of most other tissues do not. Nerve cells (not shown in the drawing of the adult at the bottom) are mainly clustered in ganglia near the anterior and posterior ends of the animal and in a ventral nerve cord that runs the length of the body.

germ line. But in any case, cell diversification starts early, as soon as the egg begins to cleave: long before terminal differentiation, the cells begin to step through a series of intermediate states of specialization, following different programs according to their locations and their interactions with their neighbors. How do these early differences between cells arise?

Products of Maternal-Effect Genes Organize the Asymmetric Division of the Egg

The worm is typical of most animals in the early specification of the cells that will eventually give rise to the germ cells (eggs or sperm). The worm's germ line is produced by a strict series of asymmetric cell divisions of the fertilized egg. The asymmetry originates with a cue from the egg's environment: the sperm entry point defines the future posterior pole of the elongated egg. The proteins in the egg then interact with one another and organize themselves in relation to this point so as to create a more elaborate asymmetry in the interior of the cell. The proteins involved are mainly translated from the accumulated mRNA products of the genes of the mother. Because this RNA is made before the egg is laid, it is only the mother's genotype that dictates what happens in the first steps of development. Genes acting in this way are called **maternal-effect genes.**

A subset of maternal-effect genes are specifically required to organize the asymmetric pattern of the nematode egg. These are called *Par* (*Partitioning-defective*) genes, and at least six have been identified, through genetic screens for mutants where this pattern is disrupted. The *Par* genes have homologs in insects and vertebrates, where they play a fundamental part in the organization of cell polarity, as discussed in Chapter 19. In fact, one of the keys to our present understanding of the general mechanisms of cell polarity was the discovery of these genes through studies of early development in *C. elegans*.

In the nematode egg, as in other cells both in the nematode and other animals, the Par proteins (the products of the *Par* genes) are themselves asymmetrically located, some at one end of the cell and some at the other. They serve in the egg to bring a set of ribonucleoprotein particles called *P granules* to the posterior pole, so that the posterior daughter cell inherits P granules and the anterior daughter cell does not. Throughout the next few cell divisions, the Par proteins operate in a similar way, orienting the mitotic spindle and segregating the P granules to one daughter cell at each mitosis, until, at the 16-cell stage, there is just one cell that contains the P granules (**Figure 22–19**). This one cell gives rise to the germ line.

The specification of the germ-cell precursors as distinct from somatic-cell precursors is a key event in the development of practically every type of animal, and the process has common features even in phyla with very different body plans. Thus, in *Drosophila*, particles similar to P granules are also segregated

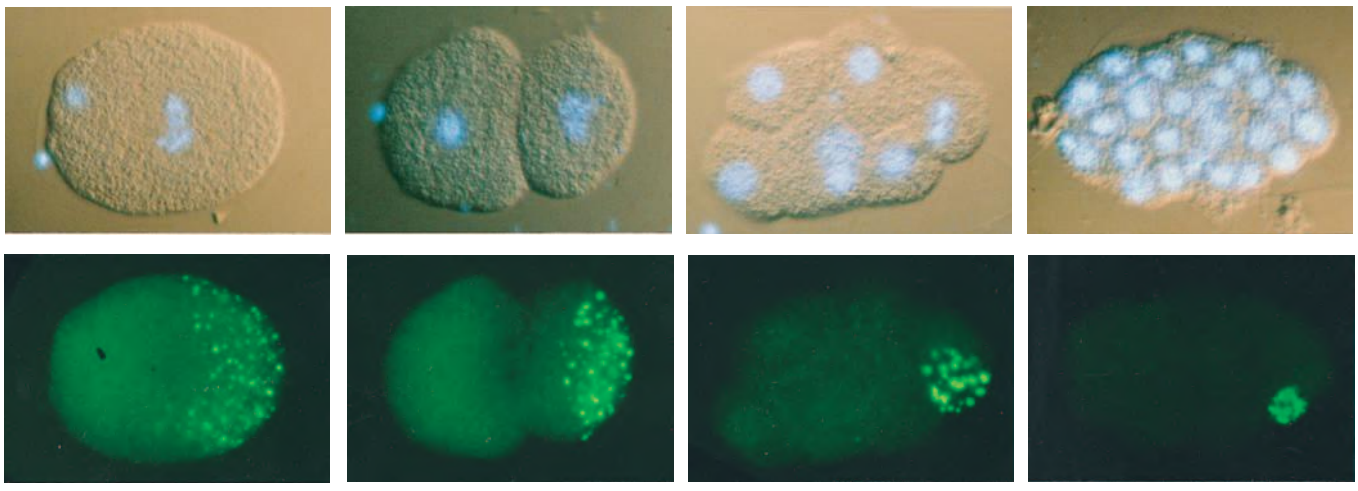


Figure 22–19 Asymmetric divisions segregating P granules into the founder cell of the *C. elegans* germ line. The micrographs in the upper row show the pattern of cell divisions, with cell nuclei stained *blue* with a DNA-specific fluorescent dye; below are the same cells stained with an antibody against P granules. These small granules (0.5–1 μm in diameter) are distributed randomly throughout the cytoplasm in the unfertilized egg (not shown). After fertilization, at each cell division up to the 16-cell stage, both they and the intracellular machinery that regulates their asymmetric localization are segregated into a single daughter cell. (Courtesy of Susan Strome.)

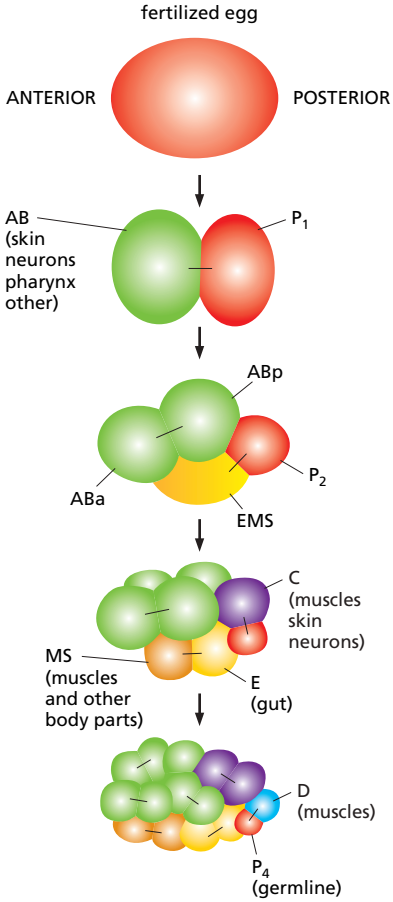


Figure 22–20 The pattern of cell divisions in the early *C. elegans* embryo, indicating the names and fates of the individual cells. Cells that are sisters are shown linked by a *short black line*. (After K. Kemphues, *Cell* 101:345–348, 2000. With permission from Elsevier.)

into one end of the egg, and become incorporated into the germ-line precursor cells to determine their fate. Similar phenomena occur in fish and frogs. In all these species, one can recognize at least some of the same proteins in the germ-cell-determining material, including homologs of an RNA-binding protein called Vasa. How Vasa and its associated proteins and RNA molecules act to define the germ line is still unknown.

Progressively More Complex Patterns Are Created by Cell–Cell Interactions

The egg, in *C. elegans* as in other animals, is an unusually big cell, with room for complex internal patterning. In addition to the P granules, other factors become distributed in an orderly way along its anteroposterior axis under the control of the Par proteins, and thus are allocated to different cells as the egg goes through its first few cell-division cycles. These divisions occur without growth (since feeding cannot begin until a mouth and a gut have formed) and therefore subdivide the egg into progressively smaller cells. Several of the localized factors are gene regulatory proteins, which act directly in the cell that inherits them to either drive or block the expression of specific genes, adding to the differences between that cell and its neighbors and committing it to a specialized fate.

While the first few differences between cells along the anteroposterior axis of *C. elegans* result from asymmetric divisions, further patterning, including the pattern of cell types along the other axes, depends on interactions between one cell and another. The cell lineages in the embryo are so reproducible that individual cells can be assigned names and identified in every animal (**Figure 22–20**); the cells at the four-cell stage, for example, are called ABa and ABp (the two anterior sister cells), and EMS and P_2 (the two posterior sister cells). As a result of the asymmetric divisions we have just described, the P_2 cell expresses a signal protein on its surface—a nematode homolog of the Notch ligand Delta—while the ABa and ABp cells express the corresponding transmembrane receptor—a homolog of Notch. The elongated shape of the eggshell forces these cells into an arrangement such that the most anterior cell, ABa, and the most posterior cell, P_2, are no longer in contact with one another. Thus only the ABp cell receives the signal from P_2, making ABp different from ABa and defining the future dorsal–ventral axis of the worm (**Figure 22–21**).

At the same time, P_2 also expresses another signal molecule, a Wnt protein, which acts on a Wnt receptor (a Frizzled protein) in the membrane of the EMS cell. This signal polarizes the EMS cell in relation to its site of contact with P_2, controlling the orientation of the mitotic spindle. The EMS cell then divides to give two daughters that become committed to different fates as a result of the Wnt signal from P_2. One daughter, the MS cell, will give rise to muscles and various other body parts; the other daughter, the E cell, is the founder cell for the gut, committed to give rise to all the cells of the gut and to no other tissues (see Figure 22–21).

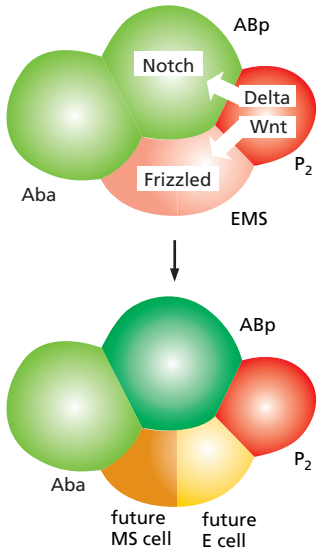


Figure 22–21 Cell signaling pathways controlling assignment of different characters to the cells in a four-cell nematode embryo. The P_2 cell uses the Notch signaling pathway to send an inductive signal to the ABp cell, causing this to adopt a specialized character. The ABa cell has all the molecular apparatus to respond in the same way to the same signal, but it does not do so because it is out of contact with P_2. Meanwhile, a Wnt signal from the P_2 cell causes the EMS cell to orient its mitotic spindle and generate two daughters that become committed to different fates as a result of their different exposure to Wnt protein—the MS cell and the E cell (the founder cell of the gut).

Having sketched the chain of cause and effect in early nematode development, we now examine some of the methods that have been used to decipher it.

Microsurgery and Genetics Reveal the Logic of Developmental Control; Gene Cloning and Sequencing Reveal Its Molecular Mechanisms

To discover the causal mechanisms, we need to know the developmental potential of the individual cells in the embryo. At what points in their lives do they undergo decisive internal changes that determine them for a particular fate, and at what points do they depend on signals from other cells? In the nematode, using laser microbeam microsurgery, one can accurately kill one or more of a cell's neighbors and then observe directly how the cell behaves in the altered circumstances. Alternatively, cells of the early embryo can be pushed around and rearranged inside the eggshell using a fine needle. For example, the relative positions of ABa and ABp can be flipped at the four-cell stage of development. The ABa cell then undergoes what would normally be the fate of the ABp cell, and vice versa, showing that the two cells initially have the same developmental potential and depend on signals from their neighbors to make them different. A third tactic is to remove the eggshell of an early *C. elegans* embryo by digesting it with enzymes, and then to manipulate the cells in culture. The existence of a polarizing signal from P_2 to EMS was demonstrated in this way.

Genetic screens were used to identify the genes involved in the P_2–EMS cell interaction. A search was made for mutant strains of worms in which no gut cells were induced (called *Mom* mutants, because they had *mo*re *m*esoderm—mesoderm being the fate of both of the EMS cell daughters when induction fails). Cloning and sequencing the *Mom* genes revealed that one encodes a Wnt signal protein that is expressed in the P_2 cell, while another encodes a Frizzled protein (a Wnt receptor) that is expressed in the EMS cell. A second genetic screen was conducted for mutant strains of worms with the opposite phenotype, in which extra gut cells are induced (called *Pop* mutants, for *po*sterior *p*harynx defect). One of the *Pop* genes (*Pop1*) turns out to encode a gene regulatory protein (a LEF1/TCF homolog) whose activity is down-regulated by Wnt signaling in *C. elegans*. When Pop1 activity is absent, both daughters of the EMS cell behave as though they have received the Wnt signal from P_2. Similar genetic methods were used to identify the genes whose products mediate the Notch-dependent signaling from P_2 to ABa.

Continuing in this way, it is possible to build up a detailed picture of the decisive events in nematode development, and of the genetically specified machinery that controls them.

Cells Change Over Time in Their Responsiveness to Developmental Signals

The complexity of the adult nematode body is achieved through repeated use of a handful of patterning mechanisms, including those we have just seen in action in the early embryo. For example, cell divisions with a molecular asymmetry dependent on the Pop1 gene regulatory proteins occur throughout *C. elegans* development, creating anterior and posterior sister cells with different characters.

As emphasized earlier, while the same few types of signals act repeatedly at different times and places, the effects they have are different because the cells are programmed to respond differently according to their age and their past history. We have seen, for example, that at the four-cell stage of development, one cell, ABp, changes its developmental potential because of a signal received via the Notch pathway. At the 12-cell stage of development, the granddaughters of the ABp cell and the granddaughters of the ABa cell both encounter another Notch signal, this time from a daughter of the EMS cell. The ABa granddaughter

changes its internal state in response to this signal and begins to form the pharynx. The ABp granddaughter does no such thing—the earlier exposure to a Notch signal has made it unresponsive. Thus, at different times in their history, both ABa lineage cells and ABp lineage cells respond to Notch, but the outcomes are different. Somehow a Notch signal at the 12-cell stage induces pharynx, but a Notch signal at the 4-cell stage has other effects—which include the prevention of pharynx induction by Notch at a later stage. This phenomenon, in which the same signaling mechanism evokes different effects at different stages and in different contexts—is seen in the development of all animals, and in all of them Notch signaling is used repeatedly in this way.

Heterochronic Genes Control the Timing of Development

A cell does not have to receive an external cue in order to change: one set of regulatory molecules inside the cell can provoke the production of another, and the cell can thus step through a series of different states through its own internal mechanisms. These states differ not only in their responsiveness to external signals, but also in other aspects of their internal chemistry, including proteins that stop or start the cell-division cycle. In this way, the internal mechanisms of the cell, together with the past and present signals received, dictate both the sequence of biochemical changes in the cell and the timing of its cell divisions.

The specific molecular details of the mechanisms governing the temporal program of development are still mysterious. Remarkably little is known, even in the nematode embryo with its rigidly predictable pattern of cell divisions, about how the sequence of cell divisions is controlled. However, for the later stages, when the larva feeds and grows and moults to become an adult, it has been possible to identify some of the genes that control the timing of cellular events. Mutations in these genes cause *heterochronic* phenotypes: the cells in a larva of one stage behave as though they belonged to a larva of a different stage, or cells in the adult carry on dividing as though they belonged to a larva (**Figure 22–22**).

Through genetic analyses, one can determine that the products of the heterochronic genes act in series, forming regulatory cascades. Curiously, two genes at the top of their respective cascades, called *Lin4* and *Let7*, do not code for proteins but for microRNAs—short untranslated regulatory RNA molecules, 21 or 22 nucleotides long. These act by binding to complementary sequences in the noncoding regions of mRNA molecules transcribed from other heterochronic genes, thereby inhibiting their translation and promoting their degradation, as discussed in Chapter 7. Increasing levels of *Lin4* RNA govern the progression from larval stage-1 cell behavior to larval stage-3 cell behavior; increasing levels

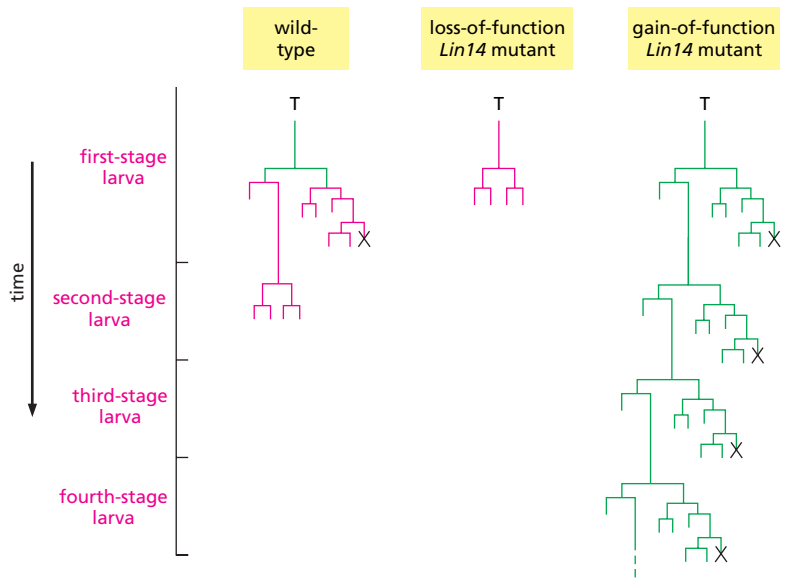


Figure 22–22 Heterochronic mutations in the *Lin14* gene of *C. elegans*. The effects on only one of the many affected lineages are shown. The loss-of-function (recessive) mutation in *Lin14* causes premature occurrence of the pattern of cell division and differentiation characteristic of a late larva, so that the animal reaches its final state prematurely and with an abnormally small number of cells. The gain-of-function (dominant) mutation has the opposite effect, causing cells to reiterate the patterns of cell divisions characteristic of the first larval stage, continuing through as many as five or six molt cycles and persisting in the manufacture of an immature type of cuticle. The cross denotes a programmed cell death. *Green lines* represent cells that contain Lin14 protein (which binds to DNA), *red lines* those that do not. In normal development the disappearance of Lin14 is triggered by the beginning of larval feeding. (After V. Ambros and H.R. Horvitz, *Science* 226:409–416, 1984, with permission from AAAS; and P. Arasu, B. Wightman and G. Ruvkun, *Growth Dev. Aging* 5:1825–1833, 1991, with permission from Growth Publishing Co., Inc.)

of *Let7* RNA govern the progression from late larva to adult. In fact, *Lin4* and *Let7* were the first microRNAs to be described in any animal: it was through developmental genetic studies in *C. elegans* that the importance of this whole class of molecules for gene regulation in animals was discovered.

RNA molecules that are identical or almost identical to the *Let7* RNA are found in many other species, including *Drosophila*, zebrafish, and human. Moreover, these RNAs appear to act in a similar way to regulate the level of their target mRNA molecules, and the targets themselves are homologous to the targets of *Let7* RNA in the nematode. In *Drosophila*, this system of molecules seems to be involved in the metamorphosis of the larva into a fly, hinting at a conserved role in governing the timing of developmental transitions.

Cells Do Not Count Cell Divisions in Timing Their Internal Programs

Since the steps of cell specialization have to be coordinated with cell divisions, it is often suggested that the cell division cycle might serve as a clock to control the tempo of other events in development. In this view, changes of internal state would be locked to passage through each division cycle: the cell would click to the next state as it went through mitosis, so to speak. Although there are indeed some cases where changes of cell state are conditional on cell cycle events, this is far from being the general rule. Cells in developing embryos, whether they be worms, flies, or vertebrates, usually carry on with their standard timetable of determination and differentiation even when progress through the cell-division cycle is artificially blocked. Necessarily, there are some abnormalities, if only because a single undivided cell cannot differentiate in two ways at once. But in most cases that have been studied, it seems that the cell changes its state with time more or less regardless of cell division, and that this changing state controls both the decision to divide and the decision as to when and how to specialize.

Selected Cells Die by Apoptosis as Part of the Program of Development

The control of cell numbers in development depends on cell death as well as cell division. A *C. elegans* hermaphrodite generates 1030 somatic cell nuclei in the course of its development, but 131 of the cells die. These programmed cell deaths occur in an absolutely predictable pattern. In *C. elegans*, they can be chronicled in detail, because one can trace the fate of each individual cell and see which dies, watching as each suicide victim undergoes apoptosis and is rapidly engulfed and digested by neighboring cells (**Figure 22–23**). In other organisms, where close observation is harder, such deaths easily go unnoticed; but cell death by apoptosis is probably the fate of a substantial fraction of the cells produced in most animals, playing an essential part in generating an individual with the right cell types in the right numbers and places, as discussed in Chapter 18.

Genetic screens in *C. elegans* have been crucial in identifying the genes that bring about apoptosis and in highlighting its importance in development. Three genes, called *Ced3*, *Ced4*, and *Egl1* (*Ced* stands for *cell death abnormal*), are found to be required for the 131 normal cell deaths to occur. If these genes are inactivated by mutation, cells that are normally fated to die survive instead, differentiating as recognizable cell types such as neurons. Conversely, over-expression or misplaced expression of the same genes causes many cells to die that

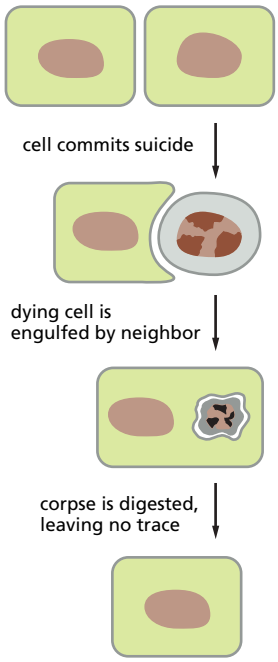


Figure 22–23 Apoptotic cell death in *C. elegans*. Death depends on expression of the *Ced3* and *Ced4* genes in the absence of *Ced9* expression—all in the dying cell itself. The subsequent engulfment and disposal of the remains depend on expression of other genes in the neighboring cells.

would normally survive, and the same effect results from mutations that inactivate another gene, *Ced9*, which normally represses the death program.

All these genes code for conserved components of the cell-death machinery. As described in Chapter 18, *Ced3* codes for a caspase homolog, while *Ced4*, *Ced9*, and *Egl1* are respectively homologs of *Apaf1*, *Bcl2*, and *Bad*. Without the insights that came from detailed analysis of the development of the transparent, genetically tractable nematode worm, it would have been very much harder to discover these genes and understand the cell-death process in vertebrates.

Summary

The development of the small, relatively simple, transparent nematode worm Caenorhabditis elegans *is extraordinarily reproducible and has been chronicled in detail, so that a cell at any given position in the body has the same lineage in every individual, and this lineage is fully known. Also, the genome has been completely sequenced. Thus, powerful genetic and microsurgical approaches can be combined to decipher developmental mechanisms. As in other organisms, development depends on an interplay of cell–cell interactions and cell-autonomous processes. Development begins with an asymmetric division of the fertilized egg, dividing it into two smaller cells containing different cell-fate determinants. The daughters of these cells interact via the Notch and Wnt cell signaling pathways to create a more diverse array of cell states. Meanwhile, through further asymmetric divisions one cell inherits materials from the egg that determine it at an early stage as progenitor of the germ line.*

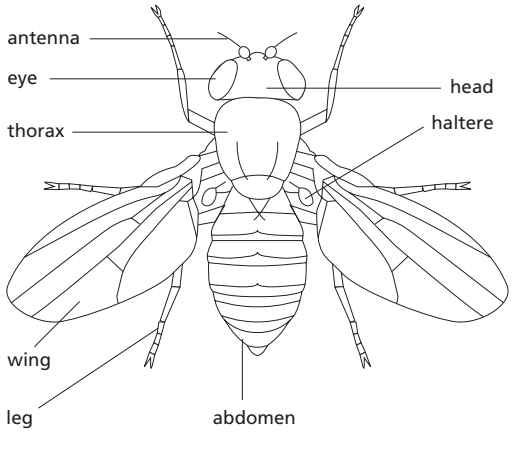
Genetic screens identify the sets of genes responsible for these and later steps in development, including, for example, cell-death genes that control the apoptosis of a specific subset of cells as part of the normal developmental program. Heterochronic genes that govern the timing of developmental events have also been found, although in general our understanding of temporal control of development is still very poor. There is good evidence, however, that the tempo of development is not set by the counting of cell divisions.

DROSOPHILA AND THE MOLECULAR GENETICS OF PATTERN FORMATION: GENESIS OF THE BODY PLAN

It is the fly *Drosophila melanogaster* (**Figure 22–24**), more than any other organism, that has transformed our understanding of how genes govern the patterning of the body. The anatomy of *Drosophila* is more complex than that of *C. elegans*,



(A)



(B)

Figure 22–24 *Drosophila melanogaster.* Dorsal view of a normal adult fly. (A) Photograph. (B) Labeled drawing. (Photograph courtesy of E.B. Lewis.)

with more than 100 times as many cells, and it shows more obvious parallels with our own body structure. Surprisingly, the fly has fewer genes than the worm—about 14,000 as compared with 20,000. On the other hand, it has almost twice as much DNA per gene (about 10,000 nucleotides on average, as compared with about 5000), most of this being noncoding sequence. The molecular construction kit has fewer types of parts, but the assembly instructions—as specified by the regulatory sequences in the non-coding DNA—seem to be more voluminous.

Decades of genetic study, culminating in massive systematic genetic screens, have yielded a catalogue of the developmental control genes that define the spatial pattern of cell types and body structures of the fly, and molecular biology has given us the tools to watch these genes in action. By *in situ* hybridization using DNA or RNA probes on whole embryos, or by staining with labeled antibodies to reveal the distribution of specific proteins, one can observe directly how the internal states of the cells are defined by the sets of regulatory genes that they express at different times of development. Moreover, by analyzing animals that are a patchwork of mutant and nonmutant cells, one can discover how each gene operates as part of a system to specify the organization of the body.

Most of the genes controlling the pattern of the body in *Drosophila* turn out to have close counterparts in higher animals, including ourselves. In fact, many of the basic devices for defining the body plan and patterning individual organs and tissues are astonishingly similar. Thus, quite surprisingly, the fly has provided the key to understanding the molecular genetics of our own development.

Flies, like nematode worms, are ideal for genetic studies: cheap to breed, easy to mutagenize, and rapid in their reproductive cycle. But there is a more fundamental reason why they have been so important for developmental geneticists. As emphasized earlier, as a result of gene duplications, vertebrate genomes often contain two or three homologous genes corresponding to a single gene in the fly. A mutation that disrupts one of these genes very often fails to reveal the gene's core function, because the other homologs share this function and remain active. In the fly, with its more economical gene set, this phenomenon of genetic redundancy is less prevalent. The phenotype of a single mutation in the fly therefore more often directly uncovers the function of the mutant gene.

The Insect Body Is Constructed as a Series of Segmental Units

The timetable of *Drosophila* development, from egg to adult, is summarized in **Figure 22–25**. The period of *embryonic development* begins at fertilization and takes about a day, at the end of which the embryo hatches out of the egg shell to become a *larva.* The larva then passes through three stages, or *instars,* separated by molts in which it sheds its old coat of cuticle and lays down a larger one. At the end of the third instar it pupates. Inside the *pupa,* a radical remodeling of the body takes place—a process called *metamorphosis.* Eventually, about nine days after fertilization, an adult fly, or *imago,* emerges.

The fly consists of a head, with mouth, eyes, and antennae, followed by three thoracic segments (numbered T1 to T3), and eight or nine abdominal segments (numbered A1 to A9). Each segment, although different from the others, is built according to a similar plan. Segment T1, for example, carries a pair of legs, T2 carries a pair of legs plus a pair of wings, and T3 carries a pair of legs plus a pair of halteres—small knob-shaped balancers important in flight, evolved from the second pair of wings that more primitive insects possess. The quasi-repetitive segmentation develops in the early embryo during the first few hours after fertilization (**Figure 22–26**), but it is more obvious in the larva (**Figure 22–27**), where the segments look more similar than in the adult. In the embryo it can be seen that the rudiments of the head, or at least the future adult mouth parts, are likewise segmental. At the two ends of the animal, however, there are highly specialized terminal structures that are not segmentally derived.

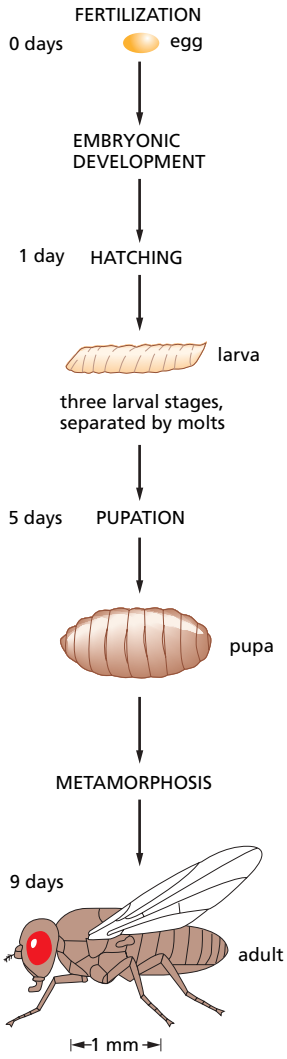


Figure 22–25 Synopsis of *Drosophila* development from egg to adult fly.

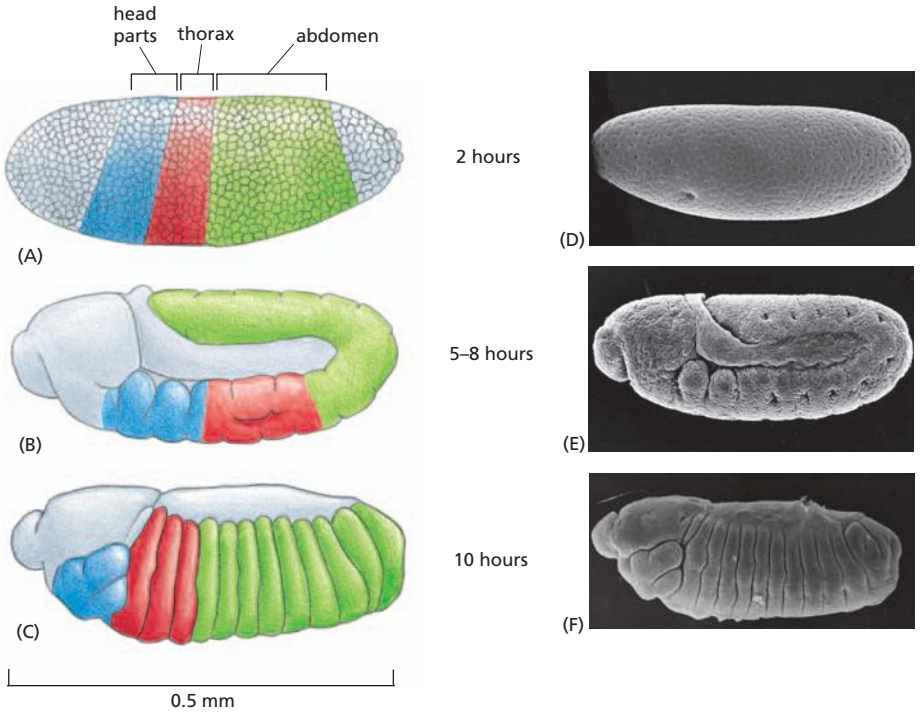


Figure 22–26 The origins of the *Drosophila* body segments during embryonic development. <AACG> The embryos are seen in side view in drawings (A–C) and corresponding scanning electron micrographs (D–F). (A and D) At 2 hours the embryo is at the syncytial blastoderm stage (see Figure 22–28) and no segmentation is visible, although a fate map can be drawn showing the future segmented regions (color in A). (B and E) At 5–8 hours the embryo is at the extended germ band stage: gastrulation has occurred, segmentation has begun to be visible, and the segmented axis of the body has lengthened, curving back on itself at the tail end so as to fit into the egg shell. (C and F) At 10 hours the body axis has contracted and become straight again, and all the segments are clearly defined. The head structures, visible externally at this stage, will subsequently become tucked into the interior of the larva, to emerge again only when the larva goes through pupation to become an adult. (D and E, courtesy of F.R. Turner and A.P. Mahowald, *Dev. Biol.* 50:95–108, 1976; F, from J.P. Petschek, N. Perrimon and A.P. Mahowald, *Dev. Biol.* 119:175–189, 1987. All with permission from Academic Press.)

The boundaries between segments are traditionally defined by visible anatomical markers; but in discussing gene expression patterns it is often convenient to draw a different set of segmental boundaries, defining a series of segmental units called *parasegments*, half a segment out of register with the traditional divisions (see Figure 22–27).

Drosophila Begins Its Development as a Syncytium

The egg of *Drosophila* is about 0.5 mm long and 0.15 mm in diameter, with a clearly defined polarity. Like the eggs of other insects, but unlike most vertebrates, it begins its development in an unusual way: a series of nuclear divisions without cell division creates a syncytium. The early nuclear divisions are synchronous and extremely rapid, occurring about every 8 minutes. The first nine divisions generate a cloud of nuclei, most of which migrate from the middle of the egg toward the surface, where they form a monolayer called the *syncytial blastoderm*. After another four rounds of nuclear division, plasma membranes

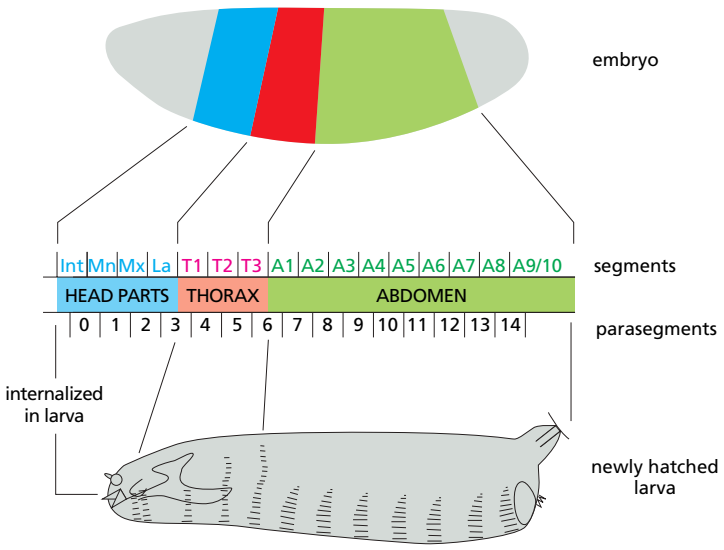
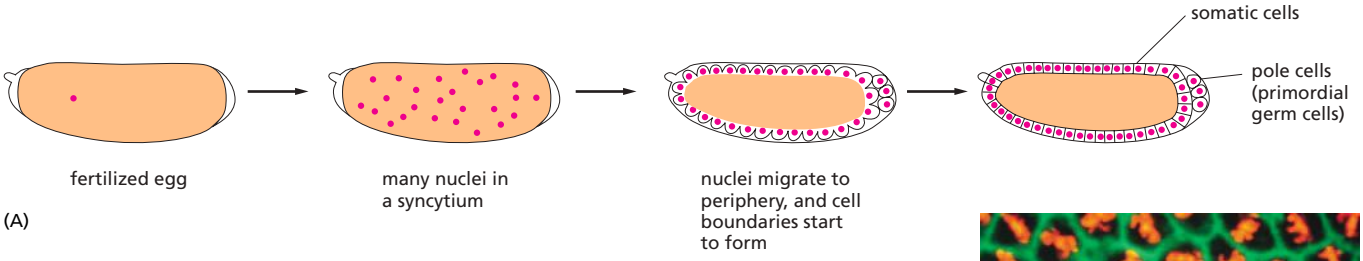


Figure 22–27 The segments of the *Drosophila* larva and their correspondence with regions of the blastoderm. The parts of the embryo that become organized into segments are shown in *color*. The two ends of the embryo, shaded *gray*, are not segmented and become tucked into the interior of the body to form the internal structures of the head and gut. (The future external, segmental structures of the adult head are also transiently tucked into the interior in the larva.) Segmentation in *Drosophila* can be described in terms of either segments or parasegments: the relationship is shown in the middle part of the figure. Parasegments often correspond more simply to patterns of gene expression. The exact number of abdominal segments is debatable: eight are clearly defined, and a ninth is present vestigially in the larva, but absent in the adult.



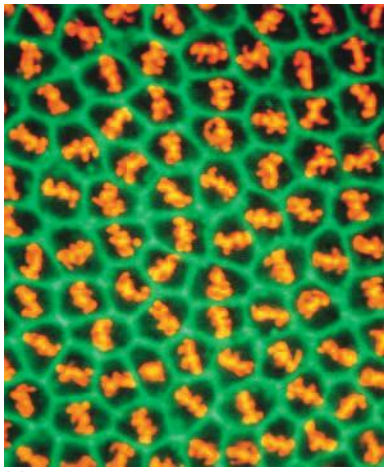


Figure 22–28 Development of the *Drosophila* egg from fertilization to the cellular blastoderm stage. (A) Schematic drawings. (B) Surface view—an optical-section photograph of blastoderm nuclei undergoing mitosis at the transition from the syncytial to the cellular blastoderm stage. Actin is stained *green,* chromosomes *orange.* (A, after H.A. Schneiderman, in Insect Development [P.A. Lawrence, ed.], pp. 3–34. Oxford, UK: Blackwell, 1976; B, courtesy of William Sullivan.)

grow inward from the egg surface to enclose each nucleus, thereby converting the syncytial blastoderm into a *cellular blastoderm* consisting of about 6000 separate cells (**Figure 22–28**). About 15 of the nuclei populating the extreme posterior end of the egg are segregated into cells a few cycles earlier; these *pole cells* are the germ-line precursors (primordial germ cells) that will give rise to eggs or sperm.

Up to the cellular blastoderm stage, development depends largely—although not exclusively—on stocks of maternal mRNA and protein that accumulated in the egg before fertilization. The frantic rate of DNA replication and nuclear division evidently gives little opportunity for transcription. After cellularization, cell division continues in a more conventional way, asynchronously and at a slower rate, and the rate of transcription increases dramatically. Gastrulation begins a little while before cellularization is complete, when parts of the sheet of cells forming the exterior of the embryo start to tuck into the interior to form the gut, the musculature, and associated internal tissues. A little later and in another region of the embryo, a separate set of cells move from the surface epithelium into the interior to form the central nervous system. By marking and following the cells through these various movements, one can draw a fate map for the monolayer of cells on the surface of the blastoderm (**Figure 22–29**).

As gastrulation nears completion, a series of indentations and bulges appear in the surface of the embryo, marking the subdivision of the body into segments along its anteroposterior axis (see Figure 22–26). Soon a fully segmented larva emerges, ready to start eating and growing. Within the body of the larva, small groups of cells remain apparently undifferentiated, forming structures called *imaginal discs.* These will grow as the larva grows, and eventually they will give rise to most of the structures of the adult body, as we shall see later.

A head end and a tail end, a ventral (belly) side and a dorsal (back) side, a gut, a nervous system, a series of body segments—these are all features of the

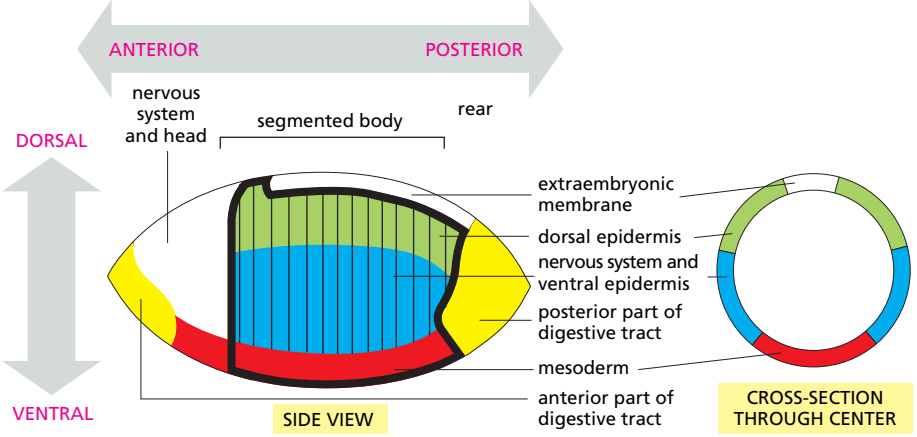


Figure 22–29 Fate map of a *Drosophila* embryo at the cellular blastoderm stage. The embryo is shown in side view and in cross section, displaying the relationship between the dorsoventral subdivision into future major tissue types and the anteroposterior pattern of future segments. A heavy line encloses the region that will form segmental structures. During gastrulation the cells along the ventral midline invaginate to form mesoderm, while the cells fated to form the gut invaginate near each end of the embryo. (After V. Hartenstein, G.M. Technau and J.A. Campos-Ortega, *Wilhelm Roux' Arch. Dev. Biol.* 194:213–216, 1985. With permission from Elsevier.)

basic body plan that *Drosophila* shares with many other animals, including ourselves. We begin our account of the mechanisms of *Drosophila* development by considering how this body plan is set up.

Genetic Screens Define Groups of Genes Required for Specific Aspects of Early Patterning

By carrying out a series of genetic screens based on saturation mutagenesis (discussed in Chapter 8), it has been possible to amass a collection of *Drosophila* mutants that appears to include changes in a large proportion of the genes affecting development. Independent mutations in the same gene can be distinguished from mutations in separate genes by a complementation test (see Panel 8–1, p. 555), leading to a catalog of genes classified according to their mutant phenotypes. In such a catalog, a group of genes with very similar mutant phenotypes will often code for a set of proteins that work together to perform a particular function.

Sometimes the developmental functions revealed by mutant phenotypes are those that one would expect; sometimes they are a surprise. A large-scale genetic screen focusing on early *Drosophila* development revealed that the key genes fall into a relatively small set of functional classes defined by their mutant phenotypes. Some—the *egg-polarity genes* (**Figure 22–30**)—are required to

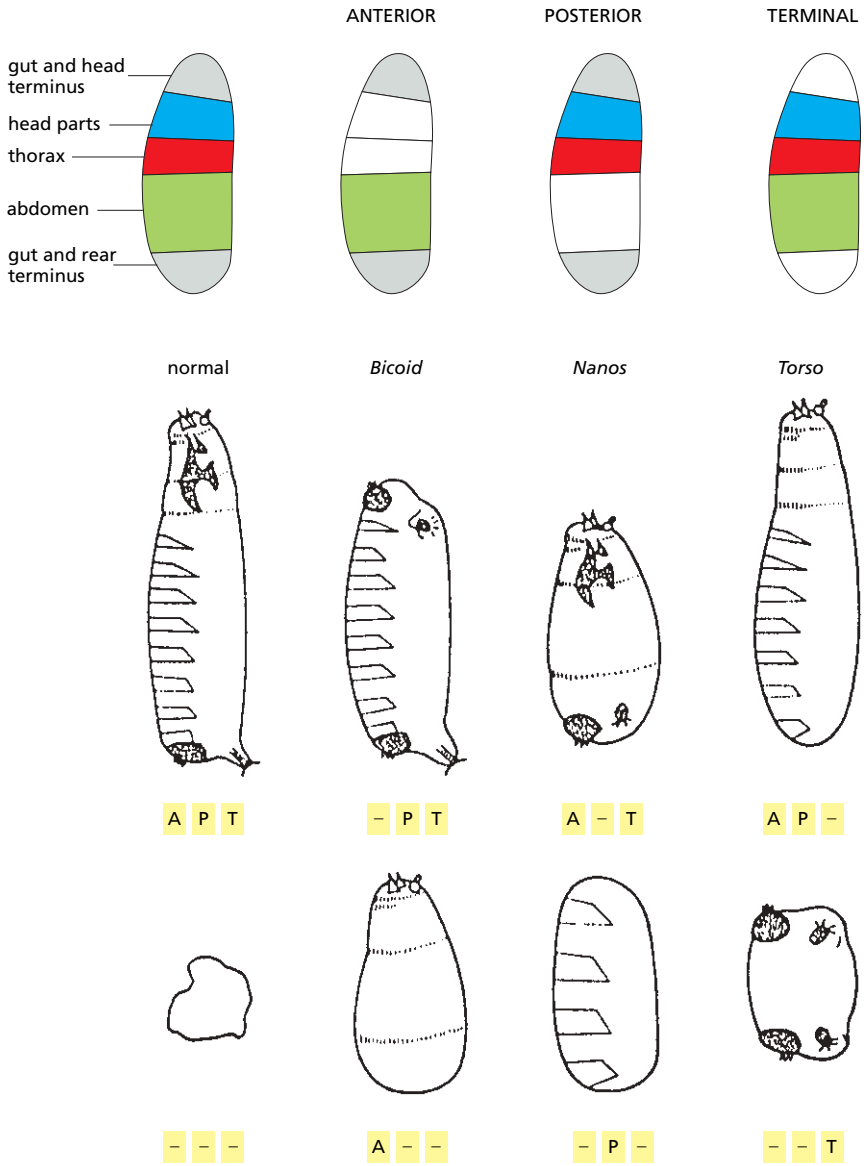


Figure 22–30 The domains of the anterior, posterior, and terminal systems of egg-polarity genes. The upper diagrams show the fates of the different regions of the egg/early embryo and indicate (in *white*) the parts that fail to develop if the anterior, posterior, or terminal system is defective. The middle row shows schematically the appearance of a normal larva and of mutant larvae that are defective in a gene of the anterior system (for example, *Bicoid*), of the posterior system (for example, *Nanos*), or of the terminal system (for example, *Torso*). The bottom row of drawings shows the appearances of larvae in which none or only one of the three gene systems is functional. The lettering beneath each larva specifies which systems are intact (A P T for a normal larva, – P T for a larva where the anterior system is defective but the posterior and terminal systems are intact, and so on). Inactivation of a particular gene system causes loss of the corresponding set of body structures; the body parts that form correspond to the gene systems that remain functional. Note that larvae with a defect in the anterior system can still form terminal structures at their anterior end, but these are of the type normally found at the rear end of the body rather than the front of the head. (Slightly modified from D. St. Johnston and C. Nüsslein-Volhard, *Cell* 68:201–219, 1992. With permission from Elsevier.)

define the anteroposterior and dorsoventral axes of the embryo and mark out its two ends for special fates, by mechanisms involving interactions between the oocyte and surrounding cells in the ovary. Others, the *gap genes*, are required in specific broad regions along the anteroposterior axis of the early embryo to allow their proper development. A third category, the *pair-rule genes*, are required, more surprisingly, for development of alternate body segments. A fourth category, the *segment polarity genes*, are responsible for organizing the anteroposterior pattern of each individual segment.

The discovery of these four systems of genes and the subsequent analysis of their functions (an enterprise that still continues) was a famous tour-de-force of developmental genetics. It had a revolutionary impact on all of developmental biology by showing the way toward a systematic, comprehensive account of the genetic control of embryonic development. In this section, we shall summarize only briefly the conclusions relating to the earliest phases of *Drosophila* development, because these are insect-specific; we dwell at greater length on the parts of the process that illustrate more general principles.

Interactions of the Oocyte With Its Surroundings Define the Axes of the Embryo: the Role of the Egg-Polarity Genes

Surprisingly, the earliest steps of animal development are among the most variable, even within a phylum. A frog, a chicken, and a mammal, for example, even though they develop in similar ways later, make eggs that differ radically in size and structure, and they begin their development with different sequences of cell divisions and cell specialization events.

The style of early development that we have described for *C. elegans* is typical of many classes of animals. In contrast, the early development of *Drosophila* represents a rather extreme variant. The main axes of the future insect body are defined before fertilization by a complex exchange of signals between the unfertilized egg, or oocyte, and the follicle cells that surround it in the ovary (**Figure 22–31**). Then, in the syncytial phase following fertilization, an exceptional amount of patterning occurs in the array of rapidly dividing nuclei, before the first partitioning of the egg into separate cells. Here, there is no need for the usual forms of cell–cell communication involving transmembrane signaling; neighboring regions of the early *Drosophila* embryo can communicate by means of gene regulatory proteins and mRNA molecules that diffuse or are actively transported through the cytoplasm of the giant multinuclear cell.

In the stages before fertilization, the anteroposterior axis of the future embryo becomes defined by three systems of molecules that create landmarks in the oocyte (**Figure 22–32**). Following fertilization, each landmark serves as a beacon, providing a signal, in the form of a morphogen gradient, that organizes the developmental process in its neighborhood. Two of these signals are generated from localized deposits of specific mRNA molecules. The future anterior end of the embryo contains a high concentration of mRNA for a gene regulatory protein called Bicoid; this mRNA is translated to produce Bicoid protein, which

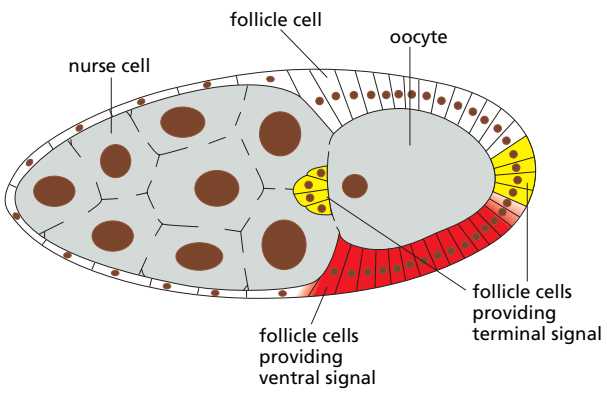


Figure 22–31 A *Drosophila* oocyte in its follicle. The oocyte is derived from a germ cell that divides four times to give a family of 16 cells that remain in communication with one another via cytoplasmic bridges *(gray)*. One member of the family group becomes the oocyte, while the others become nurse cells, which make many of the components required by the oocyte and pass them into it via the cytoplasmic bridges. The follicle cells that partially surround the oocyte have a separate ancestry. As indicated, they are the sources of terminal and ventral egg-polarizing signals.

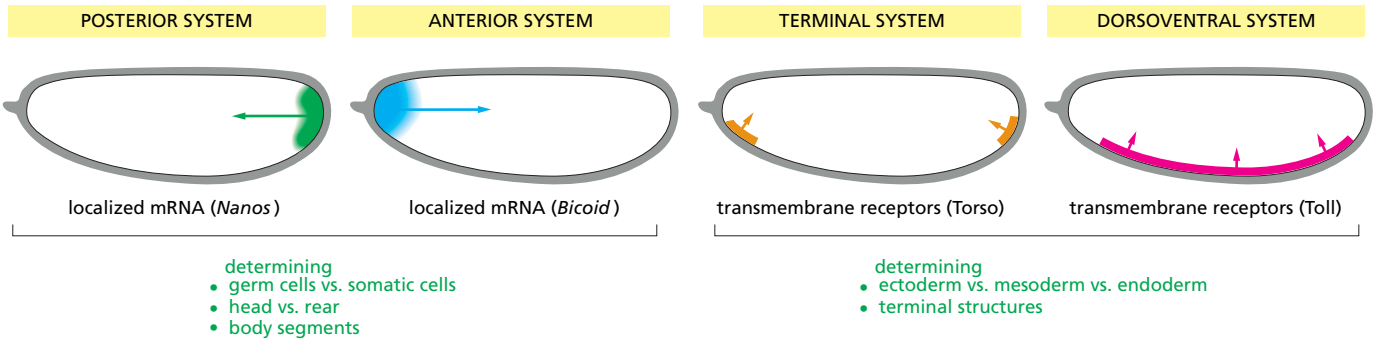


Figure 22–32 The organization of the four egg-polarity gradient systems. The receptors Toll and Torso are distributed all over the membrane; the coloring in the diagrams on the right indicates where they become activated by extracellular ligands.

diffuses away from its source to form a concentration gradient with its maximum at the anterior end of the egg. The future posterior end of the embryo contains a high concentration of mRNA for a regulator of translation called Nanos, which sets up a posterior gradient in the same way. The third signal is generated symmetrically at both ends of the egg, by local activation of a transmembrane tyrosine kinase receptor called Torso. The activated receptor exerts its effects over a shorter range, marking the sites of specialized terminal structures that will form at the head and tail ends of the future larva and also defining the rudiments of the future gut. The three sets of genes responsible for these localized determinants are referred to as the **anterior**, **posterior**, and **terminal** sets of **egg-polarity** genes.

A fourth landmark defines the dorsoventral axis (see Figure 22–32): a protein that is produced by follicle cells underneath the future ventral region of the embryo leads to localized activation of another transmembrane receptor, called Toll, in the oocyte membrane. The genes required for this function are called **dorsoventral** egg-polarity genes.

All the egg-polarity genes in these four classes are maternal-effect genes: it is the mother's genome, not the zygotic genome, that is critical. Thus, a fly whose chromosomes are mutant in both copies of the *Bicoid* gene but who is born from a mother carrying one normal copy of *Bicoid* develops perfectly normally, without any defects in the head pattern. However, if that daughter fly is a female no functional *Bicoid* mRNA can be deposited into the anterior part of her own eggs, and all of these will develop into headless embryos regardless of the father's genotype.

Each of the four egg-polarity signals—provided by Bicoid, Nanos, Torso, and Toll—exerts its effect by regulating (directly or indirectly) the expression of genes in the nuclei of the blastoderm. The use of these particular molecules to organize the egg is not a general feature of early animal development—indeed, only *Drosophila* and closely related insects possess a *Bicoid* gene. And Toll has been coopted here for dorsoventral patterning; its more ancient and universal function is in the innate immune response, as discussed in Chapter 24.

Nevertheless, the egg-polarity system shows some highly conserved features. For example, the localization of *Nanos* mRNA at one end of the egg is linked to, and dependent on, the localization of germ-cell determinants at that site, just as it is in *C. elegans*. Later in development, as the zygotic genome comes into play under the influence of the egg-polarity system, more similarities with other animal species become apparent. We shall use the dorsoventral system to illustrate this point.

The Dorsoventral Signaling Genes Create a Gradient of a Nuclear Gene Regulatory Protein

Localized activation of the Toll receptor on the ventral side of the egg controls the distribution of Dorsal, a gene regulatory protein inside the egg. The Dorsal protein belongs to the same family as the NFκB gene regulatory protein of vertebrates (discussed in Chapter 15). Its Toll-regulated activity, like that of NFκB,

depends on its translocation from the cytoplasm, where it is held in an inactive form, to the nucleus, where it regulates gene expression. In the newly laid egg, both the *Dorsal* mRNA (detected by *in situ* hybridization) and the protein it encodes (detected with antibodies) are distributed uniformly in the cytoplasm. After the nuclei have migrated to the surface of the embryo to form the blastoderm, however, a remarkable redistribution of the Dorsal protein occurs: dorsally the protein remains in the cytoplasm, but ventrally it is concentrated in the nuclei, with a smooth gradient of nuclear localization between these two extremes (**Figure 22–33**). The signal transmitted by the Toll protein controls this redistribution of Dorsal through a signaling pathway that is essentially the same as the Toll-dependent pathway involved in innate immunity.

Once inside the nucleus, the Dorsal protein turns on or off the expression of different sets of genes depending on its concentration. The expression of each responding gene depends on its regulatory DNA—specifically, on the number and affinity of the binding sites that this DNA contains for Dorsal and other regulatory proteins. In this way, the regulatory DNA can be said to *interpret* the positional signal provided by the Dorsal protein gradient, so as to define a dorsoventral series of territories—distinctive bands of cells that run the length of the embryo (**Figure 22–34**A). Most ventrally—where the concentration of Dorsal protein is highest—it switches on, for example, the expression of a gene called *Twist* that is specific for mesoderm (**Figure 22–35**). Most dorsally, where the concentration of Dorsal protein is lowest, the cells switch on *Decapentaplegic (Dpp)*. And in an intermediate region, where the concentration of Dorsal protein is high enough to repress *Dpp* but too low to activate *Twist*, the cells switch on another set of genes, including one called *Short gastrulation (Sog)*.

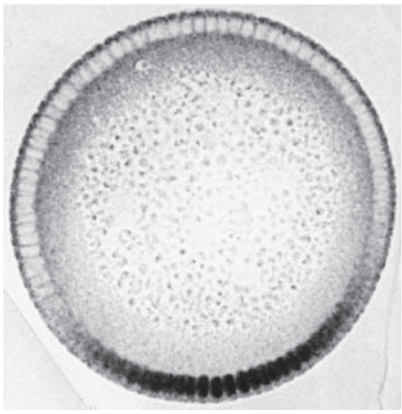


Figure 22–33 The concentration gradient of Dorsal protein in the nuclei of the blastoderm, as revealed by an antibody. Dorsally, the protein is present in the cytoplasm and absent from the nuclei; ventrally, it is depleted in the cytoplasm and concentrated in the nuclei. (From S. Roth, D. Stein and C. Nüsslein-Volhard, *Cell* 59:1189–1202, 1989. With permission from Elsevier.)

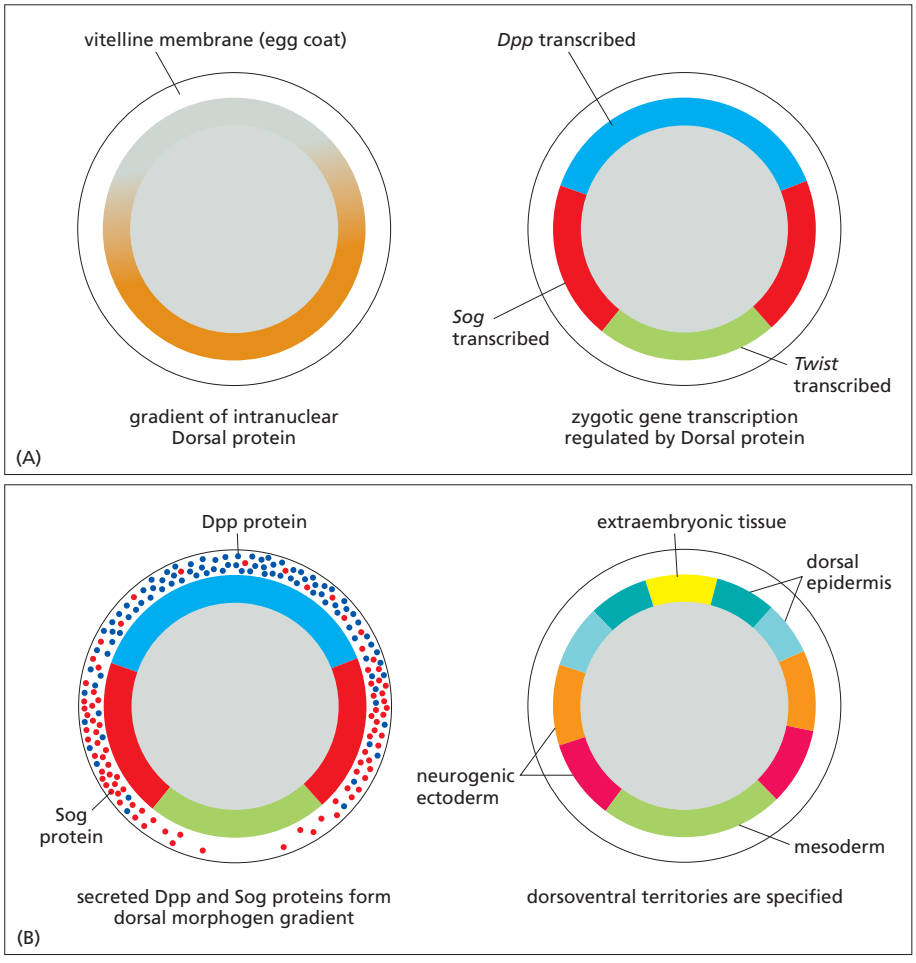


Figure 22–34 Morphogen gradients patterning the dorsoventral axis of the embryo. (A) The gradient of Dorsal protein defines three broad territories of gene expression, marked here by the expression of three representative genes—*Dpp, Sog,* and *Twist.* (B) Slightly later, the cells expressing *Dpp* and *Sog* secrete, respectively, the signal proteins Dpp (a TGFβ family member) and Sog (an antagonist of Dpp). These two proteins diffuse and interact with one another (and with certain other factors) to set up a gradient of Dpp activity that guides a more detailed patterning process.



Figure 22–35 Origin of the mesoderm from cells expressing Twist. Embryos were fixed at successive stages, cross-sectioned, and stained with an antibody against the Twist protein, a gene regulatory protein of the bHLH family. The cells that express Twist move into the interior of the embryo to form mesoderm. (From M. Leptin, J. Casal, B. Grunewald and R. Reuter, *Development Suppl.* 23–31, 1992. With permission from The Company of Biologists.)

Dpp and Sog Set Up a Secondary Morphogen Gradient to Refine the Pattern of the Dorsal Part of the Embryo

Products of the genes directly regulated by the Dorsal protein generate in turn more local signals that define finer subdivisions of the dorsoventral axis. These signals act after cellularization, and take the form of conventional extracellular signal molecules. In particular, *Dpp* codes for the secreted Dpp protein, which forms a local morphogen gradient in the dorsal part of the embryo. The gene *Sog*, meanwhile, codes for another secreted protein that is produced in the neurogenic ectoderm and acts as an antagonist of Dpp. The opposing diffusion gradients of these two proteins create a steep gradient of Dpp activity. The highest Dpp activity levels, in combination with certain other factors, cause development of the most dorsal tissue of all—extraembryonic membrane; intermediate levels cause development of dorsal ectoderm; and very low levels allow development of neurogenic ectoderm (Figure 22–34B).

The Insect Dorsoventral Axis Corresponds to the Vertebrate Ventrodorsal Axis

Dpp is a member of the TGFβ superfamily of signal molecules that is also important in vertebrates; Sog is a homolog of the vertebrate protein chordin. It is striking that a Dpp homolog, BMP4, and chordin work together in vertebrates in the same way as do Dpp and Sog in *Drosophila*. These two proteins control the dorsoventral pattern of the ectoderm, with high levels of chordin defining the region that is neurogenic and high levels of BMP4 activity defining the region that is not. This, combined with other molecular parallels, strongly suggests that this part of the body plan has been conserved between insects and vertebrates. However, the axis is inverted, so that dorsal in the fly corresponds to ventral in the vertebrate (**Figure 22–36**). At some point in its evolutionary history, it seems, the ancestor of one of these classes of animals took to living life upside down.

Three Classes of Segmentation Genes Refine the Anterior–Posterior Maternal Pattern and Subdivide the Embryo

After the initial gradients of Bicoid and Nanos are created to define the anteroposterior axis, the **segmentation genes** refine the pattern. Mutations in any one of the segmentation genes alter the number of segments or their basic internal organization without affecting the global polarity of the embryo. Segmentation genes are expressed by subsets of cells in the embryo, so their products are the

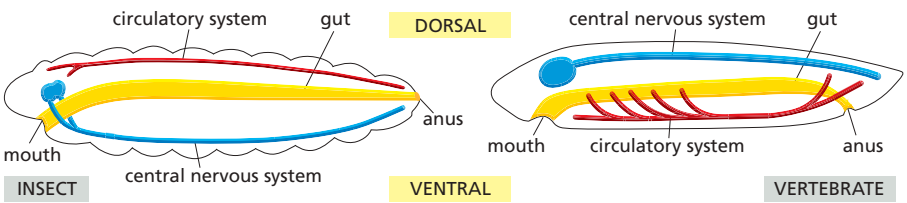


Figure 22–36 The vertebrate body plan as a dorsoventral inversion of the insect body plan. The mechanism of dorsoventral patterning in a vertebrate embryo is discussed in more detail later in this chapter. Note the correspondence with regard to the circulatory system as well as the gut and nervous system. In insects, the circulatory system is represented by a tubular heart and a main dorsal blood vessel, which pumps blood out into the tissue spaces through one set of apertures and receives blood back from the tissues through another set. In contrast with vertebrates, there is no system of capillary vessels to contain the blood as it percolates through the tissues. Nevertheless, heart development depends on homologous genes in vertebrates and insects, reinforcing the relationship between the two body plans. (After E.L. Ferguson, *Curr. Opin. Genet. Dev.* 6:424–431, 1996. With permission from Elsevier.)

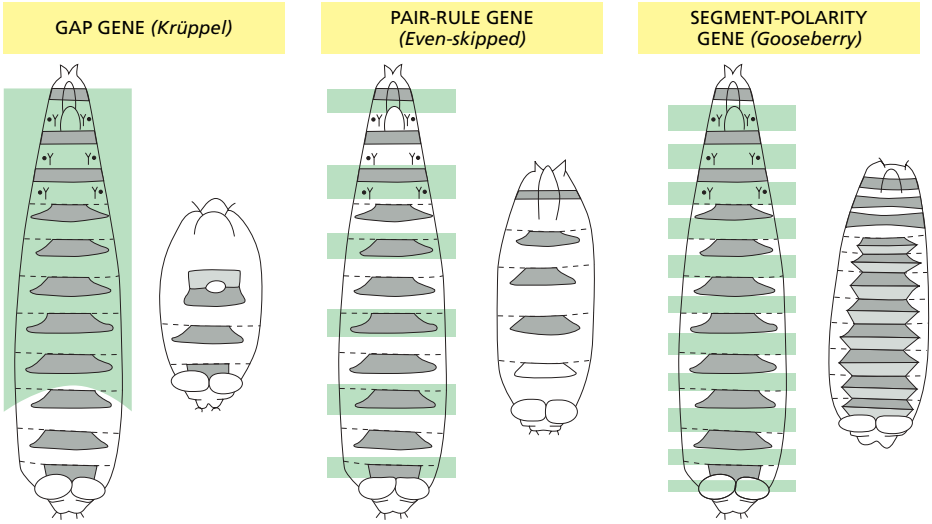


Figure 22–37 Examples of the phenotypes of mutations affecting the three types of segmentation genes. In each case the areas shaded in *green* on the normal larva *(left)* are deleted in the mutant or are replaced by mirror-image duplicates of the unaffected regions. (Modified from C. Nüsslein-Volhard and E. Wieschaus, *Nature* 287:795–801, 1980. With permission from Macmillan Publishers Ltd.)

first components that the embryo's own genome, rather than the maternal genome, contributes to embryonic development. They are therefore called *zygotic-effect genes* to distinguish them from the earlier maternal-effect genes.

The segmentation genes fall into three groups according to their mutant phenotypes (**Figure 22–37**). It is convenient to think of these three groups as acting in sequence, although in reality their functions overlap in time. First come a set of at least six **gap genes**, whose products mark out coarse subdivisions of the embryo. Mutations in a gap gene eliminate one or more groups of adjacent segments, and mutations in different gap genes cause different but partially overlapping defects. In the mutant *Krüppel,* for example, the larva lacks eight segments, from T1 to A5 inclusive.

The next category of segmentation genesis a set of eight **pair-rule genes**. Mutations in these cause a series of deletions affecting alternate segments, leaving the embryo with only half as many segments as usual. While all the pair-rule mutants display this two-segment periodicity, they differ in the precise positioning of the deletions relative to the segmental or parasegmental borders. The pair-rule mutant *Even-skipped (Eve),* for example, which is discussed in Chapter 7, lacks the whole of each odd-numbered parasegment, while the pair-rule mutant *Fushi tarazu (Ftz)* lacks the whole of each even-numbered parasegment, and the pair-rule mutant *Hairy* lacks a series of regions that are of similar width but out of register with the parasegmental units.

Finally, there are at least 10 **segment-polarity genes**. Mutations in these genes produce larvae with a normal number of segments but with a part of each segment deleted and replaced by a mirror-image duplicate of all or part of the rest of the segment. In *Gooseberry* mutants, for example, the posterior half of each segment (that is, the anterior half of each parasegment) is replaced by an approximate mirror image of the adjacent anterior half-segment (see Figure 22–37).

We see later that, in parallel with the segmentation process, a further set of genes, the *homeotic selector genes,* serve to define and preserve the differences between one segment and the next.

The phenotypes of the various segmentation mutants suggest that the segmentation genes form a coordinated system that subdivides the embryo progressively into smaller and smaller domains along the anteroposterior axis, distinguished by different patterns of gene expression. Molecular genetics has helped to reveal how this system works.

The Localized Expression of Segmentation Genes Is Regulated by a Hierarchy of Positional Signals

About three-quarters of the segmentation genes, including all of the gap genes and pair-rule genes, code for gene regulatory proteins. Their actions on one

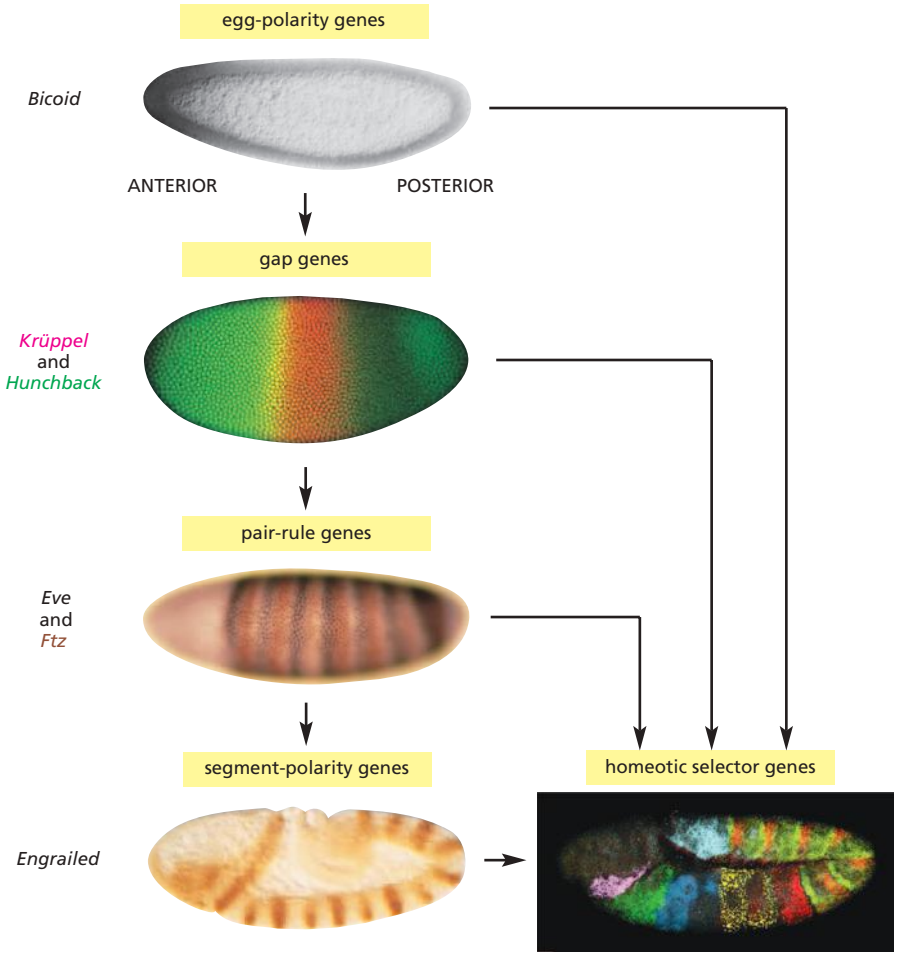


Figure 22–38 The regulatory hierarchy of egg-polarity, gap, segmentation, and homeotic selector genes. The photographs show expression patterns of representative examples of genes in each category, revealed by staining with antibodies against the protein products. The homeotic selector genes, discussed below, define the lasting differences between one segment and the next. (Photographs from top (i) from W. Driever and C. Nüsslein-Volhard, *Cell* 54:83–104, 1988. With permission from Elsevier; (ii) courtesy of Jim Langeland, Steve Paddock, Sean Carroll, and the Howard Hughes Medical Institute; (iii) from P.A. Lawrence, The Making of a Fly. Oxford, UK: Blackwell, 1992; (iv) from C. Hama, Z. Ali and T.B. Kornberg, *Genes Dev.* 4:1079–1093, 1990. With permission from Cold Spring Harbor Laboratory Press; (v) courtesy of William McGinnis, adapted from D. Kosman et al., *Science* 305:846, 2004. With permission from AAAS.)

another and on other genes can therefore be observed by comparing gene expression in normal and mutant embryos. By using appropriate probes to detect the gene transcripts or their protein products, one can, in effect, take snapshots as genes switch on and off in changing patterns. Repeating the process with mutants that lack a particular segmentation gene, one can begin to dissect the logic of the entire gene control system.

The products of the egg-polarity genes provide the global positional signals in the early embryo. These cause particular gap genes to be expressed in particular regions. The products of the gap genes then provide a second tier of positional signals that act more locally to regulate finer details of patterning through the expression of yet other genes, including the pair-rule genes (**Figure 22–38**). The pair-rule genes in turn collaborate with one another and with the gap genes to set up a regular periodic pattern of expression of segment-polarity genes, and the segment-polarity genes collaborate with one another to define the internal pattern of each individual segment. The strategy, therefore, is one of sequential induction (see Figure 22–16). By the end of the process, the global gradients produced by the egg-polarity genes have triggered the creation of a fine-grained pattern through a hierarchy of sequential, progressively more local, positional controls. Because the global positional signals that start the process do not have to directly specify fine details, the individual cell nuclei do not have to be governed with extreme precision by small differences in the concentration of these signals. Instead, at each step in the sequence, new signals come into play, providing substantial localized differences of concentration to define new details. Sequential induction is thus a robust strategy. It works reliably to produce fly embryos that all have the same pattern, despite the essential imprecision of biological control systems, and despite variations in conditions such as the temperature at which the fly develops.

The Modular Nature of Regulatory DNA Allows Genes to Have Multiple Independently Controlled Functions

The elaborate patterning process just described depends on the long stretches of noncoding DNA sequence that control the expression of each of the genes involved. These regulatory regions bind multiple copies of the gene regulatory proteins produced by the patterning genes expressed earlier. Like an input–output logic device, an individual gene is thus turned on and off according to the particular combination of proteins bound to its regulatory regions at each stage of development. In Chapter 7 we describe one particular segmentation gene—the pair-rule gene *Even-skipped (Eve)*—and discuss how the decision whether to transcribe the gene is made on the basis of all these inputs (see Figure 7–55). This example can be taken further to illustrate some important principles of developmental patterning.

Individual stripes of *Eve* expression depend on separate regulatory modules in the *Eve* regulatory DNA. Thus, one regulatory module is responsible for driving *Eve* expression in stripes 1 + 5, another for stripe 2, another for stripes 3 + 7 , and yet another for stripes 4 + 6 (**Figure 22–39**). Each regulatory module defines a different set of requirements for gene expression according to the concentrations of the products of the egg-polarity and gap genes. In this way, the *Eve* regulatory DNA serves to translate the complex nonrepetitive pattern of egg-polarity and gap proteins into the periodic pattern of expression of a pair-rule gene.

The modular organization of the *Eve* regulatory DNA just described is typical of gene regulation in multicellular animals and plants, and it has profound implications. By stringing together sequences of modules that respond to different combinations of regulatory proteins, it is possible to generate almost any pattern of gene expression on the basis of almost any other. Modularity, moreover, allows the regulatory DNA to define patterns of gene expression that are not merely complex, but whose parts are independently adjustable. A change in one of the regulatory modules can alter one part of the expression pattern without affecting the rest, and without requiring changes in regulatory proteins that would have repercussions for the expression of other genes in the genome. As described in Chapter 7, it is such regulatory DNA that contains the key to the complex organization of multicellular plants and animals, and its properties make possible the independent adaptability of each part of an organism's body structure in the course of evolution.

Most of the segmentation genes also have important functions at other times and places in the development of *Drosophila*. The *Eve* gene, for example, is expressed in subsets of neurons, in muscle precursor cells, and in various

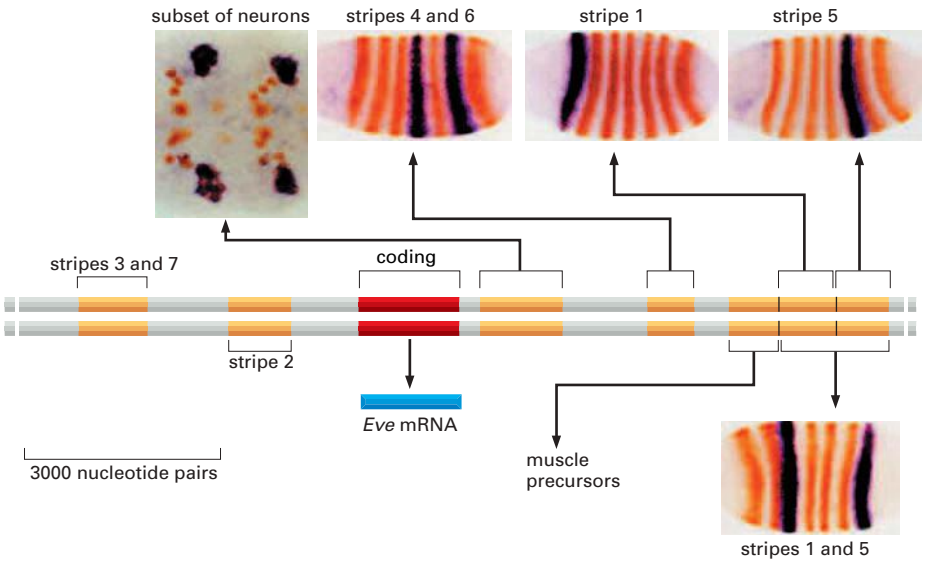


Figure 22–39 Modular organization of the regulatory DNA of the *Eve* gene. In the experiment shown, cloned fragments of the regulatory DNA were linked to a *LacZ* reporter (a bacterial gene). Transgenic embryos containing these constructs were then stained by *in situ* hybridization to reveal the pattern of expression of *LacZ (blue/black),* and counterstained with an anti-Eve antibody *(orange)* to show the positions of the normal *Eve* expression stripes. Different segments of the *Eve* regulatory DNA *(ochre)* are thus found to drive gene expression in regions corresponding to different parts of the normal *Eve* expression pattern. Two segments in tandem drive expression in a pattern that is the sum of the patterns generated by each of them individually. Separate regulatory modules are responsible for different times of gene expression, as well as different locations: the leftmost panel shows the action of a module that comes into play later than the others illustrated and drives expression in a subset of neurons. (From M. Fujioka et al., *Development* 126:2527–538, 1999. With permission from The Company of Biologists.)

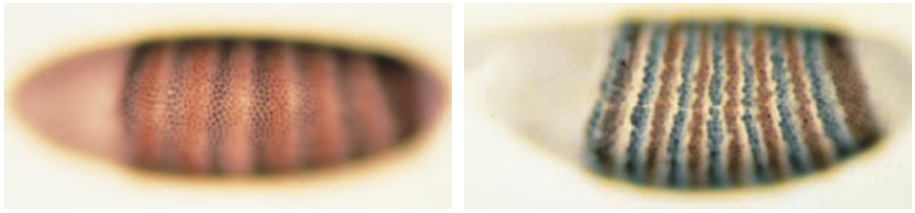


Figure 22–40 The formation of *Ftz* and *Eve* stripes in the *Drosophila* blastoderm. *Ftz* and *Eve* are both pair-rule genes. Their expression patterns (shown in *brown* for *Ftz* and in *gray* for *Eve*) are at first blurred but rapidly resolve into sharply defined stripes. (From P.A. Lawrence, The Making of a Fly. Oxford, UK: Blackwell, 1992.)

other sites, under the control of additional enhancers (see Figure 22–39). By addition of new modules to its regulatory DNA, any gene can be co-opted during evolution for new purposes at new sites in the body, without detriment to its other functions.

Egg-Polarity, Gap, and Pair-Rule Genes Create a Transient Pattern That Is Remembered by Other Genes

Within the first few hours after fertilization, the gap genes and the pair-rule genes are activated. Their mRNA products appear first in patterns that only approximate the final picture; then, within a short time—through a series of interactive adjustments—the fuzzy initial distribution of gene products resolves itself into a regular, crisply defined system of stripes (**Figure 22–40**). But this system itself is unstable and transient. As the embryo proceeds through gastrulation and beyond, the regular segmental pattern of gap and pair-rule gene products disintegrates. Their actions, however, have stamped a permanent set of labels—positional values—on the cells of the blastoderm. These positional labels are recorded in the persistent activation of certain of the segment-polarity genes and of the homeotic selector genes, which serve to maintain the segmental organization of the larva and adult. The segment-polarity gene *Engrailed* provides a good example. Its RNA transcripts are seen in the cellular blastoderm in a series of 14 bands, each approximately one cell wide, corresponding to the anteriormost portions of the future parasegments (**Figure 22–41**).

The segment-polarity genes are expressed in patterns that repeat from one parasegment to the next, and their bands of expression appear in a fixed relationship to the bands of expression of the pair-rule genes that help to induce them. However, the production of this pattern within each parasegment depends on interactions among the segment-polarity genes themselves. These interactions occur at stages when the blastoderm has already become fully partitioned into separate cells, so that cell–cell signaling of the usual sort has to come into play. A large subset of the segment-polarity genes code for components of two signal transduction pathways, the Wnt pathway and the Hedgehog pathway, including the secreted signal proteins Wingless (a Wnt family member) and Hedgehog. These are expressed in different bands of cells that serve as signaling

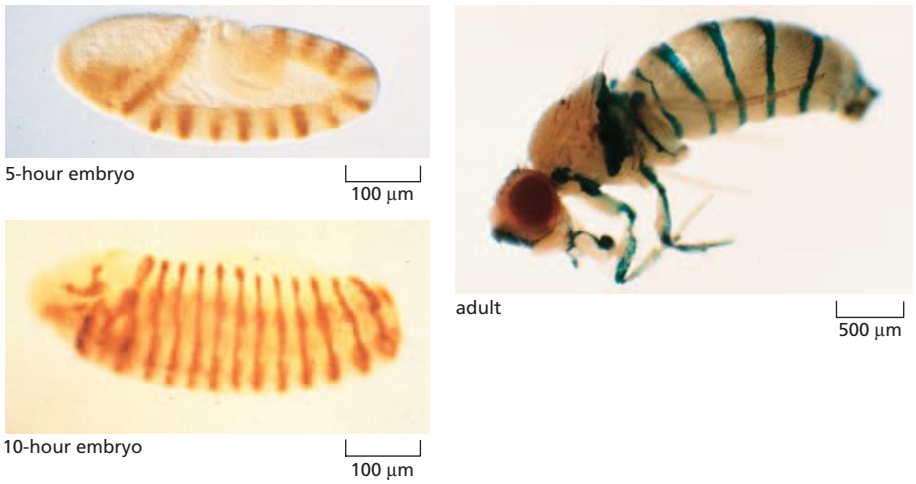


Figure 22–41 The pattern of expression of *Engrailed*, a segment-polarity gene. The *Engrailed* pattern is shown in a 5-hour embryo (at the extended germband stage), a 10-hour embryo, and an adult (whose wings have been removed in this preparation). The pattern is revealed by an antibody *(brown)* against the Engrailed protein (for the 5- and 10-hour embryos) or (for the adult) by constructing a strain of *Drosophila* containing the control sequences of the *Engrailed* gene coupled to the coding sequence of the reporter *LacZ*, whose product is detected histochemically through the blue product of a reaction that it catalyzes. Note that the *Engrailed* pattern, once established, is preserved throughout the animal's life. (From C. Hama, Z. Ali and T.B. Kornberg, *Genes Dev.* 4:1079–1093, 1990. With permission from Cold Spring Harbor Laboratory Press.)

centers within each parasegment, and they act to maintain and refine the expression of other segment-polarity genes. Moreover, although their initial expression is determined by the pair-rule genes, the two signal proteins regulate one another's expression in a mutually supportive way, and they proceed to help trigger expression of genes such as *Engrailed* in precisely the correct sites.

The *Engrailed* expression pattern will persist throughout life, long after the signals that organized its production have disappeared (see Figure 22–41). This example illustrates not only the progressive subdivision of the embryo by means of more and more narrowly localized signals, but also the transition between the transient signaling events of early development and the later stable maintenance of developmental information.

Besides regulating the segment-polarity genes, the products of pair-rule genes collaborate with the products of gap genes to cause the precisely localized activation of a further set of spatial labels—the homeotic selector genes. It is the homeotic selector genes that permanently distinguish one parasegment from another. In the next section we examine these selector genes in detail and consider their role in cell memory.

Summary

The fly Drosophila *has been the foremost model organism for study of the genetics of animal development. Like other insects, it begins its development with a series of nuclear divisions generating a syncytium, and a large amount of early patterning occurs in this single giant multinucleate cell. The pattern originates with asymmetry in the egg, organized both by localized deposits of mRNA inside the egg and by signals from the follicle cells around it. Positional information in the multinucleate embryo is supplied by four intracellular gradients that are set up by the products of four groups of maternal-effect genes called egg-polarity genes. These control four distinctions fundamental to the body plan of animals: dorsal versus ventral, endoderm versus mesoderm and ectoderm, germ cells versus somatic cells, and head versus rear.*

The egg-polarity genes operate by setting up graded distributions of gene regulatory proteins in the egg and early embryo. The gradients along the anteroposterior axis initiate the orderly expression of gap genes, pair-rule genes, segment-polarity genes, and homeotic selector genes. These, through a hierarchy of interactions, become expressed in some regions of the embryo and not others, progressively subdividing the blastoderm into a regular series of repeating modular units called segments. The complex patterns of gene expression reflect the modular organization of the regulatory DNA, with separate enhancers of an individual gene responsible for separate parts of its expression pattern.

The segment-polarity genes come into play toward the end of the segmentation process, soon after the syncytium has become partitioned into separate cells, and they control the internal patterning of each segment through cell–cell signaling via the Wnt (Wingless) and Hedgehog pathways. This leads to persistent localized activation of genes such as Engrailed, *giving cells a remembered record of their anteroposterior address within the segment. Meanwhile, a new cell–cell signaling gradient is also set up along the dorsoventral axis, with the TGFβ family member Decapentaplegic (Dpp) and its antagonist, Short gastrulation, acting as the morphogens. This gradient helps to refine the assignment of different characters to cells at different dorsoventral levels. Homologous proteins are also known to control the patterning of the ventrodorsal axis in vertebrates.*

HOMEOTIC SELECTOR GENES AND THE PATTERNING OF THE ANTEROPOSTERIOR AXIS

As development proceeds, the body becomes more and more complex. In all this growing complexity there is, however, a simplifying feature that puts an understanding of the whole developmental process within our grasp. Again and again, in every species and at every level of organization, we find that complex structures

are made by repeating a few basic themes with variations. Thus, a limited number of basic differentiated cell types, such as muscle cells or fibroblasts, recur with subtle individual variations in different sites. These cell types are organized into a limited variety of tissue types, such as muscle or tendon, which again are repeated with subtle variations in different regions of the body. From the various tissues, organs such as teeth or digits are built—molars and incisors, fingers and thumbs and toes—a few basic kinds of structure, repeated with variations.

Wherever we find this phenomenon of *modulated repetition*, we can break down the developmental biologist's problem into two kinds of question: what is the basic construction mechanism common to all the objects of the given class, and how is this mechanism modified to give the observed variations? The embryo uses a combinatorial strategy to generate its complexity, and we can use a combinatorial strategy to understand it.

The segments of the insect body provide a very clear example. We have already sketched the way in which the rudiment of a single typical segment is constructed. We must now consider how one segment is caused to be different from another.



Figure 22–42 A homeotic mutation. The fly shown here is an *Antennapedia* mutant. Its antennae are converted into leg structures by a mutation in the regulatory region of the *Antennapedia* gene that causes it to be expressed in the head. Compare with the normal fly shown in Figure 22–24. (Courtesy of Matthew Scott.)

The Hox Code Specifies Anterior–Posterior Differences

The first glimpse of a genetic answer to the question of how each segment acquires its individual identity came over 80 years ago, with the discovery of the first of a set of mutations in *Drosophila* that cause bizarre disturbances of the organization of the adult fly. In the *Antennapedia* mutant, for example, legs sprout from the head in place of antennae (**Figure 22–42**), while in the *Bithorax* mutant, portions of an extra pair of wings appear where normally there should be the much smaller appendages called halteres. These mutations transform parts of the body into structures appropriate to other positions and are called *homeotic*. A whole set of **homeotic selector genes** determines the anteroposterior character of the segments of the fly.

The genes of this set—eight of them in the fly—are all related to one another as members of a multigene family, and they all lie in one or the other of two tight gene clusters known as the **Bithorax complex** and the **Antennapedia complex**. The genes in the Bithorax complex control the differences among the abdominal and thoracic segments of the body, while those in the Antennapedia complex control the differences among thoracic and head segments. Comparisons with other species show that the same genes are present in essentially all animals, including humans. These comparisons also reveal that the Antennapedia and Bithorax complexes are the two halves of a single entity, called the **Hox complex**, that has become split in the course of the fly's evolution, and whose members operate in a coordinated way to exert their control over the head-to-tail pattern of the body.

Homeotic Selector Genes Code for DNA-Binding Proteins That Interact with Other Gene Regulatory Proteins

To a first approximation each homeotic selector gene is normally expressed in just those regions that develop abnormally when the gene is mutated or absent. The products of these genes can thus be viewed as molecular address labels possessed by the cells of each parasegment: they are the physical embodiment of the cells' positional value. If the address labels are changed, the parasegment behaves as though it were located somewhere else, and deletion of the entire complex results in a larva whose body segments are all alike (**Figure 22–43**).

A first problem, therefore, is to understand how the homeotic selector gene products act on the basic segment-patterning machinery to give each parasegment its individuality. The products of the homeotic selector genes are gene regulatory proteins, all related to one another by the possession of a highly conserved DNA-binding *homeodomain* (60 amino acids long), discussed in Chapter 7. The corresponding segment in the DNA sequence is called a *homeobox* from which, by abbreviation, the Hox complex takes its name.

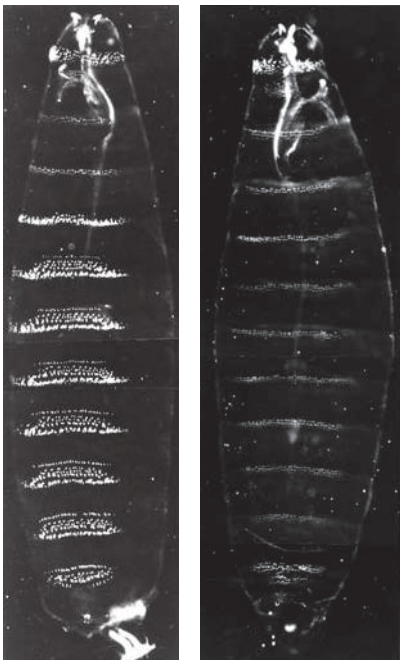


Figure 22–43 The effect of deleting most of the genes of the Bithorax complex. (A) A normal *Drosophila* larva shown in dark-field illumination; (B) the mutant larva with the Bithorax complex largely deleted. In the mutant the parasegments posterior to P5 all have the appearance of P5. (From G. Struhl, *Nature* 293:36–41, 1981. With permission from Macmillan Publishers Ltd.)

If the products of the homeotic selector genes are similar in their DNA-binding regions, how do they exert different effects so as to make one parasegment different from the next? The answer seems to lie largely in the parts of the proteins that do not bind directly to DNA but interact with other proteins in DNA-bound complexes. The different partners in these complexes act together with the homeotic selector proteins to dictate which DNA-binding sites will be recognized and whether the effect on transcription at those sites will be activation or repression. In this way, the products of the homeotic selector genes combine with other gene regulatory proteins and modulate their actions so as to give each parasegment its characteristic features.

The Homeotic Selector Genes Are Expressed Sequentially According to Their Order in the Hox Complex

To understand how the Hox complex provides cells with positional values, we also need to consider how the expression of the *Hox* genes themselves is regulated. The coding sequences of the eight homeotic selector genes in the Antennapedia and bithorax complexes are interspersed amid a much larger quantity—a total of about 650,000 nucleotide pairs—of regulatory DNA. This DNA includes binding sites for the products of egg-polarity and segmentation genes. The regulatory DNA in the Hox complex acts as an interpreter of the multiple items of positional information supplied to it by all these gene regulatory proteins. In response, a particular set of homeotic selector genes is transcribed, appropriate to the location.

In the pattern of control there is a remarkable regularity. The sequence in which the genes are ordered along the chromosome, in both the Antennapedia and the Bithorax complexes, corresponds almost exactly to the order in which they are expressed along the axis of the body (**Figure 22–44**). This suggests that the genes are activated serially by some process that is graded—in duration or intensity—along the axis of the body, and whose action spreads gradually along the chromosome. The most "posterior" of the genes expressed in a cell generally dominates, driving down expression of the previously activated "anterior" genes and dictating the character of the segment. The gene regulatory mechanisms underlying these phenomena are still not well understood, but their consequences are profound. We shall see that the serial organization of gene expression in the Hox complex is a fundamental feature that has been highly conserved in the course of evolution.

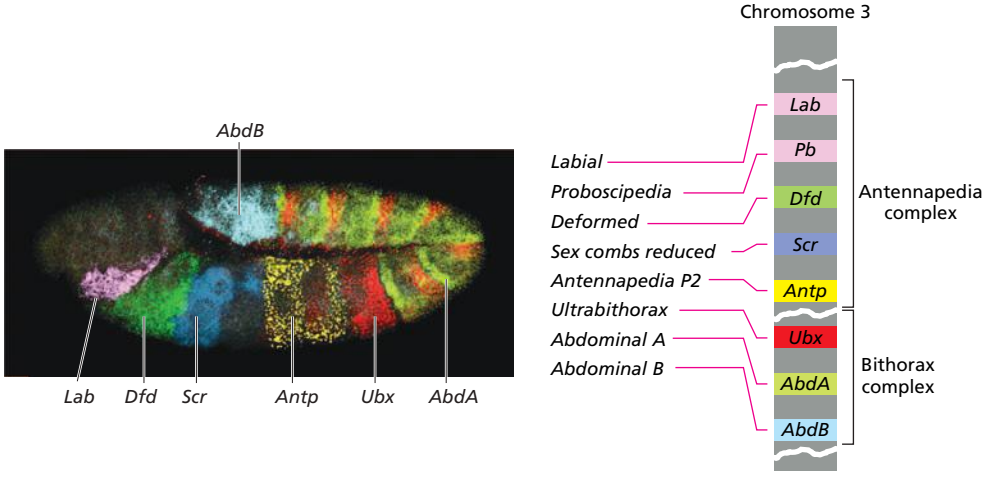


Figure 22–44 The patterns of expression compared to the chromosomal locations of the genes of the Hox complex. The diagram shows the sequence of genes in each of the two subdivisions of the chromosomal complex. This corresponds, with minor deviations, to the spatial sequence in which the genes are expressed, shown in the photograph of an embryo at the extended germ band stage, about 5 hours after fertilization. The embryo has been stained by *in situ* hybridization with differently labeled probes to detect the mRNA products of different *Hox* genes in different colors. (Photograph courtesy of William McGinnis, adapted from D. Kosman et al., *Science* 305:846, 2004. With permission from AAAS.)

There are hundreds of other homeobox-containing genes in the genome of the fly—and of other animal species—but most of them are scattered and not clustered in complexes such as the Hox complex. They have many different gene regulatory functions, but a substantial proportion of them have roles akin to that of the *Hox* genes: they control the variations on a basic developmental theme. Different classes of neurons, for example, are often distinguished from one another by expression of specific genes of this large superfamily.

The Hox Complex Carries a Permanent Record of Positional Information

The spatial pattern of expression of the genes in the Hox complex is set up by signals acting early in development, but the consequences are long-lasting. Although the pattern of expression undergoes complex adjustments as development proceeds, the Hox complex behaves in each cell as though stamped with a permanent record of the anteroposterior position that the cell occupied in the early embryo. In this way, the cells of each segment are equipped with a long-term memory of their location along the anteroposterior axis of the body—in other words, with an anteroposterior positional value. As we shall see in the next section, the memory trace imprinted on the Hox complex governs the segment-specific identity not only of the larval segments, but also of the structures of the adult fly, which are generated at a much later stage from the larval imaginal discs and other nests of imaginal precursor cells in the larva.

The molecular mechanism of the cell memory for this positional information relies on two types of regulatory inputs. One is from the homeotic selector genes themselves: many of the Hox proteins autoactivate the transcription of their own genes. Another crucial input is from two large complementary sets of proteins that control chromatin structure, called the *Polycomb group* and the *Trithorax group*. If these regulators are defective, the pattern of expression of the homeotic selector genes is set up correctly at first but is not correctly maintained as the embryo grows older.

The two sets of regulators act in opposite ways. Trithorax group proteins are needed to maintain the transcription of *Hox* genes in cells where transcription has already been switched on. In contrast, Polycomb group proteins form stable complexes that bind to the chromatin of the Hox complex and maintain the repressed state in cells where *Hox* genes have not been activated at the critical time (**Figure 22–45**). The developmental memory involves specific covalent modifications of histones in nucleosomes in the neighborhood of the *Hox* genes, leading to changes in the state of the chromatin that can be perpetuated from one cell generation to the next, as discussed in Chapters 4 and 7.

The Anteroposterior Axis Is Controlled by *Hox* Selector Genes in Vertebrates Also

Homologs of the *Drosophila* homeotic selector genes have been found in almost every animal species studied, from cnidarians (hydroids) and nematodes to mollusks and mammals. Remarkably, these genes are often grouped in complexes similar to the insect Hox complex. In the mouse there are four such complexes—called the *HoxA, HoxB, HoxC,* and *HoxD* complexes—each on a different chromosome. Individual genes in each complex can be recognized by their sequences as counterparts of specific members of the *Drosophila* set. Indeed, mammalian *Hox* genes can function in *Drosophila* as partial replacements for the corresponding *Drosophila Hox* genes. It appears that each of the four mammalian Hox complexes is, roughly speaking, the equivalent of a complete insect complex (that is, an Antennapedia complex plus a Bithorax complex) (**Figure 22–46**).

The ordering of the genes within each vertebrate Hox complex is essentially the same as in the insect Hox complex, suggesting that all four vertebrate

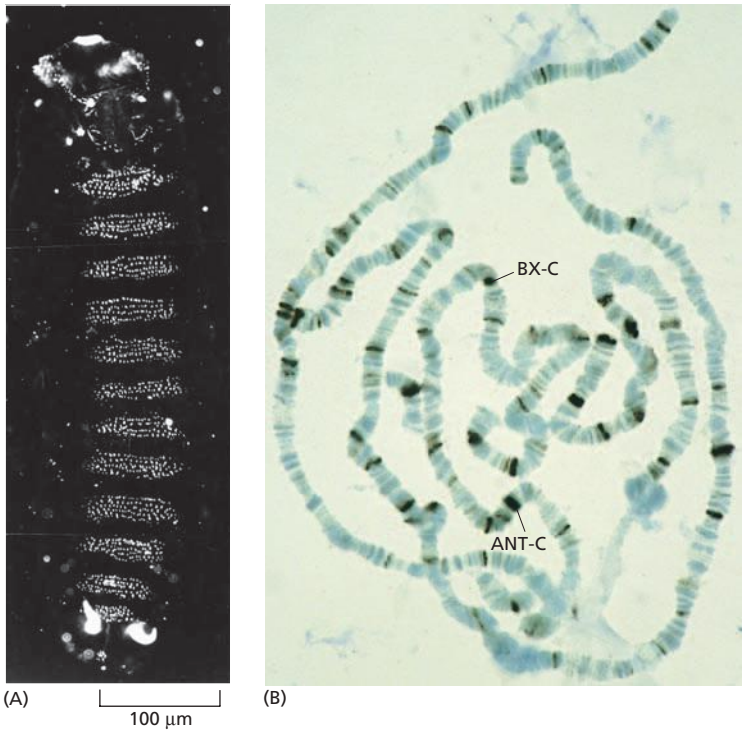


Figure 22–45 Action of genes of the Polycomb group. (A) Photograph of a mutant embryo defective for the gene *Extra sex combs (Esc)* and derived from a mother also lacking this gene. The gene belongs to the Polycomb group. Essentially all segments have been transformed to resemble the most posterior abdominal segment (compare with Figure 22–43). In the mutant the pattern of expression of the homeotic selector genes, which is roughly normal initially, is unstable in such a way that all these genes soon become switched on all along the body axis. (B) The normal pattern of binding of Polycomb protein to *Drosophila* giant chromosomes, visualized with an antibody against Polycomb. The protein is bound to the Antennapedia complex (ANT-C) and the Bithorax complex (BX-C) as well as about 60 other sites. (A, from G. Struhl, *Nature* 293:36–41, 1981. With permission from Macmillan Publishers Ltd. B, courtesy of B. Zink and R. Paro, *Trends Genet.* 6:416–421, 1990. With permission from Elsevier.)

complexes originated by duplications of a single primordial complex and have preserved its basic organization. Most tellingly, when the expression patterns of the *Hox* genes are examined in the vertebrate embryo by *in situ* hybridization, it turns out that the members of each complex are expressed in a head-to-tail series along the axis of the body, just as they are in *Drosophila* (**Figure 22–47**). The pattern is most clearly seen in the neural tube, but is also visible in other tissues, especially the mesoderm. With minor exceptions this anatomical ordering matches the chromosomal ordering of the genes in each complex, and corresponding genes in the four different Hox complexes have almost identical anteroposterior domains of expression.

The gene expression domains define a detailed system of correspondences between insect body regions and vertebrate body regions (see Figure 22–46). The parasegments of the fly correspond to a similarly labeled series of segments in the anterior part of the vertebrate embryo. These are most clearly demarcated in the hindbrain (see Figures 22–46 and 22–47), where they are called *rhombomeres.* In the tissues lateral to the hindbrain the segmentation is seen in the series of *branchial arches,* prominent in all vertebrate embryos—the precursors of the system of gills in fish and of the jaws and structures of the neck in mammals; each pair of rhombomeres in the hindbrain corresponds to one branchial arch. In the hindbrain, as in *Drosophila,* the boundaries of the expression domains of many of the *Hox* genes are aligned with the boundaries of the anatomical segments.

The products of the mammalian *Hox* genes appear to specify positional values that control the anteroposterior pattern of parts in the hindbrain, neck, and trunk (as well as some other parts of the body). As in *Drosophila,* when a posterior *Hox* gene is artificially expressed in an anterior region, it can convert the anterior tissue to a posterior character. Conversely, loss of posterior *Hox* genes allows the posterior tissue where they are normally expressed to adopt an anterior character (**Figure 22–48**). The transformations observed in mouse *Hox* mutants are not always so straightforward and are often incomplete, because of a redundancy between genes in the four *Hox* gene clusters. But it seems clear that the fly and the mouse use essentially the same molecular machinery to give individual characters to successive regions along at least a part of their anteroposterior axis.

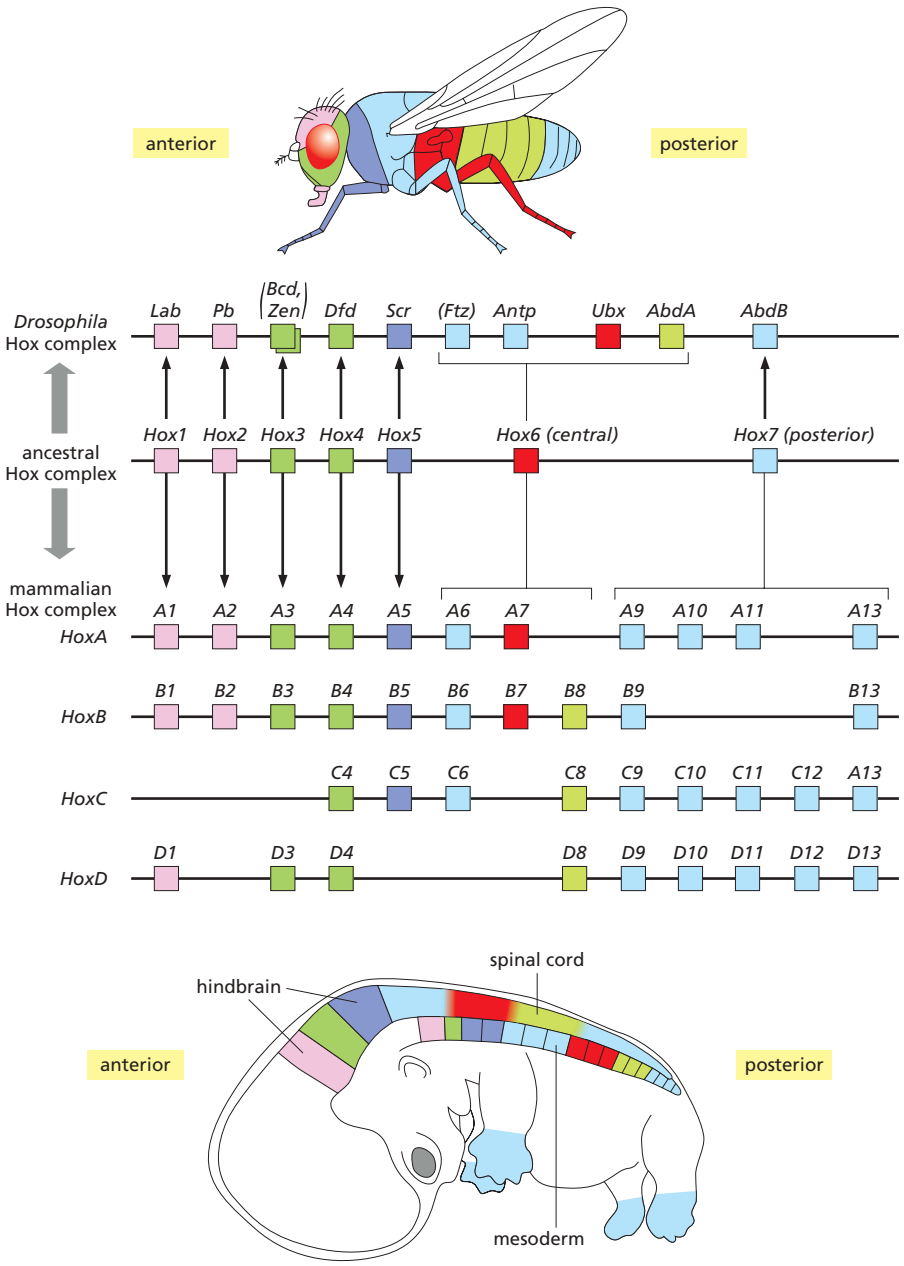


Figure 22–46 The Hox complex of an insect and the Hox complexes of a mammal compared and related to body regions. The genes of the Antennapedia and Bithorax complexes of *Drosophila* are shown in their chromosomal order in the top line; the corresponding genes of the four mammalian Hox complexes are shown below, also in chromosomal order. The gene expression domains in fly and mammal are indicated in a simplified form by *color* in the cartoons of animals above and below. However, the details of the patterns depend on developmental stage and vary somewhat from one mammalian Hox complex to another. Also, in many cases, genes shown here as expressed in an anterior domain are also expressed more posteriorly, overlapping the domains of more posterior *Hox* genes (see, for example, Figure 22–47). The complexes are thought to have evolved as follows: first, in some common ancestor of worms, flies, and vertebrates, a single primordial homeotic selector gene underwent repeated duplication to form a series of such genes in tandem—the ancestral Hox complex. In the *Drosophila* sublineage this single complex became split into separate Antennapedia and Bithorax complexes. Meanwhile, in the lineage leading to the mammals the whole complex was repeatedly duplicated to give four Hox complexes. The parallelism is not perfect because apparently some individual genes have been duplicated, others lost, and still others co-opted for different purposes (genes in parentheses in the top line) since the complexes diverged. (Based on a diagram courtesy of William McGinnis.)

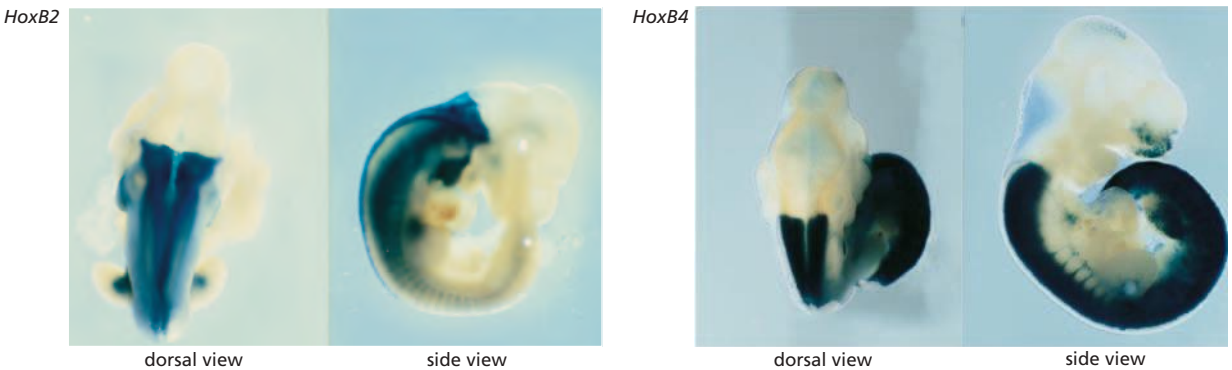
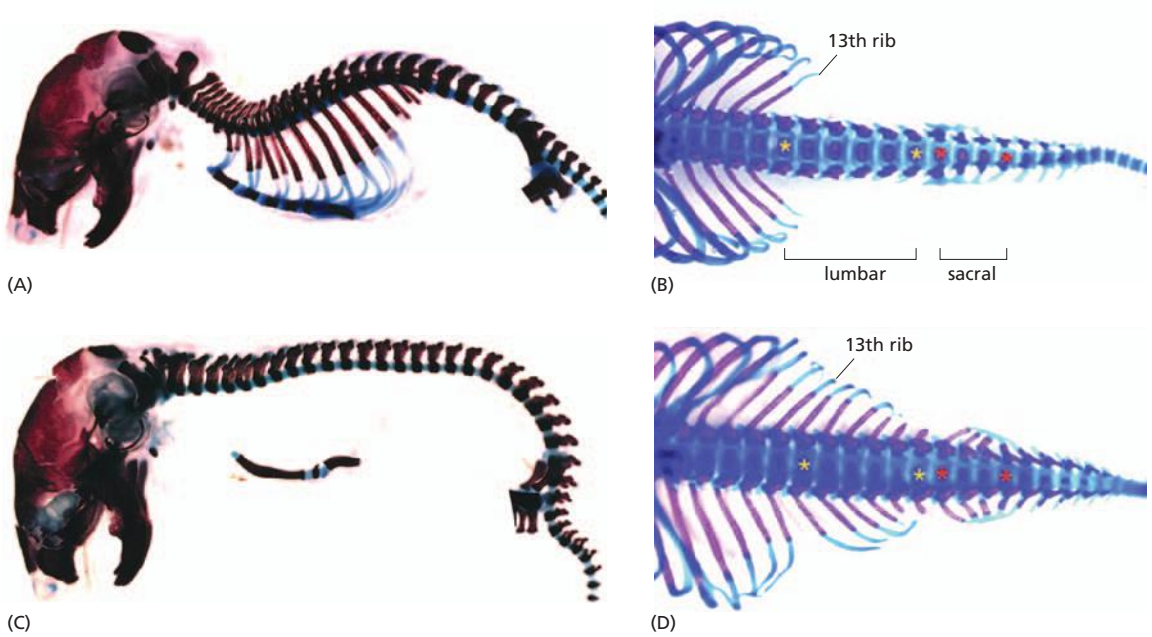


Figure 22–47 Expression domains of *Hox* genes in a mouse. The photographs show whole embryos displaying the expression domains of two genes of the HoxB complex *(blue stain)*. These domains can be revealed by *in situ* hybridization or, as in these examples, by constructing transgenic mice containing the control sequence of a *Hox* gene coupled to a *LacZ* reporter gene, whose product is detected histochemically. Each gene is expressed in a long expanse of tissue with a sharply defined anterior limit. The earlier the position of the gene in its chromosomal complex, the more anterior the anatomical limit of its expression. Thus, with minor exceptions, the anatomical domains of the successive genes form a nested set, ordered according to the ordering of the genes in the chromosomal complex. (Courtesy of Robb Krumlauf.)



Summary

The complexity of the adult body of an animal is built up by modulated repetition of a few basic types of structure. Thus, superimposed on the pattern of gene expression that repeats itself in every segment, there is a serial pattern of expression of homeotic selector genes that confer on each segment a different identity. The homeotic selector genes code for DNA-binding proteins of the homeodomain family. They are grouped in the Drosophila *genome in two clusters, called the Antennapedia and bithorax complexes, believed to be the two parts of a single primordial Hox complex that became split during evolution of the fly. In each complex, the genes are arranged in a sequence that matches their sequence of expression along the axis of the body. Hox gene expression is initiated in the embryo. It is maintained subsequently by the action of DNA-binding proteins of the Polycomb and Trithorax group, which stamp the chromatin of the Hox complex with a heritable record of its embryonic state of activation. Hox complexes homologous to that of* Drosophila *are found in virtually every type of animal that has been examined, from cnidarians to humans, and they appear to have an evolutionarily conserved role in patterning the anteroposterior axis of the body. Mammals have four Hox complexes, each showing a similar relationship between a serial arrangement of the genes in the chromosome and their serial pattern of expression along the body axis.*

Figure 22–48 Control of anteroposterior pattern by *Hox* genes in the mouse. (A,B) A normal mouse has about 65 vertebrae, differing in structure according to their position along the body axis: 7 cervical (neck), 13 thoracic (with ribs), 6 lumbar (bracketed by *yellow* asterisks in (B)), 4 sacral (bracketed by *red* asterisks in (B)), and about 35 caudal (tail). (A) shows a side view; (B) shows a dorsal view; for clarity, the limbs have been removed in each picture. (C) The *HoxA10* gene is normally expressed in the lumbar region (together with its paralogs *HoxC10* and *HoxD10*); here it has been artificially expressed in the developing vertebral tissue all along the body axis. As a result, the cervical and thoracic vertebrae are all converted to a lumbar character. (D) Conversely, when *HoxA10* is knocked out along with *HoxC10* and *HoxD10*, vertebrae that should normally have a lumbar or sacral character take on a thoracic character instead. (A and C, from M. Carapuço et al., *Genes Dev.* 19:2116–2121, 2005. With permission from Cold Spring Harbor Laboratory Press; B and D, from D.M. Wellik and M.R. Capecchi, *Science* 301:363–367, 2003. With permission from AAAS.)

ORGANOGENESIS AND THE PATTERNING OF APPENDAGES

We have seen that the segments of the insect larva are all variations on the same basic theme, with segmentation genes defining the basic repetitive module and homeotic selector genes giving each segment its individual character. The same applies to the major appendages of the adult insect body—legs, wings, antennae, mouthparts and external genitalia: they too are variations on a common basic theme. At a finer level of detail, we encounter the same wonderful simplification: the appendages—and many other parts of the body—consist of substructures that are themselves variations on a small number of basic evolutionarily conserved themes.

In this section we follow the course of development in *Drosophila* through to its end, narrowing our focus at each step to examine one example of the many related structures that are developing in parallel. As we go along, we shall point

out parallels with vertebrate structures that develop similarly, using not only the same general strategies but many of the same specific molecular mechanisms. But to avoid interrupting the narrative later, we must first briefly explain some key experimental methods, required to cope with a special problem that arises when we try to discover how genes control the later stages of development.

Conditional and Induced Somatic Mutations Make it Possible to Analyze Gene Functions Late in Development

As emphasized earlier, the same gene may be used repeatedly in many different situations—in different regions of the body, and at different times. Often, loss-of-function mutations disrupt early development so severely that the embryo or larva dies, depriving us of the opportunity to see how the mutation would affect later processes.

One way around this problem is to study conditional mutations. If we have, for example, a temperature-sensitive mutation in the gene of interest, we can maintain the animal during early development at a low temperature, where the gene product functions normally, and then disable the gene product whenever we please by raising the temperature to discover the late functions.

Other methods involve actually modifying the DNA in subsets of cells at late stages of development—a sort of genetic surgery on individual cells that allows mutant groups of cells of a specified genotype to be generated at a chosen time in development. This remarkable feat can be achieved by *induced somatic recombination*, and the resulting organism is called a **genetic mosaic**. By means of genetic mosaics, we can not only bypass the problem of lethality when a gene function is perturbed in the organism as a whole; we can also explore the function of the gene in cell–cell interactions, by juxtaposing mutant and nonmutant cells. We can test, for example, whether cells use the gene product to send a signal to neighbors, or to receive a signal from them, or neither. And by inducing the genetic change at different times, we can find out precisely when the gene acts to produce a particular effect.

A current technique for induced somatic recombination uses transgenic flies that have been bred to contain two types of yeast-derived genetic elements: the FLP site-specific recombinase gene, and the FLP Recombinase Target (FRT) sequence. Typically, the animal is homozygous for an insertion of the FRT sequence close to the centromere on a chosen chromosome arm, while a construct consisting of the *Flp* gene under a heat-shock promoter is inserted elsewhere in the genome. If such a transgenic embryo or larva is given a heat shock (that is, exposed to a high temperature for a few minutes), expression of *Flp* is induced, and this enzyme catalyzes crossing-over and recombination between the maternal and paternal chromosomes at the FRT site. If the heat shock is adjusted to be sufficiently mild, this event will occur in only one or a few cells, scattered at random. As explained in **Figure 22–49**, if the animal is also heterozygous for a gene of interest in the crossed-over chromosomal region, the process can result in a pair of daughter cells that are homozygous, the one receiving two copies of the maternal allele of the gene, the other receiving two copies of the paternal allele. Each of these daughter cells will normally grow and divide to give clonal patches of homozygous progeny.

The occurrence of the crossover can be detected if the animal is chosen to be also heterozygous for a genetic marker that lies on the same chromosome arm as the gene of interest and so undergoes crossing over in company with it. In this way clearly marked homozygous mutant clones of cells can be created to order. Either FLP and FRT, or the analogous Cre and Lox pair of recombination elements, can also be used in other configurations to switch expression of a gene on or off (see Figure 5–79). With these techniques, one can discover what happens, for example, when cells are caused to produce a particular signal molecule at an abnormal site, or are deprived of a particular receptor.

Instead of using a heat-shock promoter to drive expression of the FLP recombinase, one can use a copy of the regulatory sequence of a gene in the fly's

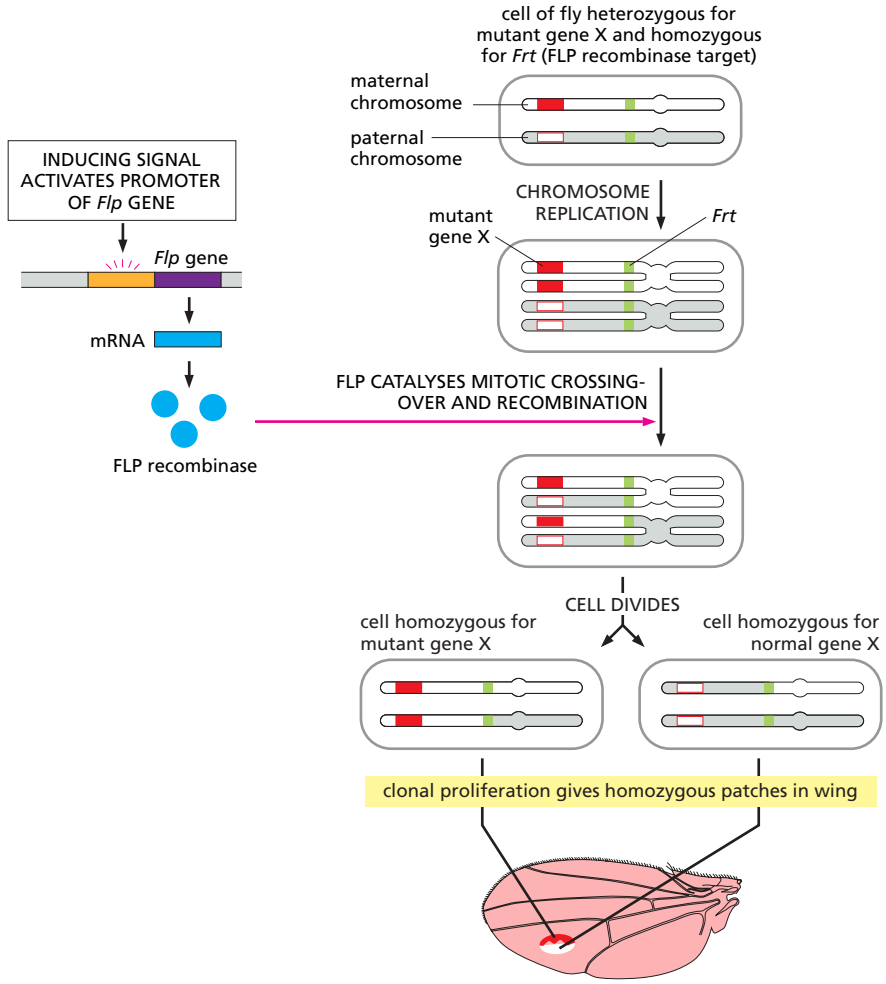


Figure 22–49 Creation of mutant cells by induced somatic recombination. The diagrams follow the fate of a single pair of homologous chromosomes, one from the father *(shaded)*, the other from the mother *(unshaded)*. These chromosomes have an *Frt* element *(green)* inserted close to their centromere, and contain a locus for a gene of interest—gene X—farther out along the same chromosome arm. The paternal chromosome (in this example) carries the wild-type allele of gene X *(open red box)* while the maternal chromosome carries a recessive mutant allele *(filled red box)*. Recombination by exchange of DNA between the maternal and paternal chromosomes, catalyzed by the FLP recombinase, can give rise to a pair of daughter cells, one containing two wild-type copies of gene X, the other containing two mutant copies. To help identify the cells where recombination has occurred, the maternal and paternal chromosomes can be chosen to carry different genetic markers (not shown here), capable of generating a visible product, and positioned on the chromosome so that recombination involving the marker locus—resulting in a visible alteration in the appearance of the cells—can be taken as a sure sign that gene X has also undergone recombination.

normal genome that is expressed at some interesting time and place. The recombination event will then be triggered, and mutant cells created, at just the sites where that gene is normally expressed. A variant of this technique uses transcriptional regulation machinery borrowed from yeast, rather than genetic recombination machinery, to switch expression of a chosen fly gene reversibly on or off according to the normal pattern of expression of some other chosen fly gene (**Figure 22–50**).

By switching gene functions off or on at specific times and places in these ways, developmental biologists can set about deciphering the system of genetically specified signals and responses that control the patterning of any organ of the body.

Body Parts of the Adult Fly Develop From Imaginal Discs

The external structures of the adult fly are formed largely from rudiments called **imaginal discs**—groups of cells that are set aside, apparently undifferentiated, in each segment of the larva. The discs are pouches of epithelium, shaped like crumpled and flattened balloons, and continuous with the epidermis (the surface layer) of the larva. There are 19 of them, arranged as 9 pairs on either side of the larva plus 1 disc in the midline (**Figure 22–51**). They grow and develop their internal pattern as the larva grows, until finally, at metamorphosis, they evert (turn inside out), extend, and differentiate overtly to form the epidermal layer of the adult. The eyes and antennae develop from one pair of discs, the wings and part of the thorax from another, the first pair of legs from another, and so on.

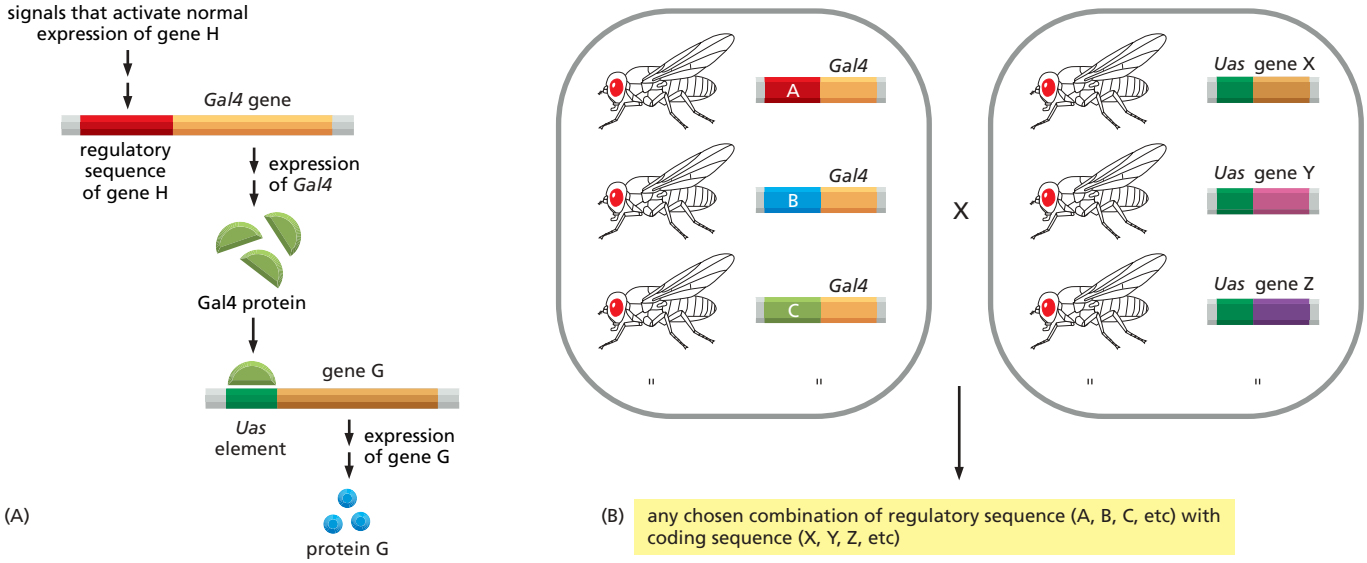


Figure 22–50 The *Gal4/Uas* technique for controlled gene misexpression in *Drosophila*. The method allows one to drive expression of a chosen gene G at the places and times where some other *Drosophila* gene H is normally expressed. (A) A transgenic animal is created, with two separate constructs inserted in its genome. One insert consists of a yeast-specific regulatory sequence, called the *Uas* (upstream activating sequence) element, coupled to a copy of the coding sequence of gene G. The other insert contains the coding sequence of the yeast *Gal4* gene, whose product is a yeast-specific gene regulatory protein that binds to the *Uas* element; this *Gal4* insert is placed next to, and controlled by, the regulatory region of gene H. Wherever gene H is normally expressed, Gal4 protein is also made and drives transcription of gene G. (B) Although one can achieve the same result by linking a copy of the H regulatory sequence directly to the G coding sequence, the *Gal4/Uas* approach allows a strategy that is more efficient in the long run. Two separate "libraries" of transgenic flies are constructed, one containing *Gal4* inserts driven by a variety of regulatory sequences of different genes A, B, C, etc., the other containing *Uas* inserts driving a variety of different coding sequences X, Y, Z, etc. By mating a fly from one library with a fly from the other, any desired coding sequence can be functionally coupled to any desired regulatory sequence. To generate the library of flies with *Gal4* insertions at useful sites, flies are first produced with *Gal4* insertions at random locations in their genome. These are then mated with flies containing a *Uas* element linked to a reporter gene with an easily detectable product. Expression of the reporter reveals whether *Gal4* has been inserted at a site that brings its expression under the control of an interesting enhancer; flies showing interesting reporter patterns are kept and studied. This is called the enhancer trap technique, because it provides a way to hunt out and characterize interesting regulatory sequences in the genome.

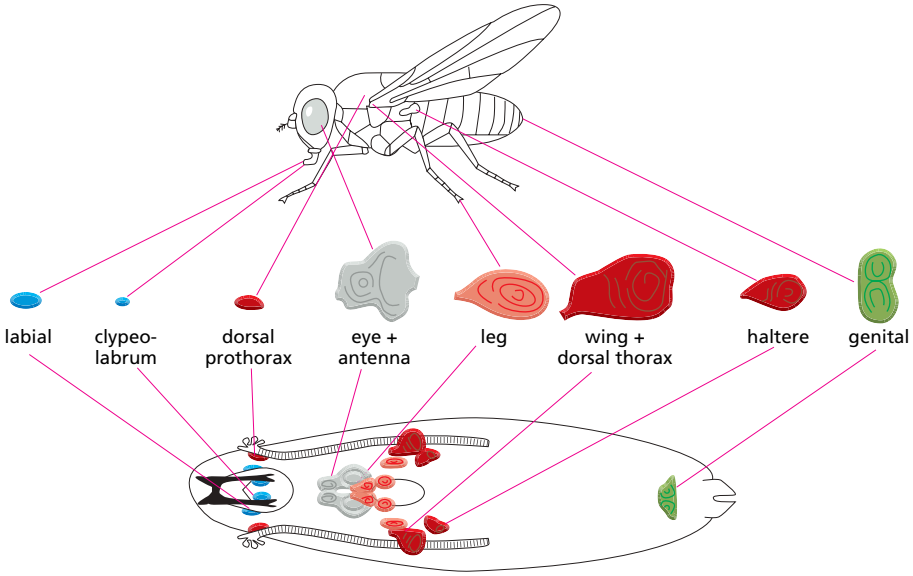


Figure 22–51 The imaginal discs in the *Drosophila* larva and the adult structures they give rise to. (After J.W. Fristrom et al., in Problems in Biology: RNA in Development [E.W. Hanley, ed.], p. 382. Salt Lake City: University of Utah Press, 1969.)

Homeotic Selector Genes Are Essential for the Memory of Positional Information in Imaginal Disc Cells

The cells of one imaginal disc look like those of another, but grafting experiments show that they are in fact already regionally determined and nonequivalent. If one imaginal disc is transplanted into the position of another in the larva and the larva is then left to go through metamorphosis, the grafted disc is found to differentiate autonomously into the structure appropriate to its origin: a wing disc will give wing structures, a haltere disc, haltere structures, regardless of its new site. This shows that the imaginal disc cells are governed by a memory of their original position. By a more complex serial grafting procedure that lets the imaginal disc cells proliferate for an extended period before differentiating, it can be shown that this cell memory is stably heritable (with rare lapses) through an indefinitely large number of cell generations.

The homeotic selector genes are essential components of the memory mechanism. If, at any stage in the long period leading up to differentiation at metamorphosis, both copies of a homeotic selector gene are eliminated by induced somatic recombination from a clone of imaginal disc cells that would normally express that gene, those cells will differentiate into incorrect structures, as though they belonged to a different segment of the body. These and other observations indicate that each cell's memory of positional information depends on the continued activity of the homeotic selector genes. This memory, furthermore, is expressed in a cell-autonomous fashion—each cell appears to maintain its state individually, depending on its own history and genome.

Specific Regulatory Genes Define the Cells That Will Form an Appendage

We must now examine how an appendage develops its internal pattern. We shall take the insect wing as our example.

The process begins with the early patterning mechanisms we have already discussed. The anteroposterior and dorsoventral systems of signals in the early embryo in effect mark out an orthogonal grid in the blastoderm, in the form of dorsoventral, anteroposterior, and periodically spaced segmental gene expression boundaries. At certain points of intersection of these boundaries, the combination of genes expressed is such as to switch a cluster of cells into the imaginal disc pathway.

In molecular terms this corresponds to switching on expression of imaginal-disc-defining regulatory genes. In most of the discs, the gene *Distal-less* is switched on. This codes for a gene regulatory protein that is essential for the sustained growth required to create an elongated appendage such as a leg or an antenna with a proximodistal axis. In its absence, such appendages fail to form, and when it is artificially expressed at abnormal sites, misplaced appendages can be produced. *Distal-less* is expressed in a similar fashion in the developing limbs and other appendages of most species of invertebrates and vertebrates that have been examined (**Figure 22–52**). For the eye disc, another gene, *Eyeless*



Figure 22–52 Expression of *Distal-less* in developing legs and related appendages of various species. (A) A sea-urchin larva. (B) A moth larva. (A, from G. Panganiban et al., *Proc. Natl Acad. Sci. U.S.A.* 94:5162–5166, 1997. With permission from National Academy of Sciences; B, from G. Panganiban, L. Nagy and S.B. Carroll, *Curr. Biol.* 4:671–675, 1994. With permission from Elsevier.)

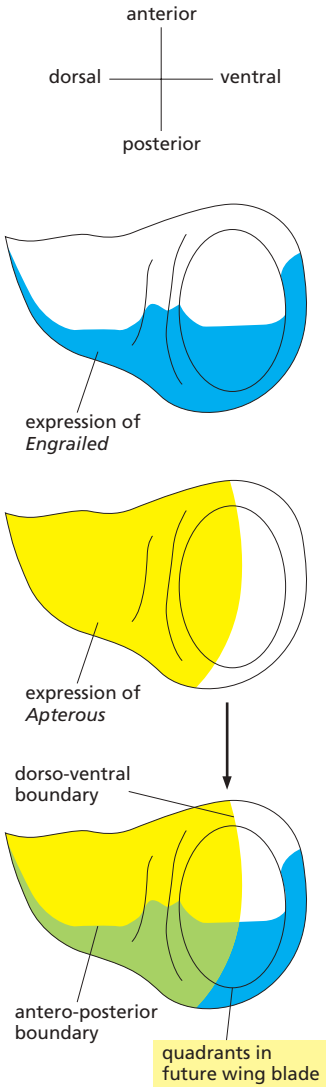


Figure 22–53 Gene expression domains in the wing imaginal disc, defining quadrants of the future wing. The wing blade itself derives from the oval-shaped domain toward the right, and it is divided into four quadrants by the expression of *Apterous* and *Engrailed,* as shown.

(together with two closely related genes), performs the corresponding role; it too has homologues with homologous functions—the *Pax6* genes that drive eye development in other species, as discussed in Chapter 7.

The Insect Wing Disc Is Divided into Compartments

From the outset, the cluster of cells forming the imaginal disc has the rudiments of an internal pattern, inherited from the earlier patterning process. For example, the cells in the posterior half of the wing-disc rudiment (and of most of the other imaginal-disc rudiments) express the segment-polarity gene *Engrailed*, while those in the anterior half do not. The initial asymmetries lay the foundations for a subsequent more detailed patterning, just as in the egg and early embryo.

The sectors of the wing disc defined by these early differences of gene expression correspond to specific parts of the future wing. The posterior, *Engrailed*-expressing region will form the posterior half of the wing, while the region that does not express *Engrailed* will form the anterior half. Meanwhile, the dorsal part of the wing disc expresses a gene called *Apterous*, while the ventral half does not. At metamorphosis, the disc folds along the line separating these domains to give a wing whose dorsal sheet of cells is derived from the *Apterous*-expressing region and whose ventral sheet is derived from the region that does not express *Apterous*. The wing margin, where these two epithelial sheets are joined, corresponds to the boundary of the *Apterous* expression domain in the disc (**Figure 22–53**).

The cells of the disc, having switched on expression of the genes that mark them as anterior or posterior, dorsal or ventral, retain this specification as the disc grows and develops. Because the cells are sensitive to these differences and selective in their choice of neighbors, sharply defined boundaries are formed between the four resultant sets of cells, with no mixing at the interfaces. The four corresponding quadrants of the disc are called **compartments**, because there is no exchange of cells between them (**Figure 22–54**).

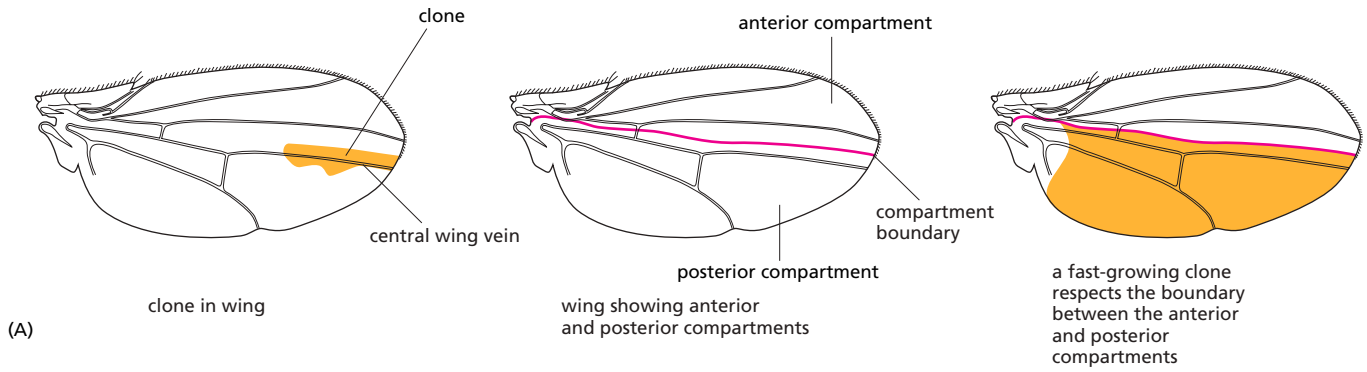
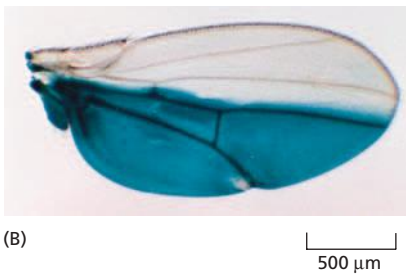


Figure 22–54 Compartments in the adult wing. (A) The shapes of marked clones in the *Drosophila* wing reveal the existence of a compartment boundary. The border of each marked clone is straight where it abuts the boundary. Even when a marked clone has been genetically altered so that it grows more rapidly than the rest of the wing and is therefore very large, it respects the boundary in the same way (drawing on right). Note that the compartment boundary does not coincide with the central wing vein. (B) The pattern of expression of the *Engrailed* gene in the wing, revealed by the same technique as for the adult fly shown in Figure 22–41. The compartment boundary coincides with the boundary of *Engrailed* gene expression. (A, after F.H.C. Crick and P.A. Lawrence, *Science* 189:340–347, 1975. With permission from AAAS; B, courtesy of Chihiro Hama and Tom Kornberg.)



Four Familiar Signaling Pathways Combine to Pattern the Wing Disc: Wingless, Hedgehog, Dpp, and Notch

Along each of the compartment boundaries—the anteroposterior boundary defined by *Engrailed* and the dorsoventral boundary defined by *Apterous*—cells in different states confront one another and interact to create narrow bands of specialized cells. These boundary cells produce new signals to organize the subsequent growth and more detailed patterning of the appendage.

Cells in the posterior wing compartment express the Hedgehog signal protein, but cannot respond to it. Cells in the anterior compartment can respond to Hedgehog. Because Hedgehog acts only over a short distance, the signal reception pathway is activated only in the narrow band of cells just anterior to the compartment boundary, where anterior and posterior cells are juxtaposed. These boundary cells respond by switching on expression of another signal molecule, Dpp—the same protein that we encountered previously, in the dorsoventral patterning of the early embryo (**Figure 22–55**). Dpp acts in its new context in much the same way as before: it spreads its effects outward from the boundary cells (by diffusion, via cytonemes, or through transfer from cell to cell by exocytosis or endocytosis), setting up a morphogen gradient to control the subsequent detailed pattern of growth and gene expression.

Analogous events occur at the dorsoventral compartment boundary (see Figure 22–55). Here, at the future wing margin, short-range communication mediated by the Notch pathway creates a band of boundary cells that produce another morphogen, the Wingless protein—the same signaling factor, belonging to the Wnt family, that acted earlier in the anteroposterior patterning of each embryonic segment. The Dpp and Wingless gradients, together with other signals and with the asymmetries of gene expression that we have discussed, combine to drive expression of other genes at precisely defined locations within each compartment.

The Size of Each Compartment Is Regulated by Interactions Among Its Cells

One of the most mysterious and ill-understood aspects of animal development is the control of growth: why does each part of the body grow to a precisely defined size? This problem is exemplified in remarkable way in the imaginal discs of *Drosophila*. By induced somatic recombination, one can, for example,

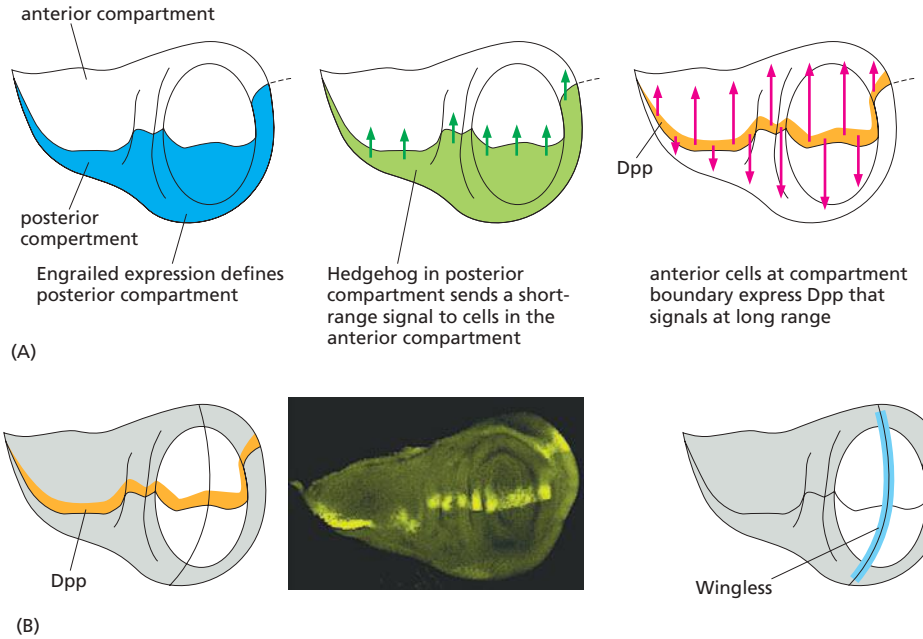


Figure 22–55 Morphogenetic signals created at compartment boundaries in the wing imaginal disc. (A) Creation of the Dpp signaling region at the anteroposterior compartment boundary through a Hedgehog-mediated interaction between the anterior and posterior cells. In an analogous way, a Notch-mediated interaction between dorsal and ventral cells creates a Wingless (Wnt) signaling region along the dorsoventral boundary. (B) The observed expression patterns of Dpp and Wingless. Although it seems clear that Dpp and Wingless act as morphogens, it is not yet certain how they spread out from their source. Cells in the imaginal disc have been seen to send out long cytonemes that may allow them to sense signals at a distance. Thus, the receiving cell may send its sensors to the source of the signal, instead of the signal moving to the receiving cell. (B, photographs courtesy of Sean Carroll and Scott Weatherbee, from S.J. Day and P.A. Lawrence, *Development* 127:2977–2987, 2000. With permission from The Company of Biologists.)

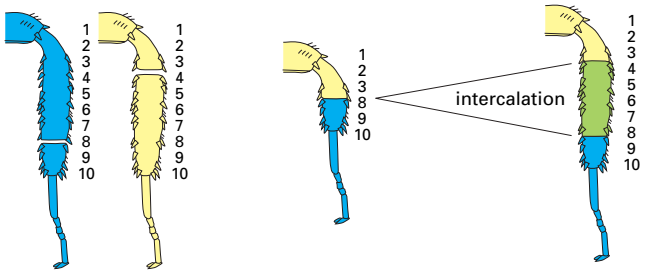


Figure 22–56 Intercalary regeneration. When mismatched portions of the growing cockroach leg are grafted together, new tissue *(green)* is intercalated (by cell proliferation) to fill in the gap in the pattern of leg structures, restoring a leg segment of normal size and pattern.

create a clonal patch of cells that proliferate more rapidly than the rest of the cells in the developing organ. The clone may grow to occupy almost the whole of the compartment in which it lies, and yet it does not overstep the boundary of the compartment. Astonishingly, its rapid growth has almost no effect on the compartment's final size, its shape, or even the details of its internal pattern (see Figure 22–54). Somehow, the cells within the compartment interact with one another to determine when their growth should stop, and each compartment behaves as a regulatory unit in this respect.

A first question is whether the size of the compartment is regulated so as to contain a set number of cells. Mutations in components of the cell-cycle control machinery can be used to speed up or slow down the rate of cell division without altering the rate of cell or tissue growth. This results in abnormally large numbers of abnormally small cells, or the converse, but the size—that is, the area—of the compartment is practically unchanged. Thus, the regulatory mechanism seems to depend on signals that indicate the physical distance between one part of the compartment and another, and on cellular responses that somehow read these signals so as to halt growth only when the spacing between the parts has reached its proper value.

This type of growth regulation is strikingly displayed in the **intercalary regeneration** that occurs when separate parts of a *Drosophila* imaginal disc or of a growing cockroach leg are surgically grafted together. After the graft, the cells in the neighborhood of the junction proliferate and fill in the parts of the pattern that should normally lie between them, continuing their growth until the normal spacing between landmarks is restored (**Figure 22–56**). The mechanisms that bring this about are a mystery, but it seems likely that they are similar to the mechanisms that regulate growth during normal development.

What mechanism could ensure that each little piece of the pattern within a compartment grows to its appropriate size, despite local disturbances in growth rate or starting conditions? The morphogen gradients (of Dpp and Wingless, for example) create a pattern by imposing different characters on cells in different positions. Could it be that the cells in each region can somehow sense how close the spacing of the pattern is—how steep the gradient of change in cell character—and continue their growth until the tissue is spread out to the right degree?

This idea has been tested by creating clones of cells in the wing disc in which downstream components of the Dpp signaling pathway are misexpressed so as to drive the level of pathway activation either higher or lower than in the neighboring cells. From the point of view of the cells, conditions at the boundary of the mutant clone are then equivalent to those produced by a very steep gradient of Dpp. The result is that cells in this neighborhood are stimulated to divide at an increased rate. Conversely, if the level of Dpp signaling is made uniform in the middle region of the developing wing disc, where it would normally be steeply graded, cell division there is inhibited. It seems that the steepness of the gradient does indeed control the rate of proliferation. But if that is so, how do cells sense the steepness of the gradient?

The answer is unknown, but there are strong hints that the mechanism depends on signals generated at cell–cell junctions, where cells with different levels of pathway activation make contact. As discussed in Chapter 19, mutations in junctional components such as the scaffold protein Discs-large (Dlg) or the cadherin superfamily member Fat can cause a dramatic failure of growth control, allowing the wing disc to grow far beyond its normal proper size. In the case of Fat, a set of other molecules, including protein kinases called Hippo and

Warts, have been identified as components of a signaling pathway that leads from Fat at the cell membrane to the control of gene expression in the nucleus. The products of the target genes include the cell-cycle regulator cyclin E and an inhibitor of apoptosis, as well as a microRNA, *Bantam,* that seems to be an essential part of the growth control mechanism. Despite these tantalizing facts, the mechanisms controlling organ size are still mysterious. If we can discover how they work in *Drosophila,* we may get some insight into the problem of the control of organ size in vertebrates, where our current perplexity on this fundamental question is even more profound. For in other aspects of organ development, as we now discuss, flies and vertebrates are unexpectedly similar at a molecular level, suggesting that their mechanisms of growth control may be similar also.

Similar Mechanisms Pattern the Limbs of Vertebrates

The limbs of vertebrates seem very different from those of insects. The insect wing, for example, consists mainly of two elaborately patterned sheets of epithelium, with very little tissue in between. In contrast, a limb of a vertebrate consists of an elaborately patterned system of muscles, bones and other connective tissues inside a thin and much more simply structured covering of epidermis. Moreover, the evolutionary evidence suggests that the last common ancestor of insects and vertebrates may have had neither legs, nor arms, nor wings, nor fins and that we have evolved these various appendages independently. And yet, when we examine the molecular mechanisms that control vertebrate limb development, we find a surprising number of similarities with the limbs of insects. We have already mentioned some of these resemblances, but there are many others: almost all the molecules we have already mentioned in the fly wing have their counterparts in the vertebrate limb, although these are expressed in different spatial relationships.

The parallels have been most thoroughly studied in the chick embryo. As we saw earlier, each leg or wing of a chick originates from a tongue-shaped limb bud, consisting of a mass of embryonic connective tissue cells, called mesenchyme cells, encased in a jacket of epithelium. In this structure, one finds expression of homologs of almost all the genes that we have mentioned in our account of *Drosophila* wing patterning, including *Distal-less*, *Wingless*, *Notch*, *Engrailed*, *Dpp*, and *Hedgehog*, mostly performing functions that seem more or less similar to their functions in the *Drosophila* wing disc (**Figure 22–57**).

The *Hox* genes likewise make an appearance in the limbs of both insects and vertebrates. In the insect appendage, the anterior and posterior compartments are distinguished by expression of different genes of the Hox complex—a result of the serial expression pattern of these genes along the anteroposterior axis of the body as a whole. In the vertebrate limb, genes of two of the vertebrate Hox complexes

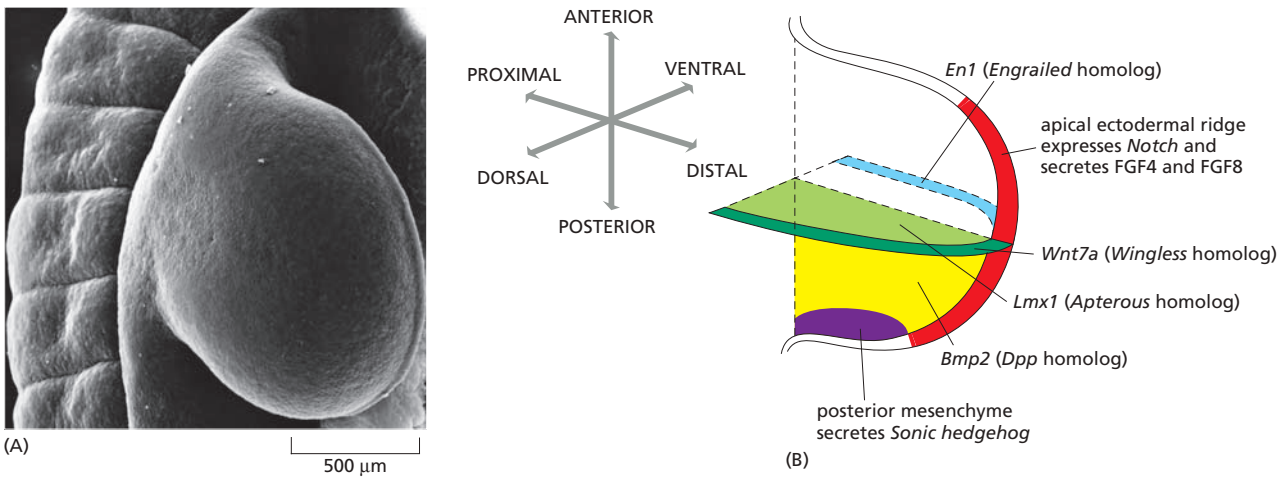


Figure 22–57 Molecules that control patterning in a vertebrate limb bud. (A) A wing bud of a chick embryo at 4 days of incubation. The scanning electron micrograph shows a dorsal view, with somites (the segments of the trunk of the embryo) visible to the left. At the distal margin of the limb bud a thickened ridge can just be seen—the apical ectodermal ridge. (B) Expression patterns of key signaling proteins and gene regulatory factors in the chick limb bud. The patterns are depicted schematically in two imaginary planes of section through the limb bud, one (horizontal) to show the dorsoventral system and the other (vertical) to show the anteroposterior and proximodistal systems. *Sonic hedgehog, Bmp2,* and *Lmx1* are expressed in the mesodermal core of the limb bud; the other molecules in the diagram are expressed in its epithelial covering. Almost all the molecules shown have homologs that are involved in patterning the *Drosophila* wing disc. (A, courtesy of Paul Martin.)

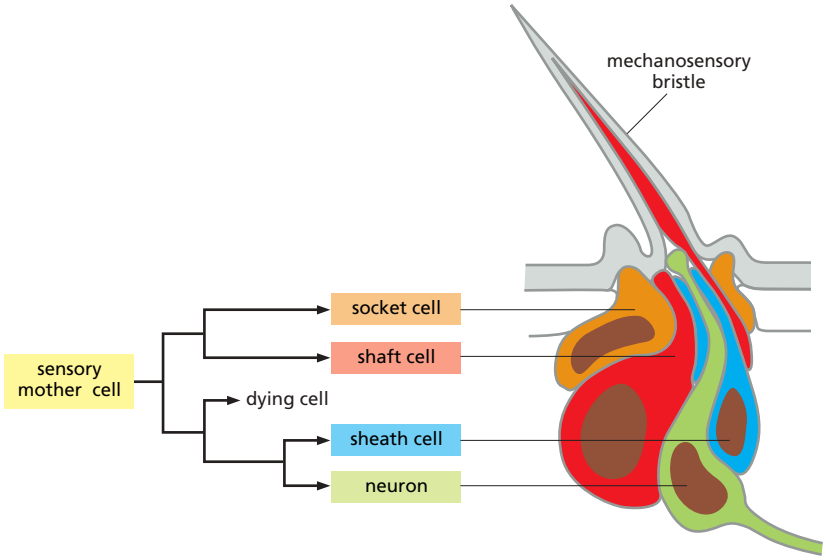


Figure 22–58 The basic structure of a mechanosensory bristle. The lineage of the four cells of the bristle—all descendants of a single sensory mother cell—is shown on the left.

(*HoxA* and *HoxD*) are expressed in a regular pattern, obedient to the usual rules of serial expression of genes in these complexes. They help, in conjunction with other factors such as the Tbx proteins mentioned earlier (see Figure 22–9), to regulate differences of cell behavior along the proximodistal limb axis.

According to one view, these molecular resemblances between developing limbs in different phyla reflect descent from a common ancestor that, while lacking limbs, had appendages of some sort built on similar principles—antennae, perhaps, or protruding mouthparts for snatching food. Modern limblike appendages, from the wings and legs of the fly to the arms and legs of a human, would then have evolved through activation of the genes for appendage formation at new sites in the body, as a result of changes in gene regulation.

Localized Expression of Specific Classes of Gene Regulatory Proteins Foreshadows Cell Differentiation

We now pick up again the thread of development in the *Drosophila* imaginal disc and follow it through to the final step at which cells become terminally differentiated. Narrowing our focus further, we take as our example the differentiation of just one type of small structure that arises in the imaginal disc epithelium: the **sensory bristle**.

The bristles that cover the body surface of an insect are miniature sense organs. Some respond to chemical stimuli, others to mechanical stimuli, but they are all constructed in a similar way. The structure is seen at its simplest in the mechanosensory bristles. Each of these consists of four cells: a shaft cell, a socket cell, a neural sheath cell, and a neuron (**Figure 22–58**). Movement of the shaft of the bristle excites the neuron, which sends a signal to the central nervous system.

The cells of the bristle of the adult fly derive from the imaginal disc epithelium, and all four of them are granddaughters or great-granddaughters (see Figure 22–58) of a single *sensory mother cell* that becomes distinct from the neighboring prospective epidermal cells during the last larval instar (**Figure 22–59**). (A fifth descendant dies, or in some tissues becomes a glial cell.) To account for the pattern of bristle differentiation, we have to explain first how the genesis of sensory mother cells is controlled and then how the five descendants of each such cell become different from one another.

Two genes, called *Achaete* and *Scute*, are crucial in initiating the formation of bristles in the imaginal disc epithelium. These genes have similar and overlapping functions and code for closely related gene regulatory proteins of the basic helix–loop–helix class (discussed in Chapter 7). As a result of disc-patterning mechanisms of the type we have already discussed, *Achaete* and *Scute* are

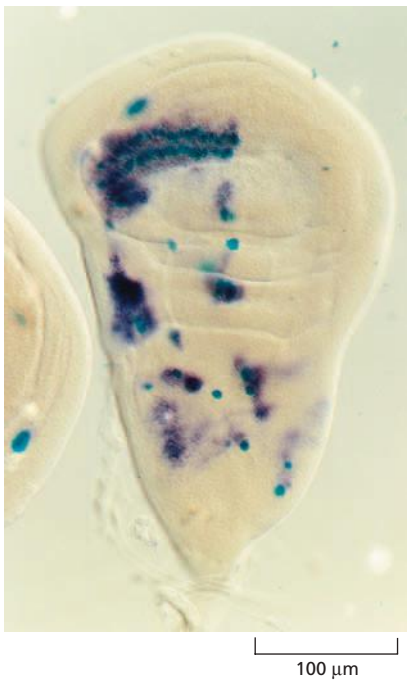


Figure 22–59 Sensory mother cells in the wing imaginal disc. The sensory mother cells (*blue* here) are easily revealed in this special strain of *Drosophila,* which contains an artificial *LacZ* reporter gene that, by chance, has inserted itself in the genome next to a control region that causes it to be expressed selectively in sensory mother cells. The *purple* stain shows the expression pattern of the *Scute* gene; this foreshadows the production of sensory mother cells and fades as the sensory mother cells successively develop. (From P. Cubas et al., *Genes Dev.* 5:996–1008, 1991. With permission from Cold Spring Harbor Laboratory Press.)

expressed in the imaginal disc in the regions within which bristles will form. Mutations that eliminate the expression of these genes at some of their usual sites block development of bristles at just those sites, and mutations that cause expression in additional, abnormal sites cause bristles to develop there. But expression of *Achaete* and *Scute* is transient, and only a minority of the cells initially expressing the genes go on to become sensory mother cells; the others become ordinary epidermis. The state that is specified by expression of *Achaete* and *Scute* is called *proneural*, and *Achaete* and *Scute* are called **proneural genes**. The proneural cells are primed to take the neurosensory pathway of differentiation, but, as we shall see, which of them will actually do so depends on competitive interactions among them.

Lateral Inhibition Singles Out Sensory Mother Cells Within Proneural Clusters

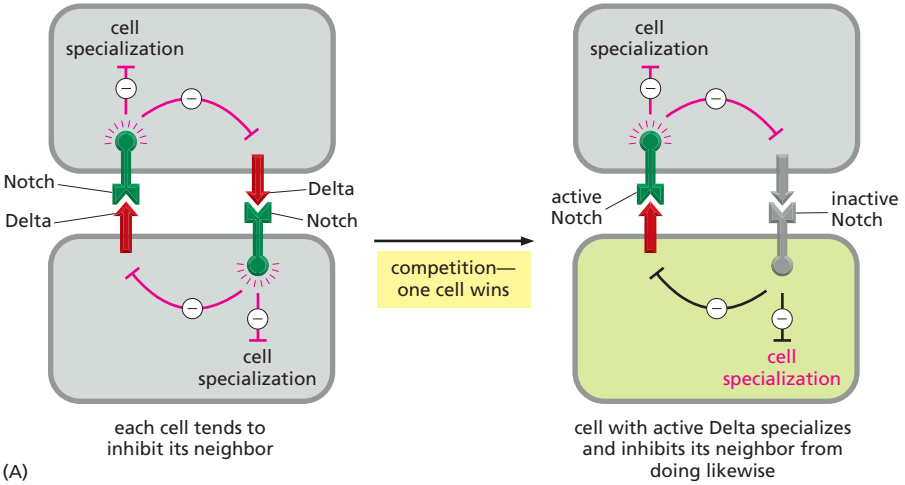
Cells expressing the proneural genes occur in groups in the imaginal disc epithelium—a small, isolated cluster of fewer than 30 cells for a big isolated bristle, a broad, continuous patch of hundreds or thousands of cells for a field of small bristles. In the former case just one member of the cluster becomes a sensory mother cell; in the latter case many cells scattered throughout the proneural region do so. In either case, each sensory mother cell becomes surrounded by cells that switch off expression of the proneural genes and become condemned to differentiate as epidermis instead. Experiments with genetic mosaics show that this is because a cell that becomes committed to the sensory-mother-cell pathway of differentiation sends a signal to its neighbors not to do the same thing: it exerts a *lateral inhibition*. If a cell that would normally become a sensory mother is genetically disabled from doing so, a neighboring proneural cell, freed from lateral inhibition, will become a sensory mother cell instead.

The lateral inhibition is mediated by the Notch signaling pathway. The cells in the cluster initially all express both the transmembrane receptor Notch and its transmembrane ligand Delta. Wherever Delta activates Notch, an inhibitory signal is sent into the Notch-expressing cell; consequently, all the cells in the cluster initially inhibit one another. However, receipt of the signal in a given cell is thought to diminish not only that cell's tendency to specialize as a sensory mother cell but also its ability to fight back by delivering the inhibitory Delta signal in return. This creates a competitive situation, from which a single cell in each small region—the future sensory mother cell—emerges as winner, sending a strong inhibitory signal to its immediate neighbors but receiving no such signal in return (**Figure 22–60**). The consequences of a failure of this regulatory mechanism are shown in **Figure 22–61**.

Lateral Inhibition Drives the Progeny of the Sensory Mother Cell Toward Different Final Fates

The same lateral inhibition mechanism dependent on Notch operates repeatedly in the formation of bristles—not only to force the neighbors of sensory mother cells to follow a different pathway and become epidermal, and but also later to make the daughters, the granddaughters, and finally the great-granddaughters of the sensory mother cell express different genes so as to form the different components of the bristle. At each stage, lateral inhibition mediates a competitive interaction that forces adjacent cells to behave in contrasting ways. Using a temperature-sensitive Notch mutation, it is possible to switch off Notch signaling after the sensory mother cell has been singled out but before it has divided. The progeny then differentiate alike, giving a cluster of neurons in place of the four different cell types of a bristle.

Like many other competitions, those mediated by lateral inhibition are often rigged: one cell starts with an advantage that guarantees it will be the winner. In the development of the different cell types of the sensory bristle, a strong initial



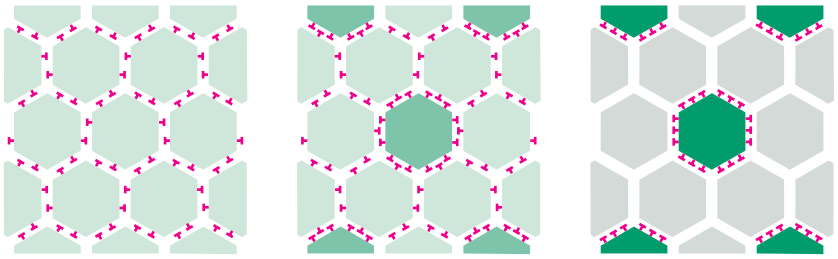


Figure 22–60 Lateral inhibition. (A) The basic mechanism of Notch-mediated competitive lateral inhibition, illustrated for just two interacting cells. In this diagram, the absence of color on proteins or effector lines indicates inactivity. (B) The outcome of the same process operating in a larger patch of cells. At first, all cells in the patch are equivalent, expressing both the transmembrane receptor Notch and its transmembrane ligand Delta. Each cell has a tendency to specialize (as a sensory mother cell), and each sends an inhibitory signal to its neighbors to discourage them from also specializing in that way. This creates a competitive situation. As soon as an individual cell gains any advantage in the competition, that advantage becomes magnified. The winning cell, as it becomes more strongly committed to differentiating as a sensory mother, also inhibits its neighbors more strongly. Conversely, as these neighbors lose their capacity to differentiate as sensory mothers they also lose their capacity to inhibit other cells from doing so. Lateral inhibition thus makes adjacent cells follow different fates. Although the interaction is thought to be normally dependent on cell–cell contacts, the future sensory mother cell may be able to deliver an inhibitory signal to cells that are more than one cell diameter away—for example, by sending out long protrusions to touch them.

bias is provided by an asymmetry in each of the cell divisions of the sensory mother cell and its progeny. A protein called Numb (together with certain other proteins) becomes localized at one end of the dividing cell, so that one daughter inherits the Numb protein and the other does not (**Figure 22–62**). Numb blocks the activity of Notch. Thus, the Numb-containing cell is deaf to inhibitory signals from its neighbors while its sister remains sensitive. Since both cells initially express the Notch ligand Delta, the cell that has inherited Numb proceeds to become neural, while driving its sister toward a nonneural fate.

Planar Polarity of Asymmetric Divisions is Controlled by Signaling via the Receptor Frizzled

For the Numb mechanism to operate, there must be machinery in the dividing cell to segregate the determinant to one side of the cell before division. In addition, as the cell enters mitosis the mitotic spindle must be aligned with this asymmetry so that the determinant is allocated to just one daughter cell, and not shared out to both daughters at the time of cell division. In the above case, the sensory mother cell, at its first division, regularly divides to give an anterior cell that inherits Numb and a posterior cell that does not. As discussed in Chapter 19, this type of polarity in the plane of the epithelium is called *planar polarity* (in



Figure 22–61 The result of switching off lateral inhibition during the singling-out of sensory mother cells. The photograph shows part of the thorax of a fly containing a mutant patch in which the neurogenic gene *Delta* has been partially inactivated. The reduction of lateral inhibition has caused almost all the cells in the mutant patch (in the center of the picture) to develop as sensory mother cells, producing a great excess of sensory bristles there. Mutant patches of cells carrying more extreme mutations in the Notch pathway, causing a total loss of lateral inhibition, form no visible bristles because all of the progeny of the sensory mother cells develop as neurons or glial cells instead of diversifying to form both neurons and the external parts of the bristle structure. (Courtesy of P. Heitzler and P. Simpson, *Cell* 64:1083–1093, 1991. With permission from Elsevier.)

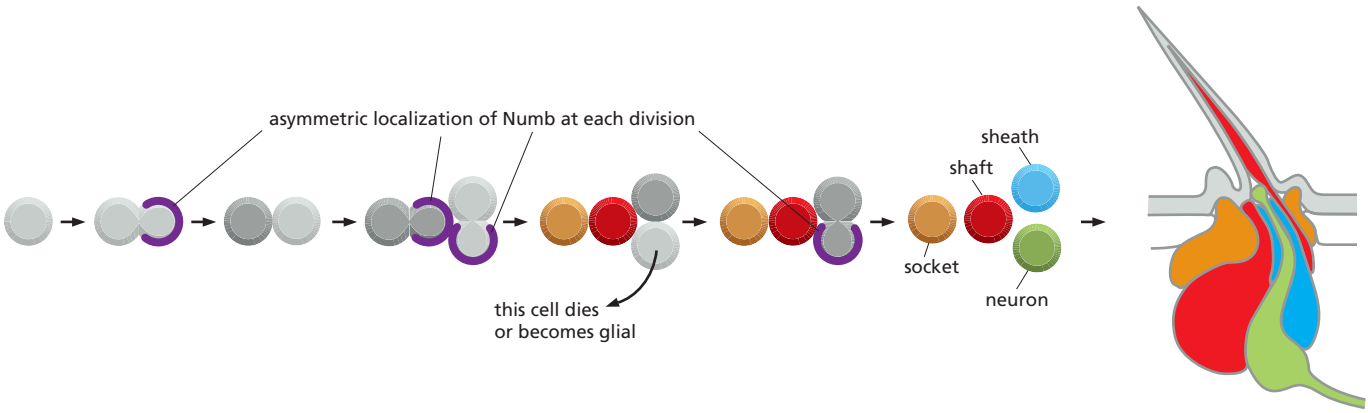


Figure 22–62 Numb biases lateral inhibition during bristle development. At each division of the progeny of the sensory mother cell, Numb protein is asymmetrically localized, producing daughter cells that differ. Note that some of the divisions are oriented with the mitotic spindle in the plane of the epithelium, others at right angles to it; the localization of Numb is controlled in different ways at these different types of division but plays a critical role at each of them in deciding cell fate. (Based on data from M. Gho, Y. Bellaiche and F. Schweisguth, *Development* 126:3573–3584, 1999. With permission from The Company of Biologists.)

contradistinction to apico-basal polarity, where the cellular asymmetry is perpendicular to the plane of the epithelium). It is manifested in the uniformly backward-pointing orientation of the bristles, giving the fly its wind-swept appearance (**Figure 22–63**).

The planar polarity in the initial division of the sensory mother cell is controlled by a signaling pathway similar to the one that we encountered controlling asymmetric divisions in the nematode (see Figure 22–21), depending on the receptor Frizzled. Frizzled proteins have been discussed in Chapter 15 as receptors for Wnt proteins, but in the control of planar polarity—in flies and probably in vertebrates too—this pathway functions in a special way: the intracellular relay mechanism exerts its main effects on the actin cytoskeleton, rather than on gene expression. The intracellular protein Dishevelled, downstream from Frizzled, is common to the gene-regulatory and the actin-regulatory branches of the signaling pathway. Separate domains of the Dishevelled molecule can be shown to be responsible for the two functions (**Figure 22–64**). Frizzled and Dishevelled both take their names from the unkempt look of flies where bristle polarity is disordered (see Figure 19–32).

Asymmetric Stem-Cell Divisions Generate Additional Neurons in the Central Nervous System

The mechanisms we have described for controlling the genesis of neurons of sensory bristles operate also, with variations, in the genesis of virtually all other neurons—not only in insects, but also in other phyla. Thus in the embryonic central nervous system, both in flies and in vertebrates, neurons are generated from regions of expression of proneural genes akin to *Achaete* and *Scute*. The nascent neurons express Delta and inhibit their immediate neighbors, which express Notch, from becoming committed to neural differentiation at the same time. When Notch signaling is blocked, inhibition fails, and in the proneural regions neurons are generated in huge excess at the expense of non-neuronal cells (**Figure 22–65**).

In the central nervous system, however, an additional mechanism comes into play to help generate the very large numbers of neurons and glial cells that are needed: a special class of cells become committed as neural precursors, but instead of differentiating directly as neurons or glial cells, these undergo a long series of asymmetric divisions through which a succession of additional neurons and glial cells are added to the population. The mechanism is best understood in *Drosophila*, although there are many hints that something similar occurs also in vertebrate neurogenesis.

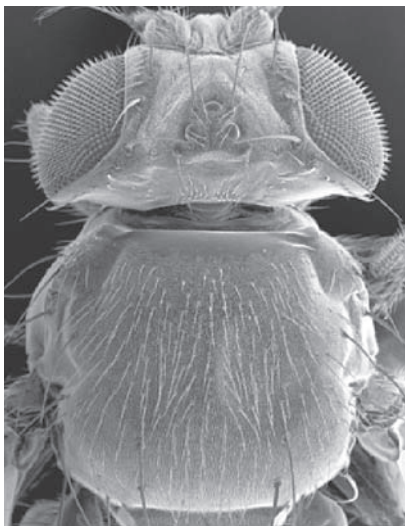
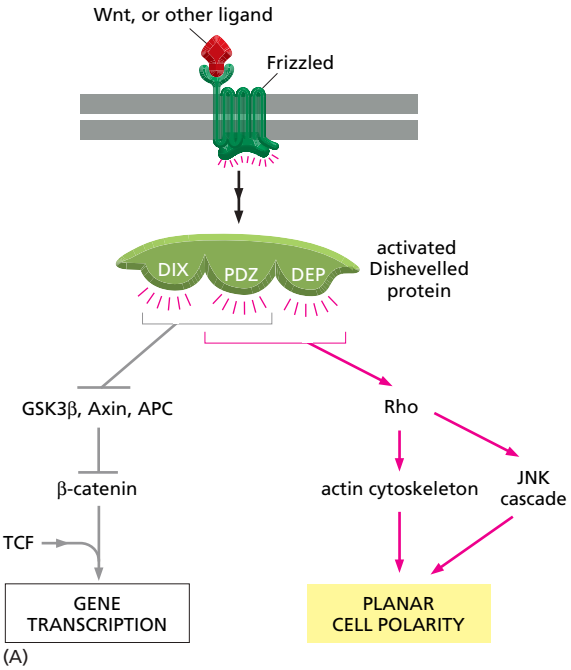


Figure 22–63 Planar cell polarity manifest in bristle polarity on a fly's back: the bristles all point backwards. (Scanning electron micrograph courtesy of S. Oldham and E. Hafen, from E. Spana and N. Perrimon, *Trends Genet.* 15:301–302, 1999. With permission from Elsevier.)



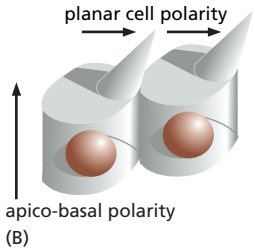


Figure 22–64 The control of planar cell polarity. (A) The two branches of the Wnt/Frizzled signaling pathway. The main branch, discussed in Chapter 15, controls gene expression via β-catenin; the planar-polarity branch controls the actin cytoskeleton via Rho GTPases. Different domains of the Dishevelled protein are responsible for the two effects. It is not yet clear which member of the Wnt signal protein family, if any, is responsible for activating the planar polarity function of Frizzled in *Drosophila*. (B) Cartoon of cells displaying planar polarity. In at least some systems, planar cell polarity is associated with asymmetric localization of the receptor Frizzled itself to one side of each cell. (See also Chapter 19, Figure 19–32.)

In the embryonic central nervous system of *Drosophila*, the nerve-cell precursors, or **neuroblasts**, are initially singled out from the neurogenic ectoderm by a typical lateral-inhibition mechanism that depends on Notch. Each neuroblast then divides repeatedly in an asymmetric fashion (**Figure 22–66**A). At each division, one daughter remains as a neuroblast, while the other, which is much smaller, becomes specialized as a *ganglion mother cell*, or *GMC*. The ganglion mother cell will divide only once, giving a pair of neurons, or a neuron plus a glial cell, or a pair of glial cells. The neuroblast becomes smaller at each division, as it parcels out its substance into one ganglion mother cell after another. Eventually, typically after about 12 cycles, the process halts, presumably because the neuroblast becomes too small to pass the cell-size checkpoint in the cell division cycle. Later, in the larva, neuroblast divisions resume, and now they are accompanied by cell growth, permitting the process to continue indefinitely, generating the much larger numbers of neurons and glial cells required in the adult fly.

The larval neuroblasts, therefore, are **stem cells**: while not terminally differentiated themselves, they behave as a self-renewing and potentially inexhaustible source of terminally differentiated cells. In Chapter 23, where we discuss stem cells in detail, we shall see that stem cells do not necessarily have to divide asymmetrically; but asymmetric division is one possible strategy, and the neuroblasts of the fly provide a beautiful example.

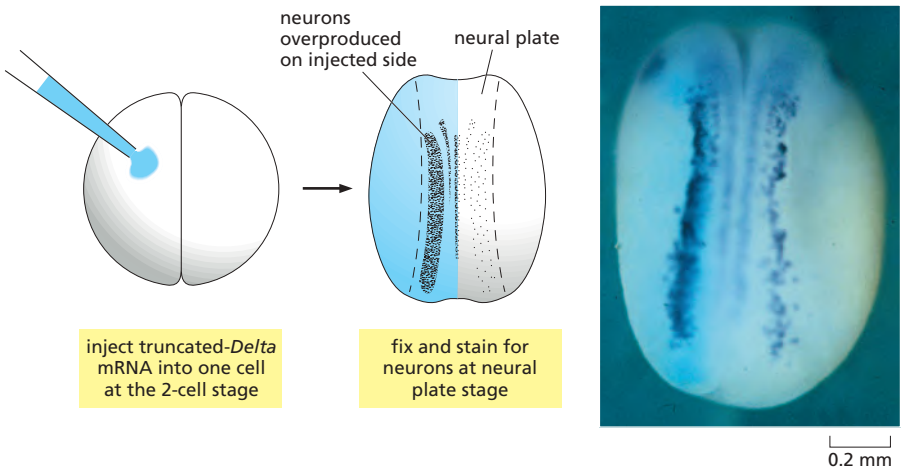
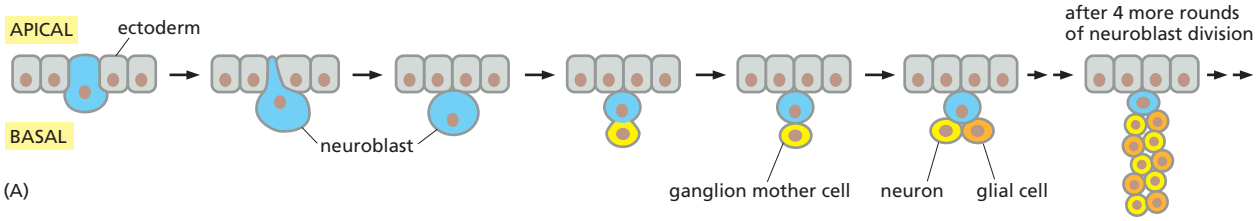


Figure 22–65 Effects of blocking Notch signaling in a *Xenopus* embryo. In the experiment shown, mRNA coding for a truncated form of the Notch ligand *Delta* is injected, together with *LacZ* mRNA as a marker, into one cell of an embryo at the two-cell stage. The truncated Delta protein produced from the mRNA blocks Notch signaling in the cells descended from the cell that received the injection. These cells lie on the left side of the embryo and are identifiable because they contain LacZ protein (*blue* stain) as well as the truncated Delta protein. The right side of the embryo is unaffected and serves as a control. The embryo is fixed and stained at a stage when the central nervous system has not yet rolled up to form a neural tube, but is still a more or less flat plate of cells—the neural plate—exposed on the surface of the embryo. The first neurons (stained *purple* in the photograph) have already begun to differentiate in elongated bands (proneural regions) on each side of the midline. On the control (right) side, they are a scattered subset of the proneural cell population. On the Notch-blocked (left) side, virtually all the cells in the proneural regions have differentiated as neurons, creating a densely stained band of neurons without intervening cells. Injections of mRNA coding for normal, functional Delta have an opposite effect, reducing the number of cells that differentiate as neurons. (Photograph from A. Chitnis et al., *Nature* 375:761–766, 1995. With permission from Macmillan Publishers Ltd.)



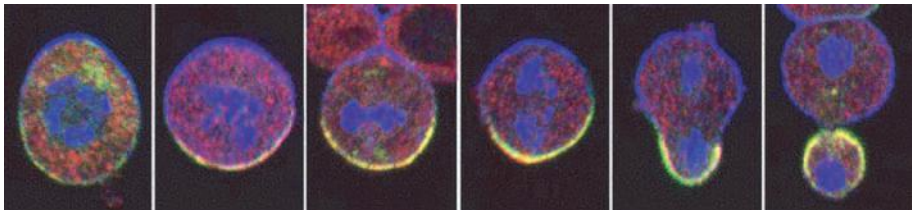


Figure 22–66 Neuroblasts and asymmetric cell division in the central nervous system of a fly embryo. (A) The neuroblast originates as a specialized ectodermal cell. It is singled out by lateral inhibition and emerges from the basal (internal) face of the ectoderm. It then goes through repeated division cycles, dividing asymmetrically to generate a series of ganglion mother cells. Each ganglion mother cell divides just once to give a pair of differentiated daughters (typically a neuron plus a glial cell). (B) The asymmetric distribution of cell fate determinants in an isolated neuroblast as it goes through mitosis. The mitotic chromosomes are stained *blue*. The Par3/Par6/aPKC complex, shown by *blue* immunostaining for aPKC, is concentrated in the apical cortex, and causes Miranda (*green*), Brat (*red,* giving *yellow* where Brat and Miranda overlap) and Prospero (not stained) to become localized in the basal cortex. As the cell divides, these latter three molecules become segregated into the ganglion mother cell, forcing it to differentiate and leaving the neuroblast free to regenerate its asymmetry and divide again in the same way. (B, from C.Y. Lee et al., *Dev. Cell* 10:441–449, 2006. With permission from Elsevier.)

Asymmetric Neuroblast Divisions Segregate an Inhibitor of Cell Division into Just One of the Daughter Cells

The divisions of the neuroblast are asymmetric in three respects: (1) physically, in that one daughter is smaller than the other; (2) biochemically, in factors controlling differentiation; and (3) biochemically, in factors controlling proliferation. These asymmetries must all be coordinated with one another and with the orientation of the axis of the mitotic spindle, if the cleavage plane is to cut the cell into the correct parts. How is this achieved?

The neuroblast has an apico-basal asymmetry that reflects its origin from the ectoderm, which, like other epithelia, has a well defined apico-basal polarity. As we saw in Chapter 19, apico-basal polarity is governed by a complex of three proteins—Par3 (also called Bazooka in *Drosophila),* Par6, and aPKC (atypical protein kinase C)—that become localized in the cortex toward the apical end of the cell. This localized Par3/Par6/aPKC complex is thought to be the primary source of asymmetry in the neuroblast. By recruiting other components, some of which exert feedback effects to maintain the localization of the complex, it coordinates the whole process of unequal division.

The Par3/Par6/aPKC complex defines the orientation of the mitotic spindle and the unequal partitioning of the cell at cytokinesis through interaction with adapter proteins called Inscuteable and Partner of Inscuteable (Pins). These in turn recruit the α subunit of a trimeric G protein (discussed in Chapter 15), which functions in this context as an intracellular messenger to guide organization of the cytoskeleton.

At the same time, the Par3/Par6/aPKC complex locally phosphorylates a regulator of intracellular architecture called Lgl (Lethal giant larvae) and thereby directs another adaptor protein called Miranda to become concentrated in the cortex at the opposite (basal) end of the cell (Figure 22–66B). Miranda binds proteins that control differentiation and proliferation, localizing them to the same site. When the neuroblast divides, Miranda and its cargo are segregated into the ganglion mother cell. One of the molecules thus carried into the ganglion mother cell is a gene regulatory protein called Prospero, which directs differentiation. Another is a posttranscriptional repressor called Brat (Brain Tumor). Brat acts as an inhibitor of cell proliferation, apparently by preventing production of the growth-promoting protein Myc, famous for its role in cancer (discussed in Chapter 20). In mutants where Brat is defective, or where it fails to become localized correctly, the smaller daughter cell of the asymmetric neuroblast division frequently fails to differentiate as a ganglion mother cell, and instead grows and divides as a neuroblast. The result is a brain tumor—a mass of neuroblasts that grows exponentially and without limit, until the fly is dead.

Whether vertebrates tissues have stem cells that behave like the fly's neuroblasts is a question of great current interest, especially in relation to cancer.

Notch Signaling Regulates the Fine-Grained Pattern of Differentiated Cell Types in Many Different Tissues

Each daughter of a normal ganglion mother cell can become either a neuron or a glial cell. This final choice, like the choice of cell fate for the progeny of a sensory mother cell in the peripheral nervous system, is controlled by Notch signaling and lateral inhibition. In fact, lateral inhibition mediated by Notch is crucial for cell diversification and fine-grained patterning in an enormous variety of different tissues. In the fly, it controls the production not only of neurons but also of many other differentiated cell types—for example, in muscle, in the lining of the gut, in the excretory system, in the tracheae, and in the eye and other sense organs. In vertebrates, homologs of Notch and its ligands are expressed in the corresponding tissues and have similar functions: mutations in the Notch pathway upset the balance not only of neurons and non-neuronal cells in the central nervous system, but also of the different specialized cell types in the lining of the gut, of endocrine and exocrine cells in the pancreas, and of sensory and supporting cells in sense organs such as the ear, to give only a few examples.

In all these tissues, a balanced mixture of different cell types is required. Notch signaling provides the means to generate the mixture, by enabling individual cells expressing one set of genes to direct their immediate neighbors to express another set.

Some Key Regulatory Genes Define a Cell Type; Others Can Activate the Program for Creation of an Entire Organ

As we mentioned at the beginning of this chapter, there are some genes whose products act as triggers for the development of a specific organ, initiating and coordinating the whole complex program of gene expression required for this. Thus, for example, when the *Eyeless* gene is artificially expressed in a patch of cells in the leg imaginal disc, a patch of well-organized eye tissue, with all its various cell types correctly arranged, will develop on the leg (see Figure 22–2). In a somewhat similar way, much later, when a cell makes its final choice of a particular mode of differentiation in the aftermath of the interactions mediated by Notch, it has to follow a complex program involving expression of a whole collection of genes, and this differentiation program is initiated and coordinated by a much smaller set of high-level regulators. Such regulators are sometimes called "master regulatory proteins" (though even they can exert their specific effect only in combination with the right partners, in a cell that is adequately primed).

An example is the MyoD/myogenin family of gene regulatory proteins. These proteins drive cells to differentiate as muscle, expressing muscle-specific actins and myosins and all the other cytoskeletal, metabolic and membrane proteins that a muscle cell needs (see Figure 7–75). The gene regulatory proteins that define particular cell types often belong (as do MyoD and its relatives) to the basic helix–loop–helix family, encoded by genes homologous to, and in some cases apparently identical to, the proneural genes that we have already mentioned. Their expression is often governed by the Notch pathway via complicated feedback loops.

Terminal cell differentiation has brought us to the end of our sketch of how genes control the making of a fly. Our account has necessarily been simplified. Many more genes than we have mentioned are involved in each of the developmental processes that we have described. Feedback loops, alternative mechanisms operating in parallel, genetic redundancy, and other phenomena complicate the full picture. Despite all this, the overriding message of developmental genetics is one of an unexpected simplicity. A limited number of genes and mechanisms, used repeatedly in different circumstances and combinations, are responsible for controlling the main features of the development of all multicellular animals.

We next turn to an essential aspect of animal development that we have so far neglected: cell movements.

Summary

The external parts of an adult fly develop from epithelial structures called imaginal discs. Each imaginal disc is divided at the outset into a small number of domains expressing different gene regulatory proteins as a result of early embryonic patterning processes. These domains are called compartments, because their cells do not mix. At the compartment boundaries, cells expressing different genes confront one another and interact, inducing localized production of morphogens that govern the further growth and internal patterning of each compartment. Thus, in the wing disc, dorsal and ventral cells interact by the Notch signaling mechanism to create a source of Wingless (Wnt) protein along the dorsoventral compartment boundary, while anterior and posterior cells interact through short-range Hedgehog signaling to create a source of Dpp protein (a TGFβ family member) along the anteroposterior compartment boundary. All these signaling molecules have homologs that play similar parts in limb patterning in vertebrates.

Each compartment of an imaginal disc, and each substructure within it, grows to a precisely predictable size, even in the face of seemingly drastic disturbances, such as mutations that alter the cell division rate. Although the morphogen gradients in the disc are clearly involved, the critical regulatory mechanisms that control organ size are not understood.

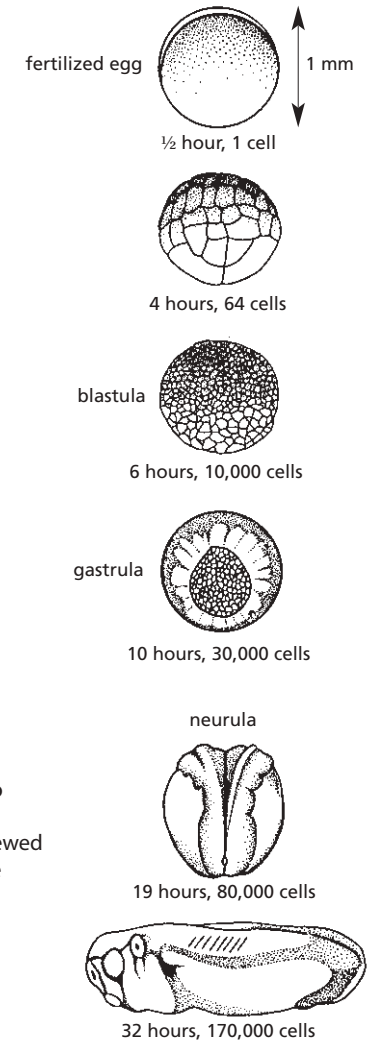
Within each compartment, the morphogen gradients control the sites of expression of further sets of genes, defining patches of cells that interact with one another yet again to create the finest details of the ultimate pattern of cell differentiation. Thus, proneural gene expression defines the sites where sensory bristles will form, and Notch-mediated interactions among the cells of the proneural cluster, together with asymmetric cell divisions, force the individual cells of the bristle to follow different paths of terminal differentiation. In the central nervous system, neuroblasts are singled out from the ectoderm by lateral inhibition in a similar way, but then go through a long series of asymmetrical divisions as stem cells to generate neurons and glia. Faults in the asymmetric distribution of the molecules that control differentiation and proliferation can convert the neuroblast stem cells into tumor cells.

Many of the same mechanisms are thought to operate in vertebrate tissues also.

CELL MOVEMENTS AND THE SHAPING OF THE VERTEBRATE BODY

Most cells of the animal body are motile, and in the developing embryo their movements are often extensive, dramatic, and surprising. <GAGC> Controlled changes of gene expression create ordered arrays of cells in different states; cell movements rearrange these cellular building blocks and put them in their proper places. The genes that the cells express determine how they move; in this sense, the control of gene expression is the primary phenomenon. But the cell movements are also crucial, and no less in need of explanation if we want to understand how the architecture of the body is created. In this section, we examine this topic in the context of vertebrate development. We take as our main example the frog *Xenopus laevis* (**Figure 22–67**), where cell movements have been well studied, though we shall also draw on evidence from chick, zebrafish, and mouse.





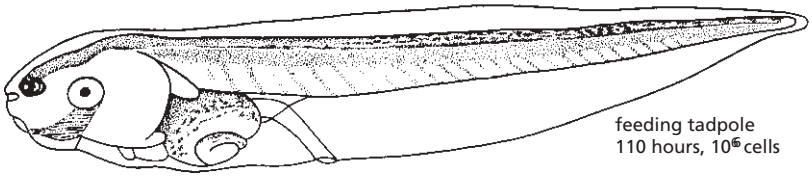
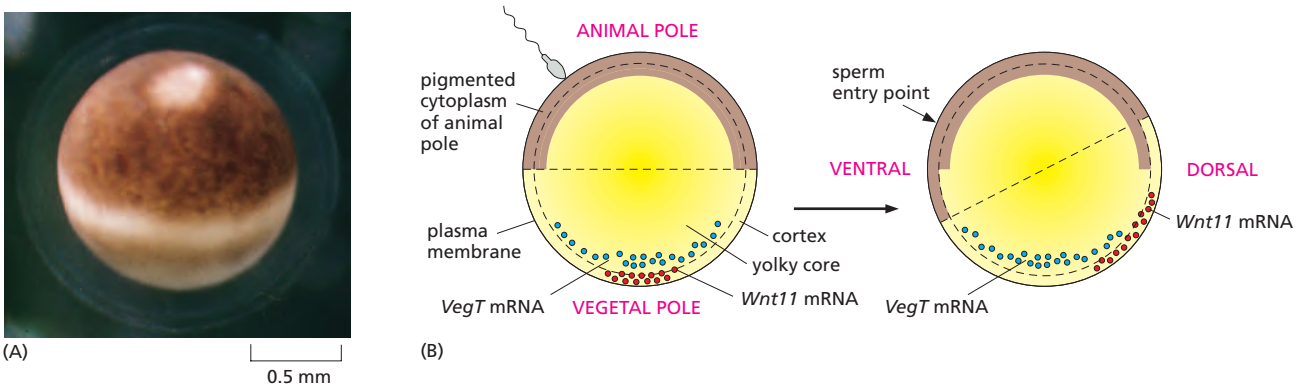


Figure 22–67 Synopsis of the development of *Xenopus laevis* from newly fertilized egg to feeding tadpole. The adult frog is shown in the photograph at the top. The developmental stages are viewed from the side, except for the 10-hour and 19-hour embryos, which are viewed from below and from above, respectively. All stages except the adult are shown at the same scale. (Photograph courtesy of Jonathan Slack; drawings after P.D. Nieuwkoop and J. Faber, Normal Table of *Xenopus laevis* [Daudin]. Amsterdam: North-Holland, 1956.)



The Polarity of the Amphibian Embryo Depends on the Polarity of the Egg

The *Xenopus* egg is a large cell, just over a millimeter in diameter (**Figure 22–68**A). The light-colored lower end of the egg is called the vegetal pole; the dark-colored upper end is called the animal pole. The animal and vegetal hemispheres contain different selections of mRNA molecules and other cell components, which become allocated to separate cells as the egg cell divides after fertilization. Near the vegetal pole, for example, there is an accumulation of mRNAs coding for the gene regulatory protein VegT (a DNA-binding protein of the T-box family) and for signal proteins of the TGFβ superfamily, as well as some ready-made protein components of the Wnt signaling pathway (Figure 22–68B). As a result, the cells that inherit vegetal cytoplasm will produce signals to organize the behavior of adjacent cells. They are committed to form the gut—the innermost tissue of the body; the cells that inherit animal cytoplasm will form the outer tissues. Thus, crudely speaking, the animal–vegetal axis of the egg corresponds to the external-to-internal (or skin-to-gut) dimension of the future organism.

Fertilization initiates a series of cell divisions and movements that will eventually tuck the vegetal cells and cells from the equatorial (middle) region of the animal–vegetal axis into the interior. In the course of these complex movements, the three principal axes of the body become established: *anteroposterior,* from head to tail; *dorsoventral,* from back to belly; and *mediolateral,* from the midline outward to the left or to the right. The orientation of these axes is determined by the asymmetries of the early embryo. The unfertilized egg, has only one axis of asymmetry—the animal–vegetal—but fertilization triggers an intracellular movement that gives the egg an additional asymmetry defining a second axis at right angles to this. Following entry of the sperm, the outer, actin-rich cortex of the egg cytoplasm rotates relative to the central core of the egg, so that the animal pole of the cortex is slightly shifted to one side. Treatments that block the rotation allow cleavage to occur normally but produce an embryo with a central gut and no dorsal structures or dorsoventral asymmetry. Thus, the cortical rotation is required to define the dorso–ventral axis of the future body, and the axis of asymmetry created in the egg by the rotation is called the dorso–ventral axis of the egg. Note, however, that the subsequent cell movements mean that the relationship between the egg axes and the future body axes is more complicated than this terminology would suggest. The direction of the cortical rotation is biased according to the point of sperm entry, perhaps through the centrosome that the sperm brings into the egg, and the movement is associated with a reorganization of microtubules in the egg cytoplasm. This leads to a microtubule-based transport of several components, including mRNA coding for Wnt11, a member of the Wnt family of signal molecules, toward the future dorsal side (see Figure 22–68B). This mRNA is soon translated, producing Wnt11 protein in the dorsal vegetal region. The Wnt11 secreted from cells that form in that region is crucial in triggering the cascade of subsequent events that will organise the dorsoventral axis of the body.

Figure 22–68 The *Xenopus* egg and its asymmetries. (A) Side view of an egg photographed just before fertilization. (B) The asymmetric distribution of molecules inside the egg, and how this changes following fertilization so as to define a dorsoventral as well as an animal–vegetal asymmetry. Fertilization, through a reorganization of the microtubule cytoskeleton, triggers a rotation of the egg cortex (a layer a few μm deep) through about 30° relative to the core of the egg in a direction determined by the site of sperm entry. Some components are carried still further to the future dorsal side by active transport along microtubules. The resulting dorsal concentration of *Wnt11* mRNA leads to dorsal production of the Wnt11 signal protein and defines the dorsoventral polarity of the future embryo. (A, courtesy of Tony Mills.)

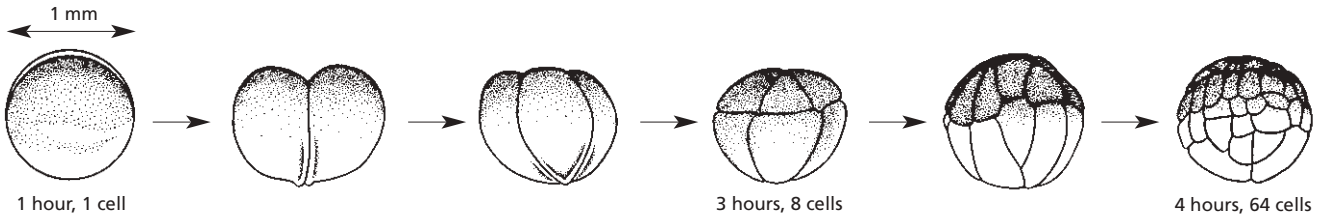


Figure 22–69 The stages of cleavage in *Xenopus*. The cleavage divisions rapidly subdivide the egg into many smaller cells. All the cells divide synchronously for the first 12 cleavages, but the divisions are asymmetric, so that the lower, vegetal cells, encumbered with yolk, are fewer and larger.

Cleavage Produces Many Cells from One

The cortical rotation is completed in about an hour after fertilization and is followed by cleavage, in which the single large egg cell rapidly subdivides by repeated mitosis into many smaller cells, or *blastomeres*, without any change in total mass (**Figure 22–69**). <ATTT> In this way, the determinants distributed asymmetrically in the egg become partitioned into separate cells, with different fates (**Figure 22–70**).

These first cell divisions in *Xenopus* have a cycle time of about 30 minutes, with a direct alternation of S and M phases, as discussed in Chapter 17. The very high rate of DNA replication and mitosis seems to preclude almost all gene transcription (although protein synthesis occurs), and the cleaving embryo is almost entirely dependent on reserves of RNA, protein, membrane, and other materials that accumulated in the egg while it developed as an oocyte in the mother. After about 12 cycles of cleavage (7 hours), the cell division rate slows down, the cell cycles begin to follow the standard pattern with G_1 and G_2 phases intervening between the S and M phases, and widespread transcription of the embryo's genome begins. This event is called the *mid-blastula transition*, and it occurs with roughly similar timing in most animal species (mammals being an exception). Studies in zebrafish show that the newly synthesized transcripts include micro-RNAs that recognize many of the transcripts deposited in the egg by the mother and direct their rapid degradation. The midblastula transition thus marks the point at which the embryo's own genome largely takes over control of development.

Gastrulation Transforms a Hollow Ball of Cells into a Three-Layered Structure with a Primitive Gut

During the period of cleavage, the frog embryo becomes transformed from a solid sphere of cells into something more like a hollow ball, with an internal fluid-filled cavity surrounded by cells that cohere to form an epithelial sheet. The embryo is now termed a **blastula** (**Figure 22–71**).

Soon after this, the coordinated movements of gastrulation begin. <TCCC> This dramatic process transforms the simple hollow ball of cells into a multilayered structure with a central gut tube and bilateral symmetry: by a more elaborate version of the process outlined earlier for the sea urchin (see Figure 22–3), many of the cells on the outside of the embryo are moved inside it. Subsequent development depends on the interactions of the inner, outer, and middle layers of cells thus formed: the *endoderm* on the inside, consisting of the cells that have moved into the interior to form the primitive gut; the *ectoderm* on the outside, consisting of cells that have remained external; and the *mesoderm* between

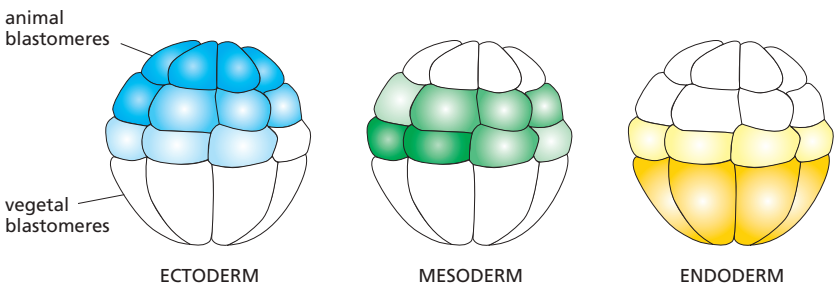


Figure 22–70 The origins of the three germ layers can be traced back to distinct blastomeres of the embryo in its early cleavage stages. The endoderm derives from the most vegetal blastomeres, the ectoderm from the most animal, and the mesoderm from a middle set that contribute also to endoderm and ectoderm. The coloring in each picture is the more intense, the higher the proportion of cell progeny that will contribute to the given germ layer. (After L. Dale, *Curr. Biol.* 9:R812–R815, 1999. With permission from Elsevier.)

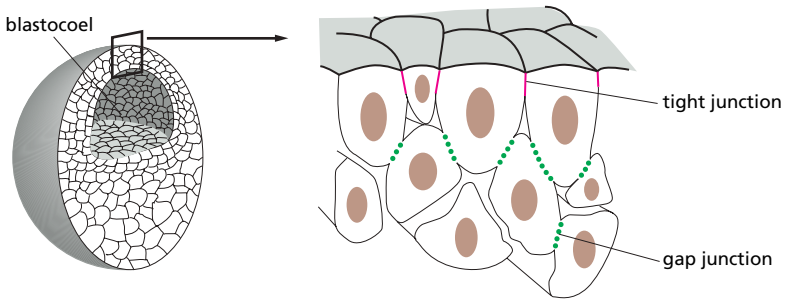


Figure 22–71 The blastula. In the outermost regions of the embryo, tight junctions between the blastomeres begin to create an epithelial sheet that isolates the interior of the embryo from the external medium. Na^+ is pumped across this sheet into the spaces in the interior of the embryo, and water follows into these spaces because of the resulting osmotic pressure gradient. As a result, the intercellular crevices inside the embryo enlarge to form a single cavity, the blastocoel. In *Xenopus* the wall of the blastocoel is several cells thick, and only the outermost cells are tightly bound together as an epithelium.

them, consisting of cells that detach from the epithelium to form a more loosely organized embryonic connective tissue (**Figure 22–72**). From these three *germ layers,* the tissues of the adult vertebrate body will be generated, preserving the basic body plan established through gastrulation.

The Movements of Gastrulation Are Precisely Predictable

The pattern of gastrulation movements that creates the germ layers and establishes the body axes is described for *Xenopus* in **Figure 22–73**. The details are complex, but the principles are simple.

Cells of the future endoderm are folded into the interior, or *involuted*, in succession. The process begins with a downward movement of cells from the animal hemisphere to cover and enclose the yolky vegetal hemisphere, which represents the food supply of the embryo. Cells that are in the vanguard of this movement, at the vegetal margin of the advancing cell sheet, are the first to involute, turning inward and then moving up toward the animal pole to form the most anterior part of the gut. As they near the animal pole, these leading endoderm cells will signal to the overlying ectoderm to define the anterior extremity of the head. The mouth will eventually develop as a hole formed at an anterior site where endoderm and ectoderm come into direct contact. Meanwhile, future mesoderm cells, destined to detach from the epithelial sheet to form the sandwich filling between endoderm and ectoderm, tuck into the interior along with the endoderm cells, and also move up toward the animal pole. The cells that are first to involute go to form parts of the head, and those that are last form parts of the tail. In this way, the anteroposterior axis of the final embryo is laid down sequentially.

The anteroposterior movements go hand in hand with movements that organize the dorsoventral axis of the body. Gastrulation begins on the side of the blastula that has been marked out as dorsal by the cortical rotation. Here, involution of cells into the interior starts with a short indentation that rapidly extends to form the *blastopore*—a line of invagination that curves around to encircle the vegetal pole. The site where the invagination starts defines the *dorsal lip of the blastopore.* As we shall see, this tissue plays a leading part in subsequent events and gives rise to the central dorsal structures of the main body axis.

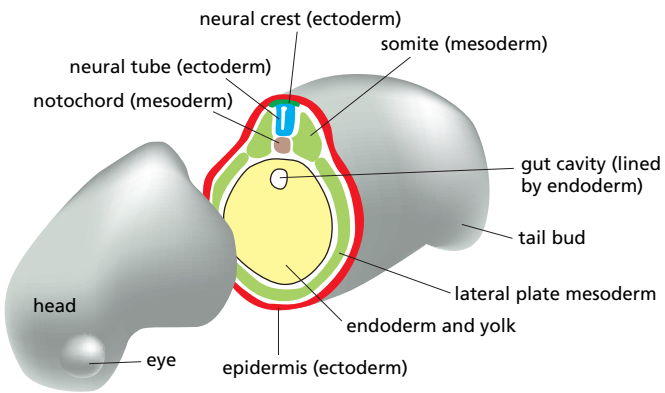


Figure 22–72 A cross section through the trunk of an amphibian embryo after the end of gastrulation, showing the arrangement of endodermal, mesodermal and ectodermal tissues. The endoderm will form the epithelial lining of the gut, from the mouth to the anus. It gives rise not only to the pharynx, esophagus, stomach, and intestines, but also to many associated glands. The salivary glands, the liver, the pancreas, the trachea, and the lungs, for example, all develop from extensions of the wall of the originally simple digestive tract and grow to become systems of branching tubes that open into the gut or pharynx. The endoderm forms only the epithelial components of these structures—the lining of the gut and the secretory cells of the pancreas, for example. The supporting muscular and fibrous elements arise from the mesoderm. The mesoderm gives rise to the connective tissues—at first to the loose, space-filling, three-dimensional mesh of cells in the embryo known as mesenchyme, and ultimately to cartilage, bone, and fibrous tissue, including the dermis (the inner layer of the skin). The mesoderm also forms the muscles, the entire vascular system—including the heart, the blood vessels, and the blood cells—and the tubules, ducts, and supporting tissues of the kidneys and gonads. The ectoderm will form the epidermis (the outer, epithelial layer of the skin) and epidermal appendages such as hair, sweat glands, and mammary glands. It will also give rise to the whole of the nervous system, central and peripheral, including not only neurons and glia but also the sensory cells of the nose, the ear, the eye, and other sense organs. (After T. Mohun et al., *Cell* 22:9–15, 1980. With permission from Elsevier.)

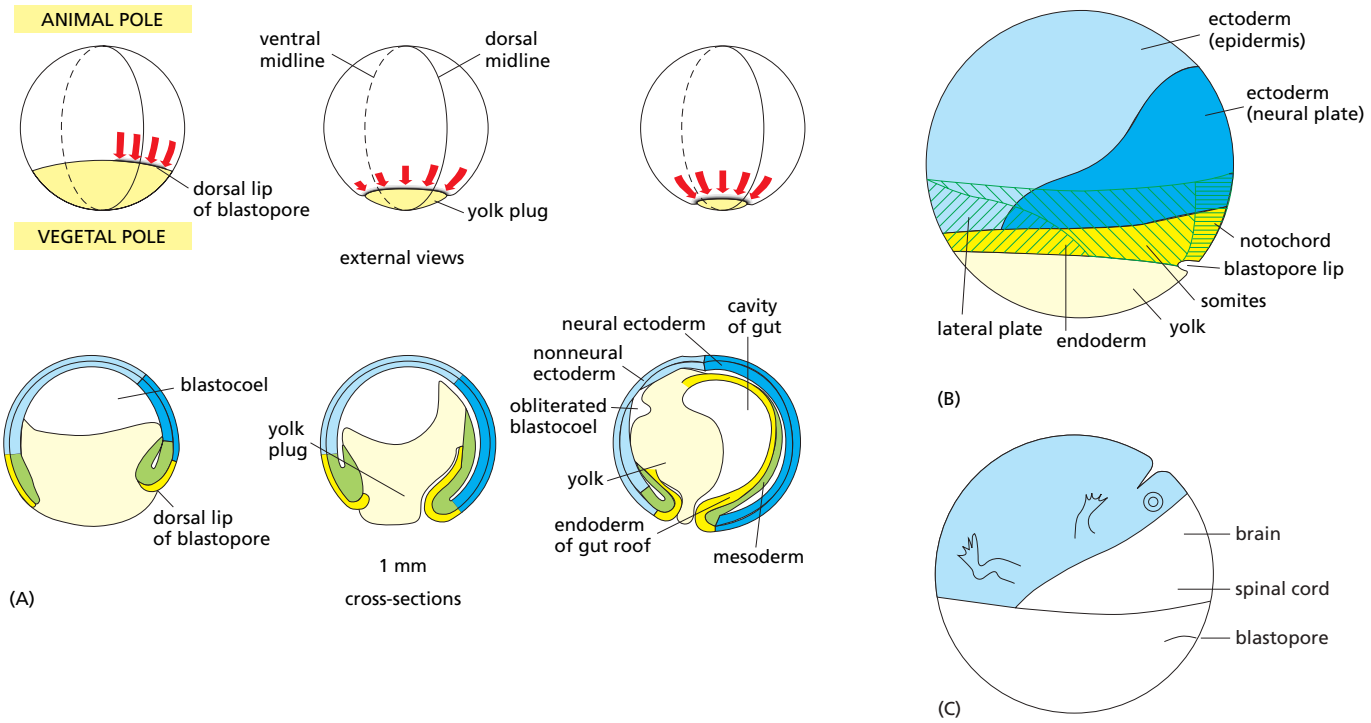
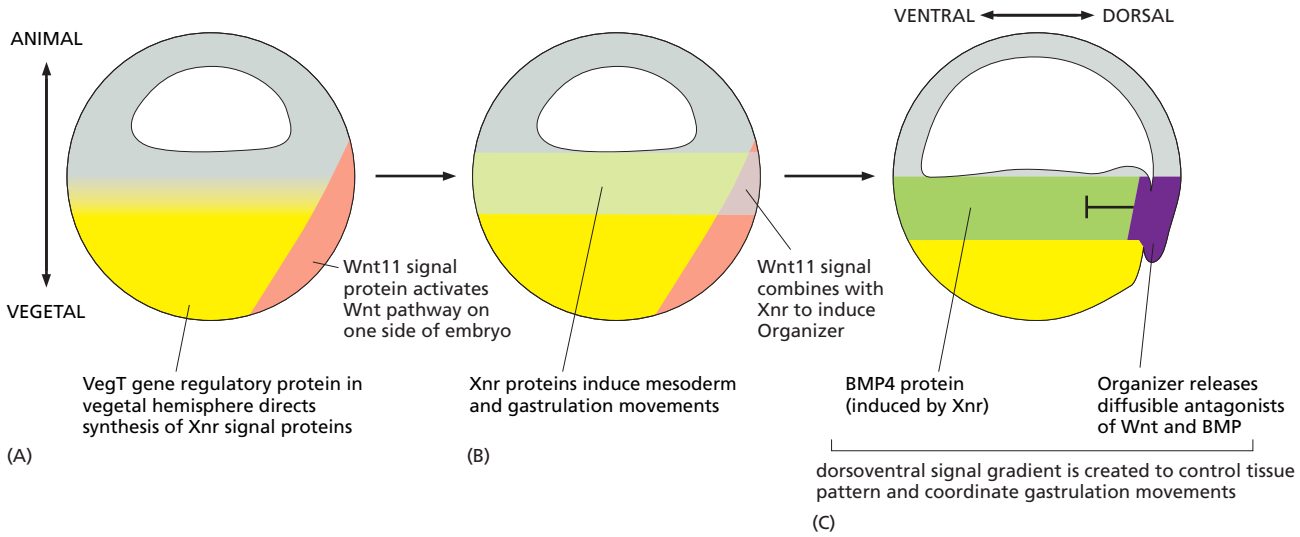


Figure 22–73 Gastrulation in *Xenopus*. <TCCC> (A) The external views *(above)* show the embryo as a semitransparent object, seen from the side; with the directions of cell movement indicated by *red arrows,* cross sections *(below)* are cut in the median plane (the plane of the dorsal and ventral midlines). Gastrulation begins when a short indentation, the beginning of the blastopore, becomes visible in the exterior of the blastula. This indentation gradually extends, curving around to form a complete circle surrounding a plug of very yolky cells (destined to be enclosed in the gut and digested). Sheets of cells meanwhile turn in around the lip of the blastopore and move deep into the interior of the embryo. At the same time the external epithelium in the region of the animal pole actively spreads to take the place of the cell sheets that have turned inward. Eventually, the epithelium of the animal hemisphere spreads in this way to cover the whole external surface of the embryo, and, as gastrulation reaches completion, the blastopore circle shrinks almost to a point. (B) A fate map for the early *Xenopus* embryo (viewed from the side) as it begins gastrulation, showing the origins of the cells that will come to form the three germ layers as a result of the movements of gastrulation. The various parts of the mesoderm (lateral plate, somites, and notochord) derive from deep-lying cells that segregate from the epithelium in the cross-hatched region. The other cells, including the more superficial cells in the cross-hatched region, will give rise to ectoderm *(blue, above)* or endoderm *(yellow, below)*. Roughly speaking, the first cells to turn into the interior, or involute, will move forward inside the embryo to form the most anterior endodermal and mesodermal structures, while the last to involute will form the most posterior structures. (C) Cartoon (not to be taken too literally) showing roughly how the different regions of the ectoderm map into the body surface of the adult animal. (After R.E. Keller, *J. Exp. Zool.* 216:81–101, 1981, with permission from John Wiley & Sons, Inc. and *Dev. Biol.* 42:222–241, 1975, with permission from Academic Press.)

Chemical Signals Trigger the Mechanical Processes

The *VegT*, *Wnt11*, and other mRNA molecules localized in the vegetal cytoplasm of the egg produce localized distributions of their protein products. These act in and on the cells in the lower and middle part of the embryo to give them specialized characters and set them moving, both by direct effects and by stimulating the production of other secreted signal molecules, in particular proteins of the TGFβ superfamily. If these latter signals are blocked, no mesodermal cell types are generated and gastrulation is disrupted. The local activation of the Wnt signaling pathway on the dorsal side of the embryo (as a result of the earlier cortical rotation; see Figure 22–68) modifies the action of the other signals so as to induce development of the special cells that form the dorsal lip of the blastopore (**Figure 22–74**).

The dorsal lip of the blastopore plays a central role in gastrulation not just in a geometrical sense, but as a powerful new source of control. If the dorsal lip of the blastopore is excised from an embryo at the beginning of gastrulation and grafted into another embryo but in a different position, the host embryo initiates gastrulation both at the site of its own dorsal lip and at the site of the graft. The movements of gastrulation at the second site entail the formation of a second



whole set of body structures, and a double embryo (Siamese twins) results (see Figure 22–6B).

Evidently, the dorsal lip of the blastopore is the source of a signal (or signals) coordinating both the movements of gastrulation and the pattern of specialization of the tissues in its neighborhood. Because of this crucial role in organizing the formation of the main body axis, the dorsal lip of the blastopore is known as the **Organizer** (or Spemann's Organizer, after its co-discoverer). It is the oldest and most famous example of an *embryonic signaling center*.

Active Changes of Cell Packing Provide a Driving Force for Gastrulation

The Organizer controls the dorsoventral pattern of cell differentiation in its neighborhood by secreting at least six different signal proteins. These act as diffusible antagonists of the two main types of signals we have already mentioned, coming from the more vegetal cells—that is, of Wnt signals and of TGFβ-like signals (specifically BMP proteins). These inhibitors released from the Organizer may help to limit the size of the Organizer by preventing neighboring cells from also adopting an Organizer character. At the same time, they create a gradient of signaling activity—a morphogen gradient, whose local value reflects the distance from the Organizer (Figure 22–74C). As the cells move during gastrulation, they experience different doses of BMP (and other) signals, delivered with different timing, evoking different cell behaviors and entailing different ultimate fates. But how is the pattern of cell movements organized in mechanical terms, and what are the forces that bring it about?

Gastrulation begins with changes in the shape of the cells at the site of the blastopore. In the amphibian these are called bottle cells: they have broad bodies and narrow necks that anchor them to the surface of the epithelium (**Figure 22–75**), and they may help to force the epithelium to curve and so to tuck inward, producing the initial indentation seen from outside. Once this first tuck has formed, cells can continue to pass into the interior as a sheet to form the gut and mesoderm. The movement seems to be driven mainly by an active repacking of the cells, especially those in the involuting regions around the Organizer (see Figure 22–75). Here **convergent extension** occurs. Small square fragments of tissue from these regions, isolated in culture, will spontaneously narrow and elongate through a rearrangement of the cells, just as they would in the embryo in the process of converging toward the dorsal midline, turning inward around the blastopore lip, and then elongating to form the main axis of the body.

Figure 22–74 A current view of the main inductive signals organizing the events of gastrulation. (A) The distribution of axis-determining molecules in the blastula results from inheritance of different parts of the cytoplasm of the fertilized frog egg. The VegT gene regulatory protein in the vegetal blastomeres is translated from *VegT* mRNA that was localized at the vegetal pole before fertilization. The Wnt11 protein on the future dorsal side is translated from mRNA localized there as a result of the cortical rotation that follows fertilization. (B) VegT drives expression of Xnr (*Xenopus* nodal-related) proteins and other members of the TGFβ superfamily, which induce formation of a band of mesoderm in the middle part of the embryo, while Wnt11 modifies the outcome on the dorsal side, collaborating with Xnr to induce formation of the Organizer. (C) A morphogen gradient that organizes the dorsoventral axis is set up by a combination of signals, including BMP4 (another TGFβ superfamily member) secreted by the mesoderm, and antagonists of the Wnt and BMP pathways, secreted by the Organizer cells at the dorsal lip of the blastopore.

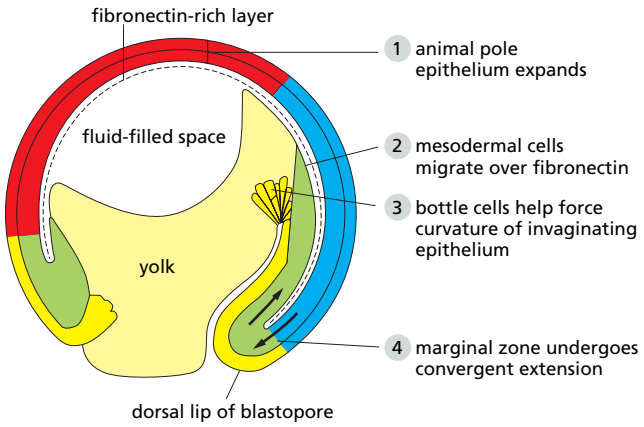


Figure 22–75 Cell movements in gastrulation. A section through a gastrulating *Xenopus* embryo, cut in the same plane as in Figure 22–73, indicating the four main types of movement that gastrulation involves. The animal pole epithelium expands by cell rearrangement, becoming thinner as it spreads. Migration of mesodermal cells over a fibronectin-rich matrix lining the roof of the blastocoel may help to pull the invaginated tissues forward. But the main driving force for gastrulation in *Xenopus* is convergent extension in the marginal zone. (After R.E. Keller, *J. Exp. Zool.* 216:81–101, 1981. With permission from Wiley-Liss.)

To bring about this remarkable transformation, the individual cells have to crawl over one another in a coordinated way (**Figure 22–76**). The alignment of their movements appears to depend on the same machinery we encountered in the worm and the fly controlling planar cell polarity: the Frizzled/Dishevelled polarity-signaling pathway. When this pathway is blocked—for example, by a dominant-negative form of Dishevelled—convergent extension fails to occur.

Changing Patterns of Cell Adhesion Molecules Force Cells Into New Arrangements

Patterns of gene expression govern embryonic cell movements in many different ways. They regulate cell motility, cell shape, and the production of signals for guidance. Very importantly, they also determine the sets of adhesion molecules that the cells display on their surfaces. Through changes in its surface molecules, a cell can break old attachments and make new ones. Cells in one region may develop surface properties that make them cohere with one another and become segregated from a neighboring group of cells whose surface chemistry is different.

Experiments done half a century ago on early amphibian embryos showed that the effects of selective cell–cell adhesion can be so powerful that they can bring about an approximate reconstruction of the normal structure of an early

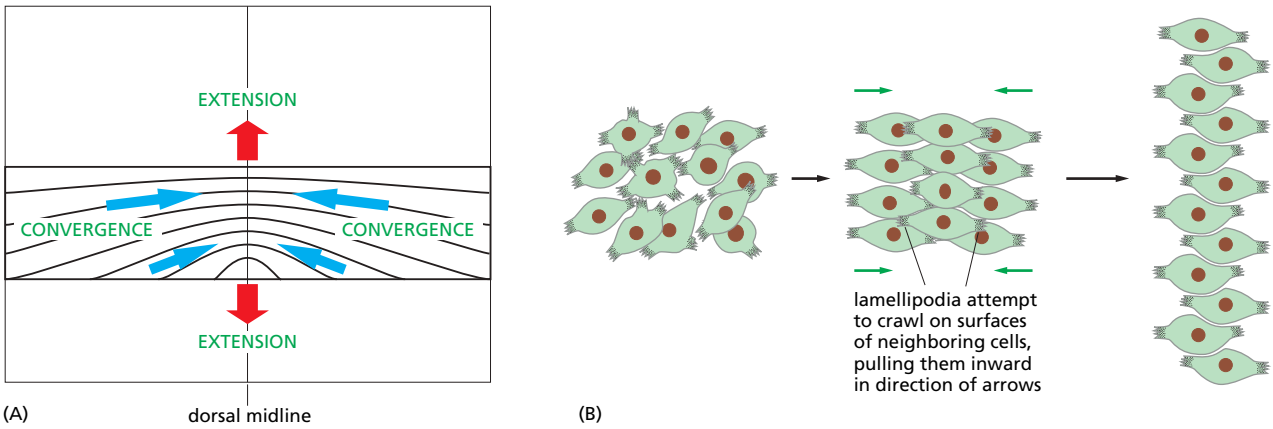


Figure 22–76 Convergent extension and its cellular basis. (A) The pattern of convergent extension in the marginal zone of a gastrula as viewed from the dorsal aspect. *Blue arrows* represent convergence toward the dorsal midline, *red arrows* represent extension of the anteroposterior axis. The simplified diagram does not attempt to show the accompanying movement of involution, whereby the cells are tucking into the interior of the embryo. (B) Schematic diagram of the cell behavior that underlies convergent extension. The cells form lamellipodia, with which they attempt to crawl over one another. Alignment of the lamellipodial movements along a common axis leads to convergent extension. The process depends on the Frizzled/Dishevelled polarity-signaling pathway and is presumably cooperative because cells that are already aligned exert forces that tend to align their neighbors in the same way. (B, after J. Shih and R. Keller, *Development* 116:901–914, 1992. With permission from The Company of Biologists.)

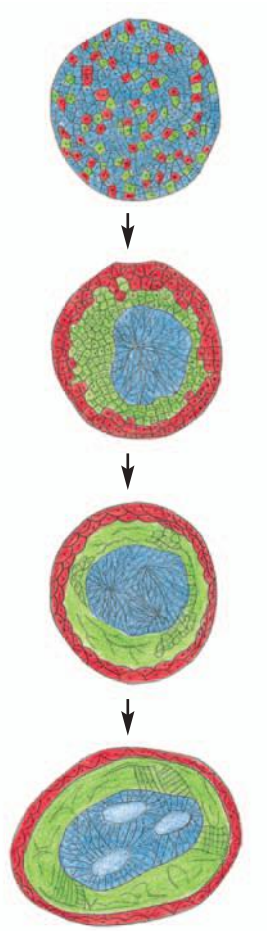


Figure 22–77 Sorting out. Cells from different parts of an early amphibian embryo will sort out according to their origins. In the classical experiment shown here, mesoderm cells *(green),* neural plate cells *(blue),* and epidermal cells *(red)* have been disaggregated and then reaggregated in a random mixture. They sort out into an arrangement reminiscent of a normal embryo, with a "neural tube" internally, epidermis externally, and mesoderm in between. (Modified from P.L. Townes and J. Holtfreter, *J. Exp. Zool.* 128:53–120, 1955. With permission from Wiley-Liss.)

postgastrulation embryo even after the cells have been artificially dissociated. When these cells are reaggregated into a random mixture, the cells sort out spontaneously according to their original characters (**Figure 22–77**). As discussed in Chapter 19, a central role in such phenomena is played by the *cadherins*—a large and varied family of evolutionarily related Ca^{2+}-dependent cell–cell adhesion proteins. These and other cell–cell adhesion molecules are differentially expressed in the various tissues of the early embryo, and antibodies against them interfere with the normal selective adhesion between cells of a similar type.

Changes in the patterns of expression of the various cadherins correlate closely with the changing patterns of association among cells during gastrulation, neurulation, and somite formation (see Figure 19–25). These rearrangements are likely to be regulated and driven in part by the cadherin pattern. In particular, cadherins appear to have a major role in controlling the formation and dissolution of epithelial sheets and clusters of cells. They not only glue one cell to another but also provide anchorage for intracellular actin filaments at the sites of cell–cell adhesion. In this way, the pattern of stresses and movements in the developing tissue is regulated according to the pattern of adhesions.

The Notochord Elongates, While the Neural Plate Rolls Up to Form the Neural Tube

Gastrulation is only the first—though perhaps the most dramatic—of a dizzying variety of cell movements that shape the parts of the body. We have space to discuss only a few of these.

In the embryo just after gastrulation, the layer of mesoderm is divided into separate slabs on the left and right sides of the body. Defining the central body axis, and effecting this separation, is the very early specialization of the mesoderm known as the **notochord**. This slender rod of cells, with ectoderm above it, endoderm below it, and mesoderm on either side (see Figure 22–72), derives from the cells of the Organizer itself. The notochordal cells are characterized by expression of a gene regulatory protein called Brachyury (Greek for "short-tail", from the mutant phenotype); this belongs to the same T-box family as the VegT protein in the vegetal blastomeres.

As the notochordal cells pass around the dorsal lip of the blastopore and move into the interior of the embryo, they form a column of tissue that elongates dramatically by convergent extension. The cells of the notochord also become swollen with vacuoles, so that the rod elongates still further and stretches out the embryo. The notochord is the defining peculiarity of the chordates—the phylum to which the vertebrates belong. It is one of the major vertebrate features that do not have any apparent counterpart in *Drosophila*. In the most primitive chordates, which have no vertebrae, the notochord persists as a primitive substitute for a vertebral column. In vertebrates it serves as a core around which other mesodermal cells will eventually gather to form the vertebrae.

In the overlying sheet of ectoderm, meanwhile, other movements are occurring to form the rudiments of the nervous system. In a process known as *neurulation*, a broad central region of ectoderm, called the *neural plate*, thickens, rolls up into a tube, and pinches off from the rest of the cell sheet. The tube thus created from the ectoderm is called the **neural tube**; it will form the brain and the spinal cord (**Figure 22–78**).

The mechanics of neurulation depend on changes of cell packing and cell shape that make the epithelium roll up into a tube (**Figure 22–79**). Signals initially

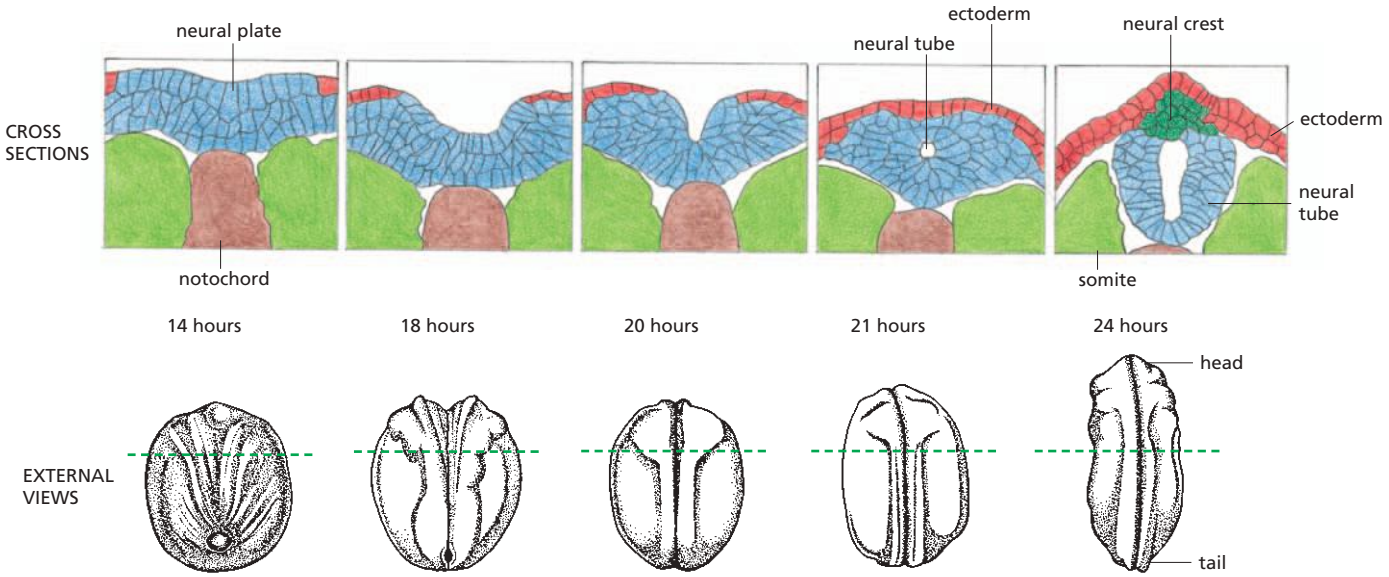


Figure 22–78 Neural tube formation in *Xenopus*. The external views are from the dorsal aspect. The cross sections are cut in a plane indicated by the broken lines. (After T.E. Schroeder, *J. Embryol. Exp. Morphol.* 23:427–462, 1970. With permission from The Company of Biologists.)

from the Organizer and later from the underlying notochord and mesoderm define the extent of the neural plate, induce the movements that make it roll up, and help to organize the internal pattern of the neural tube. The notochord in particular secretes Sonic hedgehog protein—a homolog of the *Drosophila* signal protein Hedgehog—and this acts as a morphogen to control gene expression in the neighboring tissues (**Figure 22–80**).

A Gene-Expression Oscillator Controls Segmentation of the Mesoderm Into Somites

Genetically regulated changes in cell adhesion underlie one of the most striking and characteristic processes in vertebrate development—the formation of the segments of the body axis.

On either side of the newly formed neural tube lies a slab of mesoderm (see Figure 22–72). To form the repetitive series of vertebrae, ribs, and segmental muscles, this slab breaks up into separate blocks, or **somites**—cohesive groups of cells, separated by clefts. **Figure 22–81**A shows the process as it occurs in the chick embryo. The somites form one after another, starting in the head and ending in the tail. Depending on the species, the final number of somites ranges from less than 50 (in a frog or a bird) to more than 300 (in a snake). The posterior, most immature part of the mesodermal slab, called the *presomitic mesoderm*, supplies the necessary tissue: as it retreats tailward, extending the embryo, it deposits a trail of somites. The special character of the presomitic mesoderm is maintained by FGF signaling: *Fgf8* mRNA is synthesized at the tail end of the embryo and slowly degraded as cells move away from this region. Translation of the message results in a gradient of secreted FGF8 protein, with its high point at the tail end.

Formation of the cleft between one somite and the next is foreshadowed by an alternating spatial pattern of gene expression in the presomitic mesoderm: cells about to form the posterior part of a new somite switch on expression of one set of genes, while those destined to form the anterior part of the next somite switch on expression of another set. Selective cohesion resulting from differential gene expression seems to be the underlying cause of the physical segmentation observed.

The problem then is to understand how the repetitive alternating pattern of gene expression is set up. Studies done originally in the chick embryo have provided the beginnings of an answer. In the posterior part of the presomitic mesoderm, expression of certain genes is found to oscillate in time. The first such somite oscillator gene to be discovered was *Hes1*, a homolog of the *Drosophila* pair-rule gene *Hairy* and of the *E(spl)* genes that mediate responses to Notch

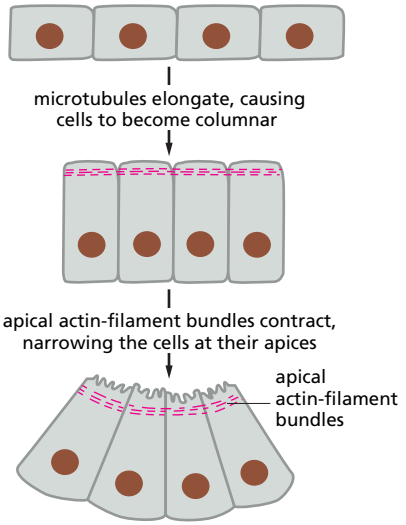


Figure 22–79 The bending of an epithelium through cell shape changes mediated by microtubules and actin filaments. The diagram is based on observations of neurulation in newts and salamanders, where the epithelium is only one cell layer thick. As the apical ends of the cells become narrower, their upper-surface membrane becomes puckered.

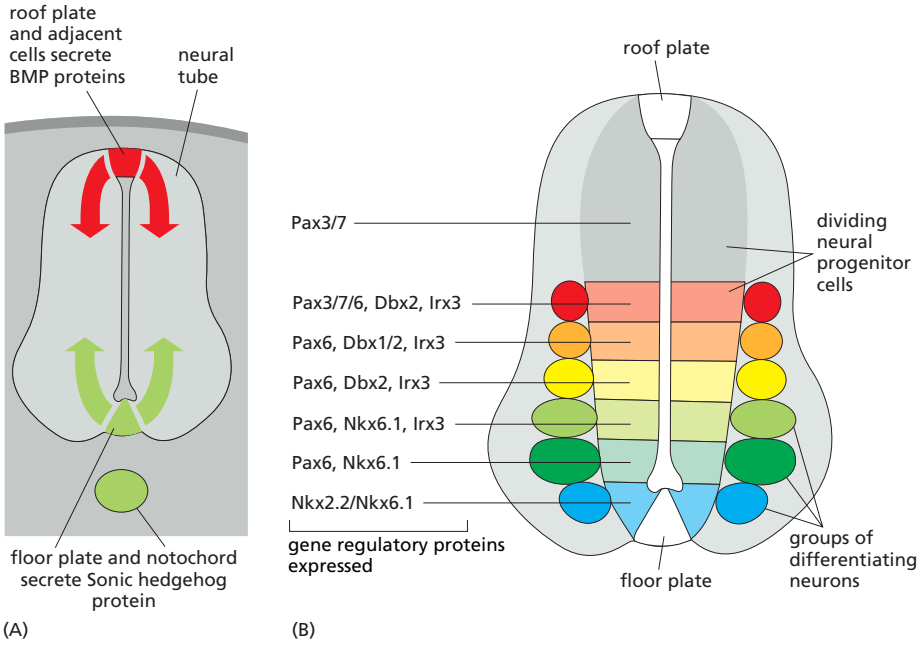


Figure 22–80 A schematic cross section of the spinal cord of a chick embryo, showing how cells at different levels along the dorsoventral axis express different gene regulatory proteins. (A) Signals that direct the dorsoventral pattern: Sonic hedgehog protein from the notochord and the floor plate (the ventral midline of the neural tube) and BMP proteins from the roof plate (the dorsal midline) act as morphogens to control gene expression. (B) The resulting patterns of gene expression in the ventral part of the developing spinal cord. Different groups of proliferating neural progenitor cells (in the ventricular zone, close to the lumen of the neural tube) and of differentiating neurons (in the mantle zone, further out) express different combinations of gene regulatory proteins. Those indicated in this diagram are almost all members of the homeodomain superfamily; various other genes of the same superfamily (including Islet/Lim proteins) are expressed in the differentiating neurons. Neurons expressing different gene regulatory proteins will form connections with different partners and may make different combinations of neurotransmitters and receptors.

signaling. The length of one complete oscillation cycle of this **segmentation clock** (90 minutes in the chick) equals the time taken to lay down one further somite. As cells emerge from the presomitic mesoderm to form somites—in other words, as they lose exposure to the FGF8 signal—their oscillation slows down and finally comes to a halt. Some become arrested in one state, some in another, according to the phase of their oscillation cycle at their time of exit from the presomitic mesoderm. *Hes1* and several of the other oscillating genes code for gene regulatory proteins; thus, the cells that drop below the critical level of FGF8 when they are at the peak of their oscillation cycle switch on one set of regulatory genes, while those passing the threshold at the trough of the cycle switch on another (Figure 22–81B). In this way, it is thought, the temporal oscillation of gene expression in the presomitic mesoderm leaves its trace in a spatially periodic pattern of gene expression in the maturing mesoderm, and this in turn dictates how the tissue will break up into physically separate blocks.

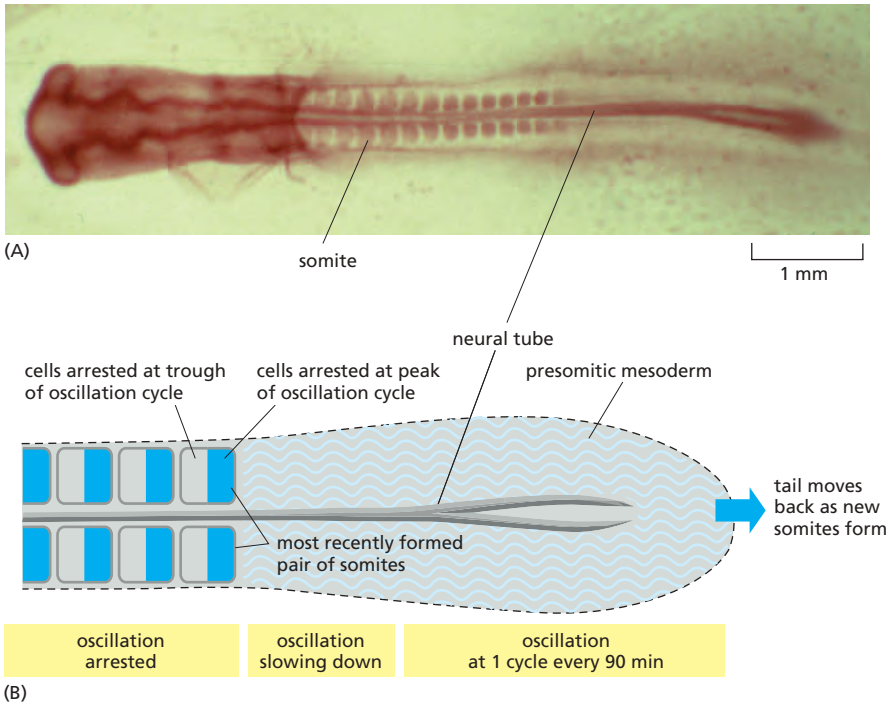
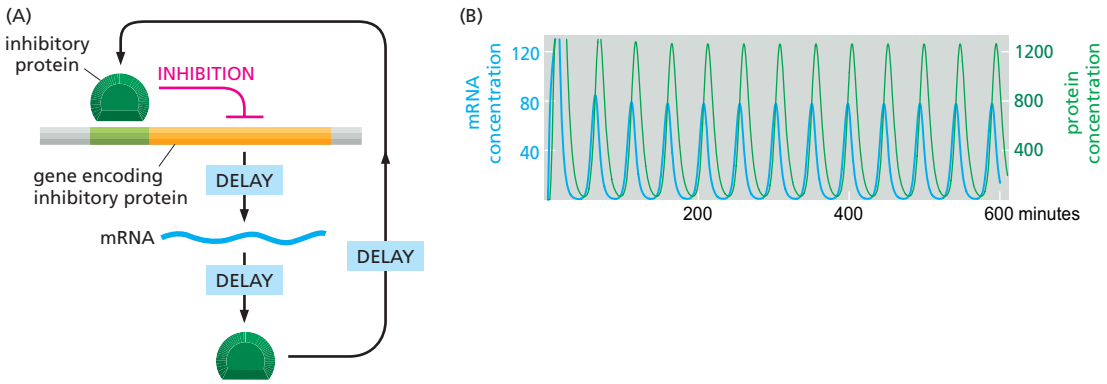


Figure 22–81 Somite formation in the chick embryo. (A) A chick embryo at 40 hours of incubation. (B) How the temporal oscillation of gene expression in the presomitic mesoderm becomes converted into a spatial alternating pattern of gene expression in the formed somites. In the posterior part of the presomitic mesoderm, each cell oscillates with a cycle time of 90 minutes. As cells mature and emerge from the presomitic region, their oscillation is gradually slowed down and finally brought to a halt, leaving them in a state that depends on the phase of the cycle they happen to be in at the critical moment. In this way, a temporal oscillation of gene expression traces out an alternating spatial pattern. (A, from Y.J. Jiang, L. Smithers and J. Lewis, *Curr. Biol.* 8:R868–R871, 1998. With permission from Elsevier.)



Delayed Negative Feedback May Generate the Oscillations of the Segmentation Clock

What, then, is the mechanism that generates the temporal oscillation? How does the clock work? In the mouse, at least three classes of genes have been found to show oscillating expression in the presomitic mesoderm, coding respectively for components of the Notch pathway, the Wnt pathway, and the Fgf pathway; but most of the mutations that are known to break the clock and disrupt somite segmentation lie in components of the Notch pathway. These include genes (such as *Hes1* and more importantly its relative *Hes7*) that are regulated by Notch and code for inhibitory gene regulatory proteins. Some of these proteins act directly on the regulatory DNA of their own gene so as to inhibit their own expression. According to one theory, this simple negative feedback loop could be the basic generator of the oscillations **Figure 22–82**: when the gene is transcribed, the amount of its protein product builds up until transcription is inhibited and synthesis of the protein ceases; the protein then decays, permitting transcription to begin again; and so on. There is a time-lag from the beginning of a new bout of transcription to the first appearance in the nucleus of the resulting regulatory protein molecules, because it takes time for the RNA polymerase to traverse the gene, for the resulting RNA transcript then to mature, leave the nucleus, and direct synthesis of a protein molecule, and for the protein then to enter the nucleus to control transcription. This delay in the feedback loop is proposed to be the main determinant of the period of oscillation of the clock and thus of the size of each somite.

Most of the cells of each newly formed somite will rapidly differentiate to form a block of muscle, corresponding to one muscle segment of the main body axis. The embryo can (and does) now begin to wriggle. Separate subsets of the somite cells will go to form the vertebrae and other connective tissues such as dermis. A further subset detach from the somite and migrate away into the lateral unsegmented mesoderm, crawling through the spaces between other cells: these emigrants will give rise to almost all the other skeletal muscle cells in the body, including those of the limbs.

Figure 22–82 Delayed negative feedback giving rise to oscillating gene expression. (A) A single gene, coding for a gene regulatory protein that inhibits its own expression, can behave as an oscillator. For oscillation to occur, there must be a delay (or several delays) in the feedback circuit, and the lifetimes of the mRNA and protein must be short compared with the total delay. The delay determines the period of oscillation. According to one theory, a feedback circuit like this, based on a gene called *Her7* in the zebrafish, or *Hes7* in the mouse (a relative of *Hes1*), is the pacemaker of the segmentation clock governing somite formation. (B) The predicted oscillation of *Her7* mRNA and protein, computed using rough estimates of the feedback circuit parameters appropriate to this gene in the zebrafish. Concentrations are measured as numbers of molecules per cell. The predicted period is close to the observed period, which is 30 minutes per somite in the zebrafish.

Embryonic Tissues Are Invaded in a Strictly Controlled Fashion by Migratory Cells

The muscle-cell precursors, or *myoblasts*, that emigrate from the somites are determined but not overtly differentiated. In the tissues that they colonize they will mingle with other classes of cells from which they appear practically indistinguishable; but they will maintain expression of myoblast-specific gene regulatory proteins (such as Pax3 and members of the MyoD family), and when the time comes for differentiation, they, and they alone, will turn into muscle cells (**Figure 22–83**).

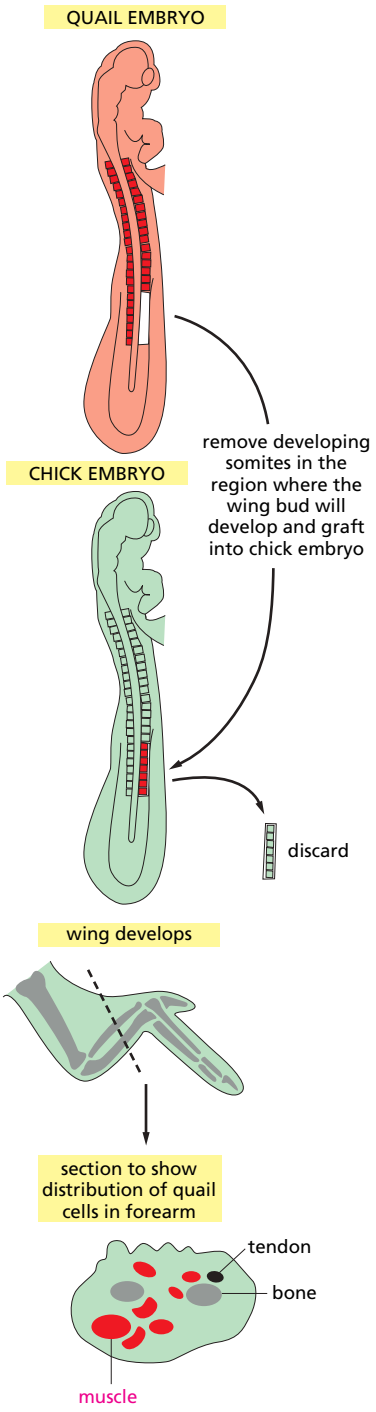


Figure 22–83 The migratory origin of limb muscle cells. The migrations can be traced by grafting cells from a quail embryo into a chick embryo; the two species are very similar in their development, but the quail cells are recognizable by the distinctive appearance of their nucleoli. If quail somite cells are substituted for the somite cells of a chick embryo at 2 days of incubation and the wing of the chick is sectioned a week later, it is found that the muscle cells in the chick wing derive from the transplanted quail somites.

The eventual pattern of muscles—in the limbs, for example—is determined by the routes that the migrant cells follow and the selection of sites that they colonize. The embryonic connective tissues form the framework through which the myoblasts travel and provide signals that guide their distribution. No matter which somite they come from, myoblasts that migrate into a forelimb bud will form the pattern of muscles appropriate to a forelimb, and those that migrate into a hindlimb bud will form the pattern appropriate to a hindlimb.

Other classes of migrant cells, meanwhile, select different routes for their travels. Along the line where the neural tube pinches off from the future epidermis, a number of ectodermal cells break loose from the epithelium and also migrate as individuals out through the mesoderm (**Figure 22–84**). These are the cells of the **neural crest**; they will give rise to almost all of the neurons and glial cells of the peripheral nervous system, as well as the pigment cells of the skin and many connective tissues in the head, including bones of the skull and jaws. Other important migrants are the precursors of the blood cells, of the germ cells, and of many groups of neurons within the central nervous system, as well as the *endothelial cells* that form blood vessels. Each of these classes of travelers will colonize a different set of sites. As a result of such invasions, most tissues in the vertebrate body are mixtures of cells of different characters derived from widely separate parts of the embryo.

As a migrant cell travels through the embryonic tissues, it repeatedly extends projections that probe its immediate surroundings, testing for subtle cues to which it is particularly sensitive by virtue of its specific assortment of cell-surface receptor proteins. Inside the cell these receptor proteins are connected to the cytoskeleton, which moves the cell along. Some extracellular matrix materials,

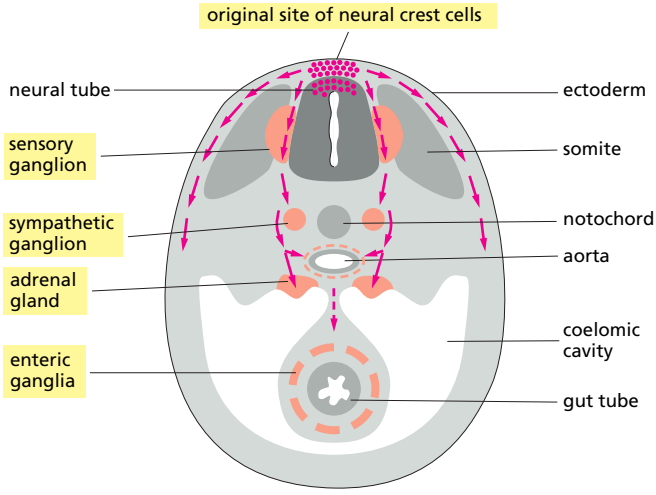


Figure 22–84 The main pathways of neural crest cell migration. A chick embryo is shown in a schematic cross section through the middle part of the trunk. Deep-lying neural crest derivatives are indicated by *yellow* text boxes. The cells that take the pathway just beneath the ectoderm will form pigment cells of the skin; those that take the deep pathway via the somites will form the neurons and glial cells of sensory and sympathetic ganglia, and parts of the adrenal gland. The neurons and glial cells of the enteric ganglia, in the wall of the gut, are formed from neural crest cells that migrate along the length of the body, originating from either the neck region or the sacral region. In *Drosophila,* neurons in the wall of the gut originate in a similar way, by migration from the head end of the embryo. (See also Figure 19–23.)

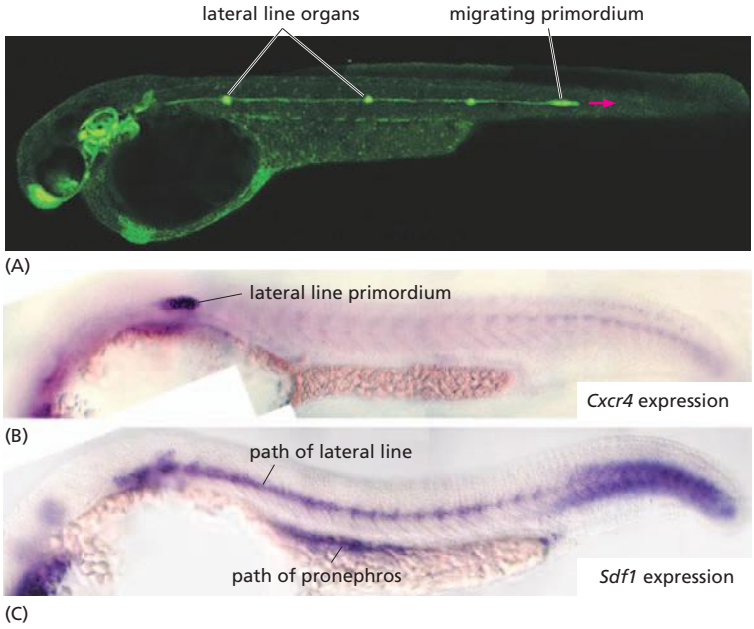


Figure 22–85 Migration of the lateral line primordium in a zebrafish larva, guided by SDF1 and CXCR4. The lateral line is a row of mechanosensory organs, closely similar to the sensory patches in the inner ear, which detect the movement of water over the surface of a fish or amphibian. (A) They originate as clusters of cells deposited by a primordium that migrates along the the flank of the larva, from a site in the head all the way down to the tail, as shown in this 2-day larva in which the lateral line cells are labeled by expression of Green Fluorescent Protein. (B) Cells in the primordium express the chemotaxis receptor CXCR4, shown here by *in situ* hybridization in a 1-day larva. (C) The track that they will follow is marked by expression of the ligand SDF1, shown by *in situ* hybridization in another 1-day specimen. If the ligand is lacking along the normal route (as a result of a mutation), the primordium departs from its proper route to follow an alternative more ventral track marked by another stripe of SDF1, defining the normal path of another migratory structure, the pronephros. (A, courtesy of David Gilmour; B and C, from N.B. David et al., *Proc. Natl Acad. Sci. U.S.A.* 99:16297–16302, 2002. With permission from National Academy of Sciences.)

such as the protein fibronectin, provide adhesive sites that help the cell to advance; others, such as chondroitin sulfate proteoglycan, inhibit locomotion and repel immigration. The nonmigrant cells along the pathway may likewise have inviting or repellent surfaces, or may even extend filopodia that touch the migrant cell and affect its behavior.

Among this mass of different guiding influences, a few stand out as particularly important. In particular, cells of many different types are guided by chemotaxis that depends on a receptor called CXCR4. This cell-surface protein belongs to the family of G-protein-coupled receptors, and it is activated by an extracellular ligand called SDF1. Cells expressing CXCR4 can snuffle their way along tracks marked out for them by production of SDF (**Figure 22–85**). Chemotaxis towards sources of SDF1 plays a major part in guiding the migrations of lymphocytes and of various other white blood cells; of neurons in the developing brain; of muscle progenitor cells entering limb buds; of primordial germ cells as they travel toward the gonads; and of cancer cells when they metastasize.

The Distribution of Migrant Cells Depends on Survival Factors as Well as Guidance Cues

The final distribution of migrant cells depends not only on the routes they take, but also on whether they survive the journey and thrive in the environment they find at the journey's end. Specific sites provide survival factors needed by specific types of migrant. For example, the neural crest cells that give rise to the pigment cells of the skin and the nerve cells of the gut depend on a peptide factor called *endothelin-3* that is secreted by tissues on the migration pathways; mutant mice and humans defective in the gene for this factor or its receptor have nonpigmented (albino) patches and potentially lethal gut malformations resulting from the lack of gut innervation (a condition called megacolon, because the colon becomes hugely distended).

Germ cells, blood cell precursors, and neural-crest-derived pigment cells all appear to share at least one common requirement for survival. This involves a transmembrane receptor, called the *Kit protein*, in the membrane of the migrant cells, and a ligand, called the *Steel factor*, produced by the cells of the tissue through which the cells migrate and/or in which they come to settle. Individuals with mutations in the genes for either of these proteins are deficient in their pigmentation, their supply of blood cells, and their production of germ cells (**Figure 22–86**).





Figure 22–86 Effect of mutations in the *Kit* gene. Both the baby and the mouse are heterozygous for a loss-of-function mutation that leaves them with only half the normal quantity of Kit gene product. In both cases pigmentation is defective because pigment cells depend on the Kit product as a receptor for a survival factor. (Courtesy of R.A. Fleischman, from *Proc. Natl Acad. Sci. U.S.A.* 88:10885–10889, 1991. With permission from National Academy of Sciences.)

Left–Right Asymmetry of the Vertebrate Body Derives From Molecular Asymmetry in the Early Embryo

Vertebrates may look bilaterally symmetrical from the outside, but many of their internal organs—the heart, the stomach, the liver, and so on—are highly asymmetric. This asymmetry is quite reproducible: 99.98% of people have their heart on the left. We have seen how a vertebrate embryo develops its internal and external tissue layers and its anteroposterior and dorsoventral axes. But how does the left–right asymmetry arise?

Genetic studies in mammals show that this problem can be broken down into two distinct questions—one concerning the creation of asymmetry and the other concerning its orientation. Several mutations are known, in humans and in mice, that cause a randomization of the left–right axis: 50% of the mutant individuals have their internal organs arranged in the normal way, while the other 50% have an inverted anatomy, with the heart on the right. In these individuals, it seems, the mechanism that makes the left and right sides different has functioned correctly, but the mechanism that decides between the two possible orientations of the left–right axis is defective.

A key to the basis of these phenomena comes from the discovery of molecular asymmetries that precede the first gross anatomical asymmetries. The earliest signs are seen in patterns of gene expression in the neighborhood of the *node*—the homolog in mouse and chick of the frog Organizer. In particular, the gene *Nodal*, coding for a member of the TGFβ superfamily, is expressed asymmetrically in this region (not only in the mouse, but also in chick, frog and zebrafish) (**Figure 22–87**). Asymmetry of *Nodal* expression in the immediate neighborhood of the node is relayed outward to create a broad stripe of *Nodal*

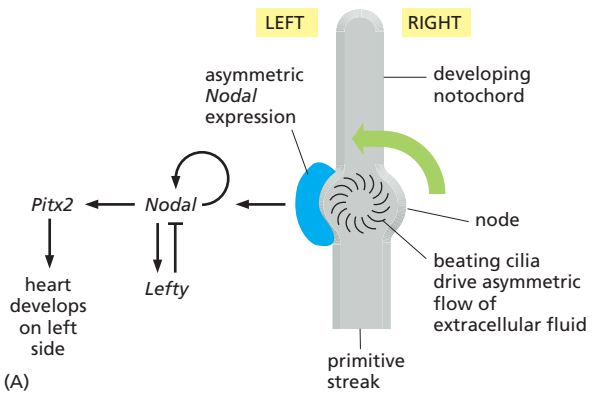




Figure 22–87 Helical beating of cilia at the node, and the origins of left-right asymmetry. (A) The beating of the cilia drives a fluid flow toward one side of the node, and this leads to asymmetric gene expression in the neighborhood of the node. According to one theory, the flow exerts this effect by carrying extracellular signal proteins to one side. Another theory notes that cilia can also function as mechanosensors, and proposes that a subset of cilia at the node respond to deflection due to the fluid flow by opening Ca^{2+} channels so as to create an increased Ca^{2+} concentration in the cells on one side. (B) The resulting asymmetric expression pattern of *Nodal*, coding for a signal protein belonging to the TGFβ superfamily, in the neighborhood of the node *(lower two blue spots)* in a mouse embryo at 8 days of gestation, as shown by *in situ* hybridization. At this stage, the asymmetry has already been relayed outward to the lateral plate mesoderm, where *Nodal* is expressed on the left side *(large elongated blue patch)* but not the right. (B, courtesy of Elizabeth Robertson.)

expression in the mesoderm along the left side—and only the left side—of the embryo's body. The mechanism that relays the asymmetry from the node and localizes *Nodal* expression is not understood and may vary from one class of vertebrates to another. In all species, however, it seems to depend on feedback loops involving *Nodal* together with a second set of genes, the *Lefty* genes. These, like *Nodal* itself, are directly regulated by the Nodal signaling pathway and their products, the Lefty proteins, are related to Nodal; but Lefty proteins diffuse more widely and act oppositely, as Nodal antagonists. Mice with a knockout mutation in the *Lefty1* gene frequently have the right side converted into a mirror image of the left, so that left–right asymmetry is lost.

Another gene that is directly regulated by the Nodal pathway, *Pitx2*, coding for a gene regulatory protein, links the outcome of the Nodal/Lefty interactions to subsequent anatomical development. Nodal drives *Pitx2* expression on the left side of the body and thereby confers asymmetry on the heart and other internal organs.

This leaves us with the puzzle of how the initial asymmetry of *Nodal* expression originates. Whatever the mechanism, the outcome of events at the node in a normal animal must be biased so that left-specific genes are regularly expressed on the left side: there has to be a link between the mechanism that creates asymmetry and the mechanism that orients it. A clue to the orienting mechanism first came to light in a Swedish infertility clinic. A small subset of infertile men were found to have sperm that were immotile because of a defect in the dynein molecules needed for beating of cilia and flagella. These men also suffered from chronic bronchitis and sinusitis because the cilia in their respiratory tract were defective. And strikingly, 50% of them had their internal organs left–right inverted, with the heart on the right. The findings originally seemed completely mysterious; but similar effects are seen in mammals with other mutations resulting in defective cilia. This suggests that ciliary beating somehow controls which way the left–right axis is oriented.

Time-lapse videomicroscopy in the living mouse embryo reveals that the cells at the node, on its internal face, have cilia that beat in a helical fashion: like a screw-thread, they have a definite handedness, and at the node they are set in a little hollow that is shaped so that their beating drives a current of fluid towards the left side (see Figure 22–87A). According to one theory, signal proteins carried in this current toward the left side provide the bias that orients the left–right axis of the mouse body. Another theory proposes that cilia in this system, as in certain other contexts, act not only as drivers of fluid flow but also as mechanical sensors, responding to deflection by generating an asymmetric current of Ca^{2+} ions across the node to influence adjacent tissue.

The handedness of the ciliary beating reflects the handedness—the left–right asymmetry—of the organic molecules of which all living things are made. It seems that this, therefore, is the ultimate director of the left–right asymmetry of our anatomy.

Summary

Animal development involves dramatic cell movements. Thus, in gastrulation, cells from the exterior of the early embryo tuck into the interior to form a gut cavity and create the three germ layers—endoderm, mesoderm, and ectoderm—from which higher animals are constructed. In vertebrates, the movements of gastrulation are organized by signals from the Organizer (the dorsal lip of the amphibian blastopore, corresponding to the node in a chick or mouse embryo). These signals specify the dorsoventral axis of the body and govern convergent extension, in which the sheet of cells moving into the interior of the body lengthens along the head-to-tail axis while narrowing at right angles to this axis. The active repacking movements of individual cells that drive convergent extension are coordinated through the Frizzled/Dishevelled planar-polarity signaling pathway—a branch of the Wnt signaling pathway that regulates the actin cytoskeleton.

Subsequent development involves many further cell movements. Part of the ectoderm thickens, rolls up, and pinches off to form the neural tube and neural crest. In the

midline, a rod of specialized cells called the notochord elongates to form the central axis of the embryo. The long slabs of mesoderm on either side of the notochord become segmented into somites. Migrant cells, such as those of the neural crest, break loose from their original neighbors and travel through the embryo to colonize new sites. Primordial germ cells and many other migrants are guided by chemotaxis dependent on the receptor CXCR4 and its ligand SDF1. Specific cell adhesion molecules, such as cadherins and integrins, help to guide the migrations and control the selective cohesion of cells in new arrangements.

Ultimately, the pattern of cell movements is directed by the pattern of gene expression, which determines cell surface properties and motility. Thus, the formation of somites depends on a periodic pattern of gene expression, which is laid down by a biochemical oscillator—the segmentation clock—in the mesoderm and dictates the way the mass of cells will break up into separate blocks. Similarly, the left–right anatomical asymmetry of the vertebrate body is foreshadowed by left–right asymmetry in the pattern of gene expression in the early embryo. This asymmetry, in mammals at least, is thought to be directed ultimately by the handedness of ciliary beating in the neighborhood of the node.

THE MOUSE

The mouse embryo—tiny and inaccessible in its mother's womb—presents a hard challenge to developmental biologists. It has, however, two immediate attractions. First, the mouse is a mammal, and mammals are the animals that we, as humans, care about most. Second, among mammals, it is one of the most convenient for genetic studies, because it is small and breeds rapidly. These two factors have spurred an enormous research effort, resulting in the development of some remarkably powerful experimental tools. In this way, the mouse has become the main model organism for experimentation in mammalian genetics and the most intensively studied surrogate for humans. It is separated from humans by only about 100 million years of evolution. Its genome is the same as ours in size, and there is very nearly a one-to-one correspondence between mouse and human genes. Our proteins are typically 80–90% identical in amino acid sequence, and large blocks of close nucleotide sequence similarity are also evident when the regulatory DNA sequences are compared.

Through ingenuity and perseverance, developmental biologists have now found ways to gain access to the early mouse embryo without killing it and to generate mice to order with mutations in any chosen gene. Almost any genetic modification that can be made in a worm, a fly, or a zebrafish can now also be made in the mouse, and in some cases made better. The costs of research in the mouse are far greater, but so are the incentives. As a result, the mouse has become a rich source of information about all aspects of the molecular genetics of development—a key model system not only for mammals, but also for other animals. It has provided, for example, much of what we know about *Hox* genes, left–right asymmetry, cell death controls, the role of Notch signaling, and a host of other topics.

We have already drawn repeatedly upon data from the mouse. We shall make use of it even more in the next chapter, where we discuss adult tissues and the developmental processes that occur in them. In this section, we examine the special features of mouse development that have been exploited to make the genetic manipulations possible. By way of example, we shall also outline how the mouse has been used to illuminate one further important developmental process—the creation of organs such as lungs and glands by interactions between embryonic connective tissue and epithelium.

Mammalian Development Begins With a Specialized Preamble

The mammalian embryo begins its development in an exceptional way. Protected within the uterus, it does not have the same need as the embryos of most

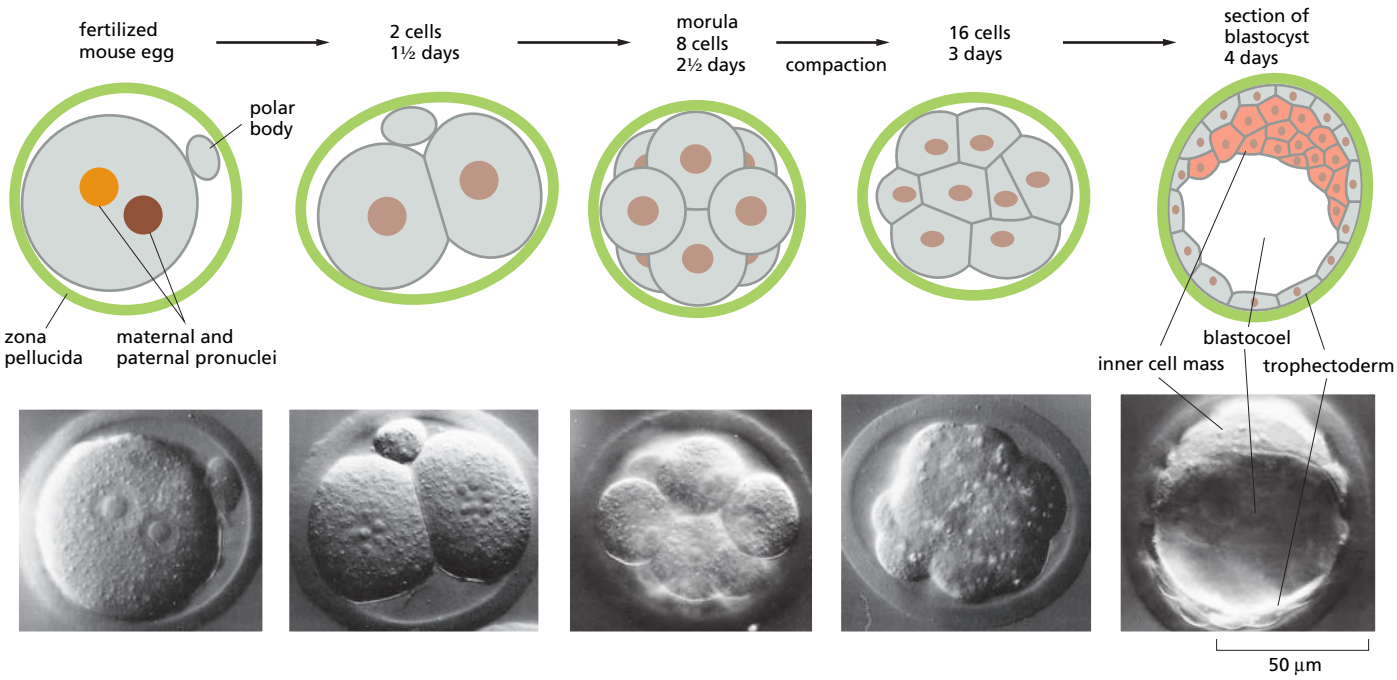
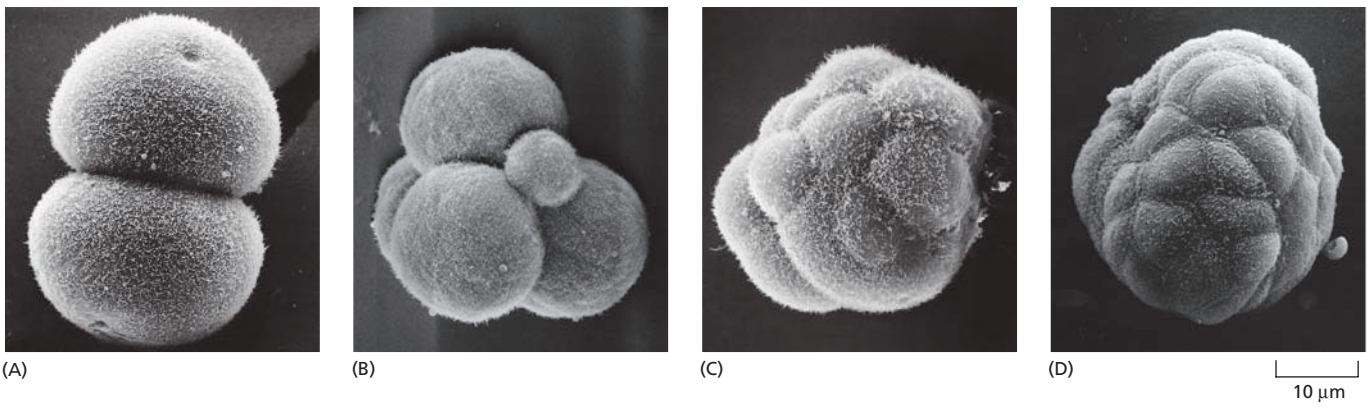


Figure 22–88 The early stages of mouse development. The zona pellucida is a jelly capsule from which the embryo escapes after a few days, allowing it to implant in the wall of the uterus. (Photographs courtesy of Patricia Calarco.)

other species to complete the early stages of development rapidly. Moreover, the development of a placenta quickly provides nutrition from the mother, so that the egg does not have to contain large stores of raw materials such as yolk. The egg of a mouse has a diameter of only about 80 μm and therefore a volume about 2000 times smaller than that of a typical amphibian egg. Its cleavage divisions occur no more quickly than the divisions of many ordinary somatic cells, and gene transcription has already begun by the 2-cell stage. Most importantly, while the later stages of mammalian development are similar to those of other vertebrates such as *Xenopus*, mammals begin by taking a large developmental detour to generate a complicated set of structures—notably the amniotic sac and the placenta—that enclose and protect the embryo proper and provide for the exchange of metabolites with the mother. These structures, like the rest of the body, derive from the fertilized egg but are called *extraembryonic* because they are discarded at birth and form no part of the adult. Similar accessory structures are formed in the development of birds and reptiles.

The early stages of mouse development are summarized in **Figure 22–88**. The fertilized egg divides to generate 16 cells by 3 days after fertilization. At first, the cells stick together only loosely, but beginning at the 8-cell stage they become more cohesive and undergo *compaction* to form a solid ball of cells called a *morula* (Latin for "little mulberry") (**Figure 22–89**). Apical tight junctions form between the cells, sealing off the interior of the morula from the external medium. Soon after this, an internal cavity develops, converting the morula into a *blastocyst*—a hollow sphere. The outer layer of cells, forming the

Figure 22–89 Scanning electron micrographs of the early mouse embryo. The zona pellucida has been removed. (A) Two-cell stage. (B) Four-cell stage (a polar body is visible in addition to the four blastomeres—see Figure 21–23). (C) Eight-to-sixteen-cell morula-compaction occurring. (D) Blastocyst. (A–C, courtesy of Patricia Calarco; D, from P. Calarco and C.J. Epstein, *Dev. Biol.* 32:208–213, 1973. With permission from Academic Press.)



wall of the sphere, is called the *trophectoderm*. It will give rise to extraembryonic tissues. An inner clump of cells, called the *inner cell mass*, is located to one side of the cavity. It will give rise to the whole of the embryo proper.

After the embryo has escaped from its jelly capsule (at about four days), the cells of the trophectoderm make close contact with the wall of the uterus, initiating the process of implantation that will lead on to formation of the placenta. Meanwhile the inner cell mass grows and begins to differentiate. Part of it gives rise to some further extraembryonic structures, such as the yolk sac, while the rest of it goes on to form the embryo proper by processes of gastrulation, neurulation, and so on, that are fundamentally similar to those seen in other vertebrates, although distortions of the geometry make some of the homologies hard to discern at first sight.

The Early Mammalian Embryo Is Highly Regulative

Localized intracellular determinants play only a small part in early mammalian development, and the blastomeres produced by the first few cell divisions are remarkably adaptable. If the early embryo is split in two, a pair of identical twins can be produced—two complete normal individuals from a single cell. Similarly, if one of the cells in a 2-cell mouse embryo is destroyed by pricking it with a needle and the resulting "half-embryo" is placed in the uterus of a foster mother to develop, in many cases a perfectly normal mouse will emerge.

Conversely, two 8-cell mouse embryos can be combined to form a single giant morula, which then develops into a mouse of normal size and structure (**Figure 22–90**). Such creatures, formed from aggregates of genetically different groups of cells, are called *chimeras*. Chimeras can also be made by injecting cells from an early embryo of one genotype into a blastocyst of another genotype. The injected cells become incorporated into the inner cell mass of the host blastocyst, and a chimeric animal develops. A single cell taken from an 8-cell embryo or from the inner cell mass of another early blastocyst can give rise in these ways to any combination of cell types in the chimera. Wherever the added cell may happen to find itself, it responds correctly to cues from its neighbors and follows the appropriate developmental pathway.

These findings have two implications. First, during the early stages, the developmental system is self-adjusting, so that a normal structure emerges even if the starting conditions are perturbed. Embryos or parts of embryos that have this property are said to be **regulative**. Second, the individual cells of the inner cell mass are initially *totipotent*, or very nearly so: though they cannot form trophoblast, they can give rise to any part of the adult body, including germ cells.

Totipotent Embryonic Stem Cells Can Be Obtained From a Mammalian Embryo

If a normal early mouse embryo is grafted into the kidney or testis of an adult, its development is disturbed beyond any possibility of proper regulation, but not halted. The result is a bizarre tumorous growth known as a *teratoma*, consisting of a disorganized mass of cells containing many varieties of differentiated tissue—skin, bone, glandular epithelium, and so on—mixed with undifferentiated stem cells that continue to divide and generate yet more of these differentiated tissues.

Investigation of the stem cells in teratomas and related types of tumors led to the discovery that their behavior reflects a remarkable property of the cells of the normal inner cell mass: given a suitable environment, they can be induced to proliferate indefinitely while retaining their totipotent character. Cultured cells with this property are called **embryonic stem cells**, or **ES cells**. They can be derived by placing a normal inner cell mass in culture and dispersing the cells as soon as they proliferate. Separating the cells from their normal neighbors and putting them in the appropriate culture medium evidently arrests the normal program of change of cell character with time and so enables the cells to carry

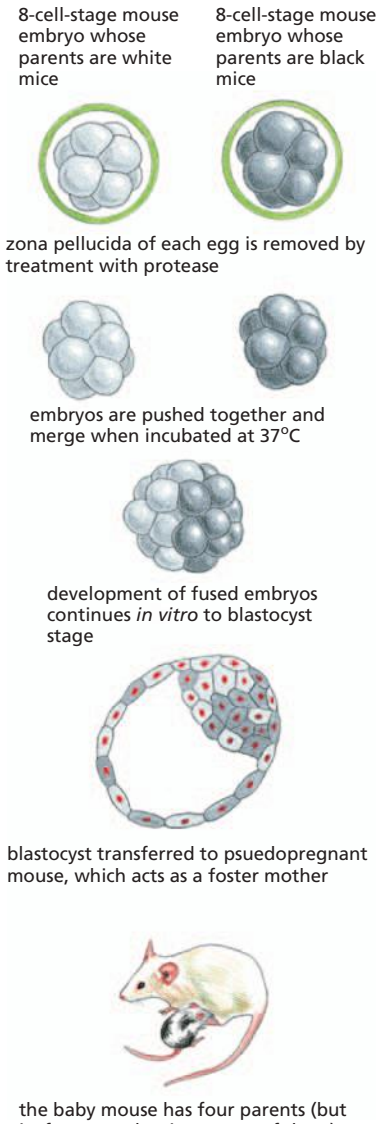


Figure 22–90 A procedure for creating a chimeric mouse. Two morulae of different genotypes are combined.

on dividing indefinitely without differentiating. Many tissues of the adult body also contain stem cells that can divide indefinitely without terminally differentiating, as we shall see in the next chapter; but these *adult stem cells,* when allowed to differentiate, normally give rise only to a narrowly restricted range of differentiated cell types.

The state in which the ES cells are arrested seems to be equivalent to that of normal inner-cell-mass cells. This can be shown by taking ES cells from the culture dish and injecting them into a normal blastocyst (**Figure 22–91**). The injected cells become incorporated in the inner cell mass of the blastocyst and can contribute to the formation of an apparently normal chimeric mouse. Descendants of the injected stem cells can be found in practically any of the tissues of this mouse, where they differentiate in a well-behaved manner appropriate to their location and can even form viable germ cells. The extraordinarily adaptable behavior of ES cells shows that cues from a cell's neighbors not only guide choices between different pathways of differentiation, but can also stop or start the developmental clock—the processes that drive a cell to progress from an embryonic to an adult state.

On a practical level, ES cells have a twofold importance. First, from a medical point of view, they offer the prospect of a versatile source of cells for repair of damaged and defective tissues in the adult body, as we shall discuss at the end of the next chapter. Second, ES cells make possible the most precisely controlled forms of genetic modification, allowing animals to be created with virtually any desired alteration introduced into their genome. As discussed in Chapter 8, the technique uses genetic recombination to substitute an artificially constructed DNA segment for the normal DNA sequence at a chosen site in the genome of an ES cell. Although only a rare cell incorporates the DNA construct correctly, selection procedures have been devised to find this cell among the thousands of cells into which the DNA construct has been transfected. Once selected, the genetically modified ES cells can be injected into a blastocyst to make a chimeric mouse. This mouse will, with luck, have some ES-derived germ cells, capable of acting as founders of a new generation of mice that consist entirely of cells carrying the carefully designed mutation. In this way, an entire mutant mouse can be resurrected from the culture dish (see Figure 8–65).

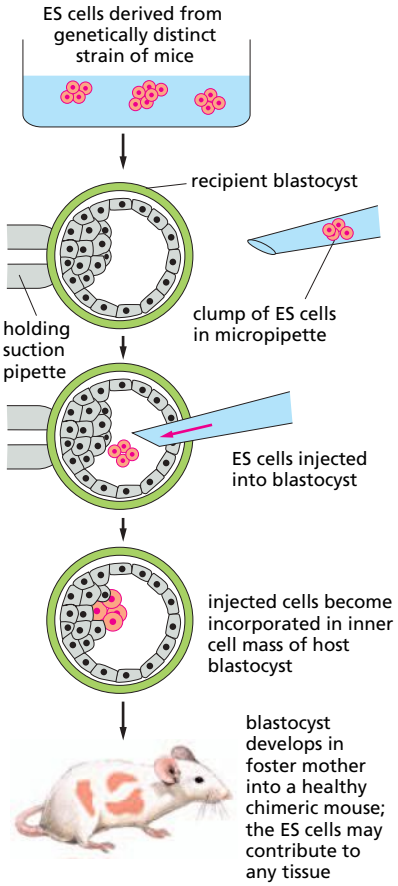


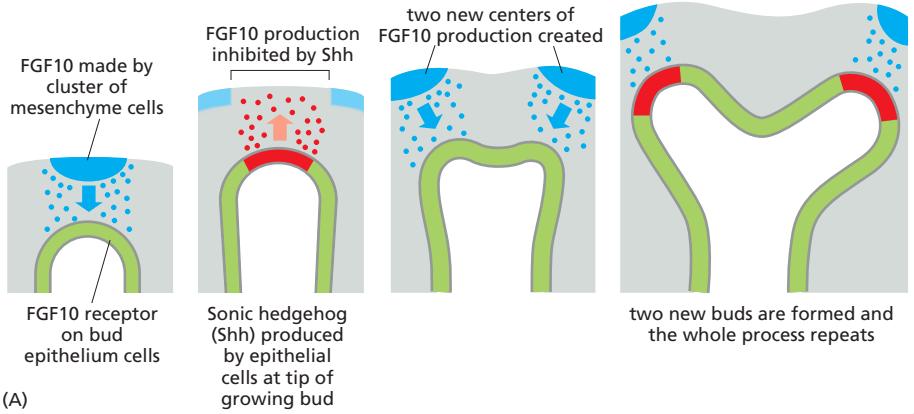
Figure 22–91 Making a chimeric mouse with ES cells. The cultured ES cells can combine with the cells of a normal blastocyst to form a healthy chimeric mouse, and can contribute to any of its tissues, including the germ line. Thus the ES cells are totipotent.

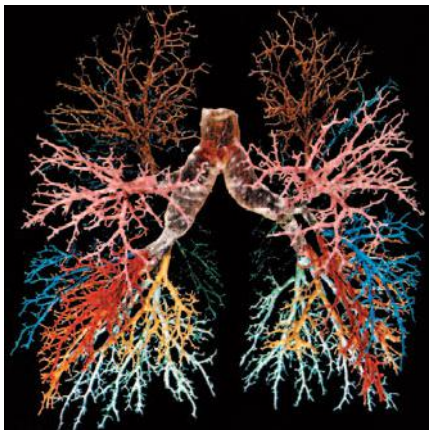
Interactions Between Epithelium and Mesenchyme Generate Branching Tubular Structures

Vertebrates are comparatively big animals, and they owe much of their bulk to connective tissues. For excretion, absorption of nutrients, and gas exchange, however, they also require large quantities of various specialized types of epithelial surfaces. Many of these take the form of tubular structures created by *branching morphogenesis,* in which an epithelium invades embryonic connective tissue (mesenchyme) to form a composite organ. The lung is a typical example. It originates from the endoderm lining the floor of the foregut. This epithelium buds and grows out into the neighboring mesenchyme to form the *bronchial tree,* a system of tubes that branch repeatedly as they extend (**Figure 22–92**). The same mesenchyme is also invaded by endothelial cells—the lining cells of blood vessels—to create the system of closely apposed airways and blood vessels required for gas exchange in the lung (discussed in Chapter 23).

The whole process depends on exchanges of signals in both directions between the growing buds of epithelium and the mesenchyme that they are invading. These signals can be analyzed by genetic manipulation in the mouse. A central part is played by signal proteins of the fibroblast growth factor (FGF) family and the receptor tyrosine kinases on which they act. This signaling pathway has various roles in development, but it seems to be especially important in the many interactions that occur between epithelium and mesenchyme.

Mammals have about 20 different *Fgf* genes, as compared with three in *Drosophila* and two in *C. elegans*. The *Fgf* that is most important in the lung is *Fgf10*. This is expressed in clusters of mesenchyme cells near the tips of the growing epithelial tubes, while its receptor is expressed in the epithelial cells





themselves. FGF10 or its receptor can be knocked out (by the standard techniques based on recombination in ES cells). In the resulting knock-out mutant mouse, the whole process of branching morphogenesis then fails—a primary bud of lung epithelium is formed but fails to grow out into the mesenchyme to create a bronchial tree. Conversely, a microscopic bead soaked in FGF10 and placed near embryonic lung epithelium in culture will induce a bud to form and grow out toward it. Evidently, the epithelium invades the mesenchyme only by invitation, in response to FGF10.

But what makes the growing epithelial tubes branch repeatedly as they invade? This seems to depend on a Sonic hedgehog signal that is sent in the opposite direction, from the epithelial cells at the tips of the buds back to the mesenchyme. In mice lacking Sonic hedgehog, the lung epithelium grows and differentiates, but forms a sac instead of a branching tree of tubules. Meanwhile, FGF10, instead of being restricted to small clusters of mesenchyme cells, with each cluster acting as a beacon to direct the outgrowth of a separate epithelial bud, is expressed in broad bands of cells immediately adjacent to the epithelium. This finding suggests that the Sonic hedgehog signal may serve to shut off FGF10 expression in the mesenchyme cells closest to the growing tip of a bud, splitting the FGF10-secreting cluster into two separate clusters, which in turn cause the bud to branch into two (see Figure 22–92A).

The branching growth of the epithelium and mesenchyme has to be coordinated with development of the associated blood vessels, and the whole process involves a large number of additional signals. Many aspects of the system are still not understood. It is known, however, that *Drosophila* uses closely related mechanisms to govern the branching morphogenesis of its tracheal system—the tubules that form the airways of an insect. Again, the process depends on the *Drosophila* FGF protein, encoded by the *Branchless* gene, and the *Drosophila* FGF receptor, encoded by the *Breathless* gene, both operating in much the same way as in the mouse. Indeed, genetic studies of tracheal development in *Drosophila* have also identified other components of the control machinery, and the *Drosophila* genes have led us to their vertebrate homologs. Genetic manipulations in the mouse have given us the means to test whether these genes have similar functions in mammals too; and to a remarkable extent they do.

Figure 22–92 Branching morphogenesis of the lung. (A) How FGF10 and Sonic hedgehog are thought to induce the growth and branching of the buds of the bronchial tree. Many other signal molecules, such as BMP4, are also expressed in this system, and the suggested branching mechanism is only one of several possibilities. (B) A cast of the adult human bronchial tree, prepared by injecting resin into the airways; resins of different colors have been injected into different branches of the tree. (B, from R. Warwick and P.L. Williams, Gray's Anatomy, 35th ed. Edinburgh: Longman, 1973.)

Summary

The mouse has a central role as model organism for study of the molecular genetics of mammalian development. Mouse development is essentially similar to that of other vertebrates, but begins with a specialized preamble to form extraembryonic structures such as the amnion and placenta. Powerful techniques have been devised for creation of gene knockouts and other targeted genetic alterations by exploiting the highly regulative properties of the cells of the inner cell mass of the mouse embryo. These cells can be put into culture and maintained as embryonic stem cells (ES cells). Under the right culture conditions, ES cells can proliferate indefinitely without differentiating, while retaining the ability to give rise to any part of the body when injected back into an early mouse embryo.

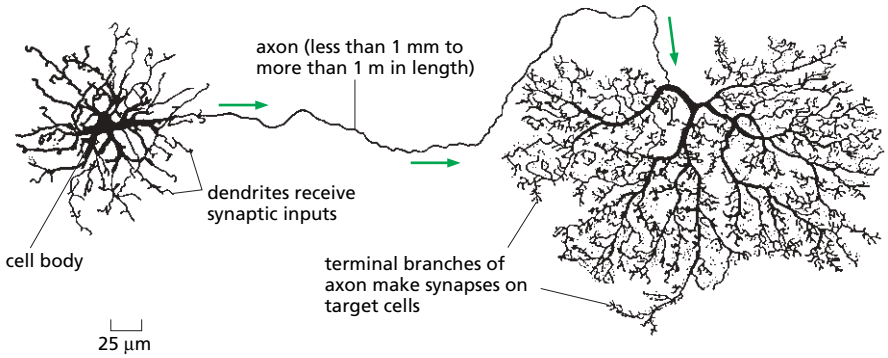


Figure 22–93 A typical neuron of a vertebrate. The arrows indicate the direction in which signals are conveyed. The neuron shown is from the retina of a monkey. The longest and largest neurons in a human extend for about 1 million μm and have an axon diameter of 15 μm. (Drawing of neuron from B.B. Boycott, in Essays on the Nervous System [R. Bellairs and E.G. Gray, eds.]. Oxford, UK: Clarendon Press, 1974.)

Many general developmental processes, including most of those discussed elsewhere in the chapter, have been illuminated by studies in the mouse. As just one example, the mouse has been used to investigate the control of branching morphogenesis. This process gives rise to structures such as lungs and glands, and is governed by exchanges of signals between mesenchyme cells and an invading epithelium. The functions of these signals can be analyzed by gene knockout experiments.

NEURAL DEVELOPMENT

Nerve cells, or **neurons**, are among the most ancient of all specialized animal cell types. Their structure is like that of no other class of cells, and the development of the nervous system poses problems that have no real parallel in other tissues. A neuron is extraordinary above all for its enormously extended shape, with a long *axon* and branching *dendrites* connecting it through synapses to other cells (**Figure 22–93**). The central challenge of neural development is to explain how the axons and dendrites grow out, find their right partners, and synapse with them selectively to create a functional network (**Figure 22–94**). The problem is formidable: the human brain contains more than 10^{11} neurons, each of which, on average, has to make connections with a thousand others, according to a regular and predictable wiring plan. The precision required is not so great as in a man-made computer, for the brain performs its computations in a different way and is more tolerant of vagaries in individual components; but the brain nevertheless outstrips all other biological structures in its organized complexity.

The components of a typical nervous system—the various classes of neurons, glial cells, sensory cells, and muscles—originate in a number of widely separate locations in the embryo and are initially unconnected. Thus, in the first phase of neural development (**Figure 22–95**), the different parts develop according to their own local programs: neurons are born and assigned specific characters according to the place and time of their birth, under the control of inductive signals and gene regulatory mechanisms similar to those we have already discussed for other tissues of the body. The next phase involves a type of morphogenesis unique to the nervous system: axons and dendrites grow out along specific routes, setting up a provisional but orderly network of connections between the separate parts of the system. In the third and final phase, which continues into adult life, the connections are adjusted and refined through interactions among the far-flung components in a way that depends on the electrical signals that pass between them.



Figure 22–94 The complex organization of nerve cell connections. This drawing depicts a section through a small part of a mammalian brain—the olfactory bulb of a dog, stained by the Golgi technique. The black objects are neurons; the thin lines are axons and dendrites, through which the various sets of neurons are interconnected according to precise rules. (From C. Golgi, *Riv. sper. freniat. Reggio-Emilia* 1:405-425, 1875; reproduced in M. Jacobson, Developmental Neurobiology, 3rd ed. New York: Plenum, 1992.)

Neurons Are Assigned Different Characters According to the Time and Place Where They Are Born

Neurons are almost always produced in association with **glial cells**, which provide a supporting framework and create an enclosed, protected environment in which the neurons can perform their functions. Both cell types, in all animals,

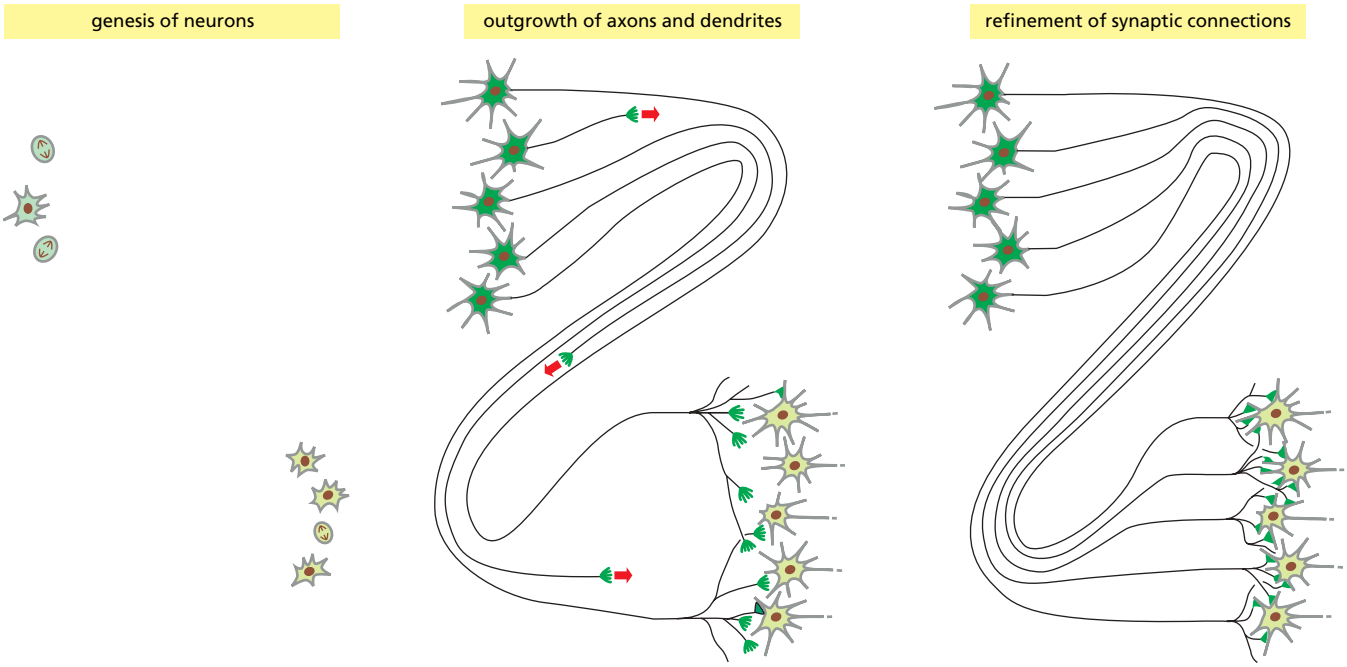


Figure 22–95 The three phases of neural development.

develop from the ectoderm, usually as sister cells or cousins derived from a common precursor. Thus, in vertebrates, the neurons and glial cells of the *central nervous system* (including the spinal cord, the brain, and the retina of the eye) derive from the part of the ectoderm that rolls up to form the neural tube, while those of the *peripheral nervous system* derive mainly from the neural crest (**Figure 22–96**).

The **neural tube**, with which we shall be mainly concerned, consists initially of a single-layered epithelium (**Figure 22–97**). The epithelial cells are the progenitors of the neurons and glia. As these cell types are generated, the epithelium becomes thickened and transformed into a more complex structure. As discussed earlier, Delta–Notch signaling controls the differentiation of the progenitor cells into neurons: the nascent neurons express Delta, and thereby inhibit their neighbors from differentiating into neurons at the same time. This ensures that the progenitors do not all differentiate simultaneously but remain as a dividing cell population from which further neurons can be generated. The progenitor and, later, glial cells also maintain the cohesiveness of the epithelium and form a scaffolding that spans its thickness. Along and between these tall cells, like animals amid the trees of the forest, the new-born neurons migrate, find their resting places, mature, and send out their axons and dendrites (**Figure 22–98**).

Signal proteins secreted from the ventral and dorsal sides of the neural tube act as opposing morphogens, causing neurons born at different dorsoventral levels to express different gene regulatory proteins (see Figure 22–80). There are differences along the head-to-tail axis as well, reflecting the anteroposterior pattern of expression of *Hox* genes and the actions of yet other morphogens. Moreover, just as in *Drosophila*, neurons continue to be generated in each region of the central nervous system over many days, weeks, or even months, and this gives rise to still greater diversity, because the cells adopt different characters

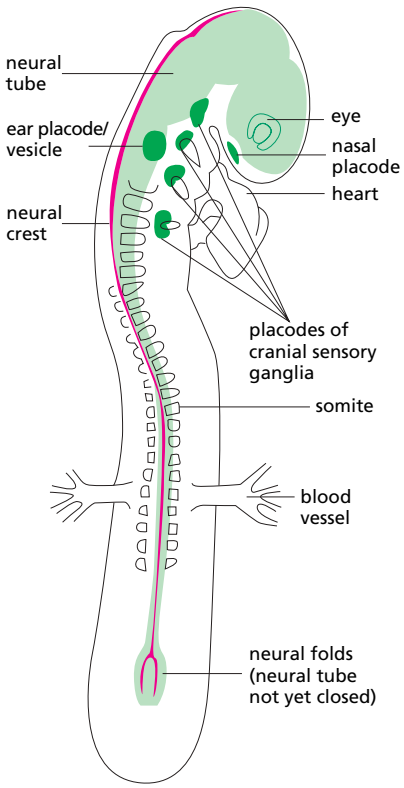


Figure 22–96 Diagram of a 2-day chick embryo, showing the origins of the nervous system. The neural tube *(light green)* has already closed, except at the tail end, and lies internally, beneath the ectoderm, of which it was originally a part (see Figure 22–78). The neural crest *(red)* lies dorsally just beneath the ectoderm, in or above the roof of the neural tube. In addition, thickenings, or placodes *(dark green),* in the ectoderm of the head give rise to some of the sensory transducer cells and neurons of that region, including those of the ear and the nose. The cells of the retina of the eye, by contrast, originate as part of the neural tube.

according to their "birthday"—the time of the terminal mitosis that marks the beginning of neuronal differentiation (**Figure 22–99**). When progenitor cells are taken from an embryonic mouse brain and maintained in culture for several days, individually isolated from their normal surroundings, they go through much the same program as in the intact tissue. That is, they divide repeatedly, producing pairs of daughters that frequently adopt different fates, such that one remains as a dividing progenitor while the other becomes committed to differentiate.

The successive divisions throw off a sequence of different neuronal and glial cell types, according to a more-or-less regular timetable. This implies that the progenitors themselves must autonomously change their intrinsic character from one cell generation to the next. The molecular mechanism of this progressive change is unknown, just as it is in other cell types where similar slow changes occur.

The Character Assigned to a Neuron at Its Birth Governs the Connections It Will Form

The differences of gene expression modulate the characters of the neurons and help to cause them to make connections with different partners. In the spinal cord, for example, ventrally located clusters of cells express genes of the *Islet/Lim* homeobox family (coding for gene regulatory proteins) and develop as motor neurons, sending out axons to connect with specific subsets of muscles—different muscles according to the particular *Islet/Lim* family members expressed. If the pattern of gene expression is artificially altered, the neurons project to different target muscles.

The different destinations reflect different pathway choices that the axons make as they grow out from the nerve cell body, as well as their selective

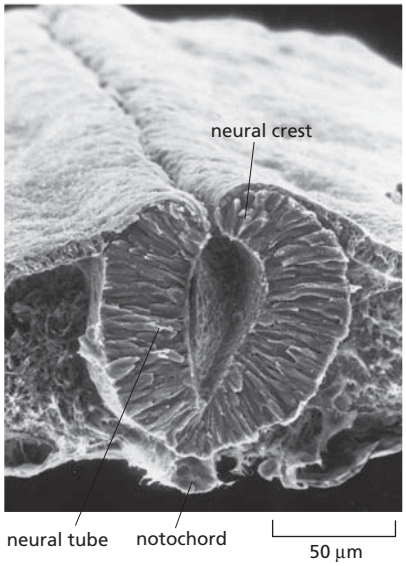


Figure 22–97 Formation of the neural tube. The scanning electron micrograph shows a cross section through the trunk of a 2-day chick embryo. The neural tube is about to close and pinch off from the ectoderm; at this stage it consists (in the chick) of an epithelium that is only one cell thick. (Courtesy of J.P. Revel and S. Brown.)

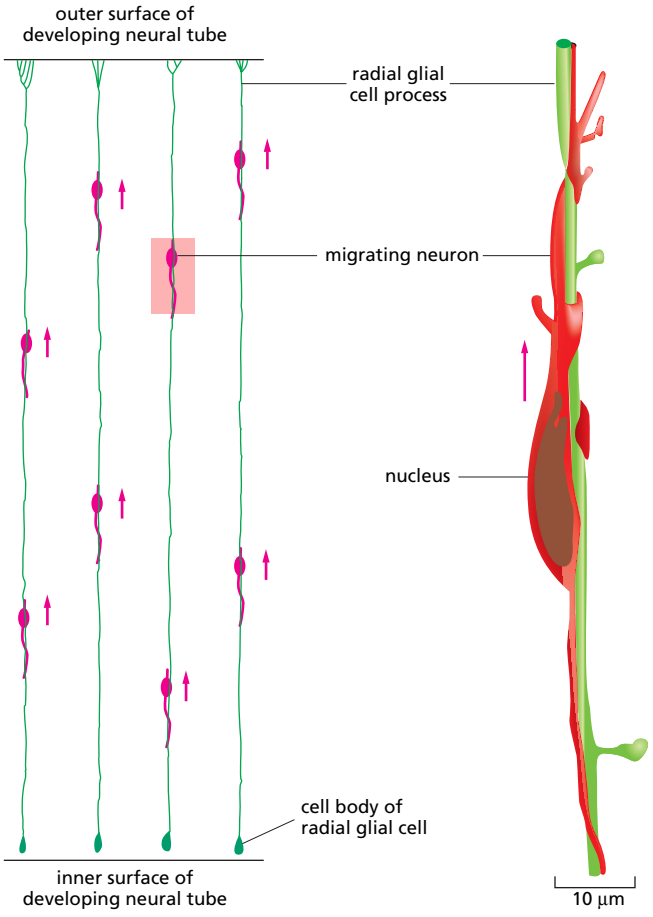


Figure 22–98 Migration of immature neurons. Before sending out axons and dendrites, newborn neurons often migrate from their birthplace and settle in some other location. The diagrams are based on reconstructions from sections of the cerebral cortex of a monkey (part of the neural tube). The neurons go through their final cell division close to the inner, luminal face of the neural tube and then migrate outward by crawling along radial glial cells. Each of these cells extends from the inner to the outer surface of the tube, a distance that may be as much as 2 cm in the cerebral cortex of the developing brain of a primate. The radial glial cells can be considered as persisting cells of the original columnar epithelium of the neural tube that become extraordinarily stretched as the wall of the tube thickens. (After P. Rakic, *J. Comp. Neurol.* 145:61–84, 1972. With permission from John Wiley & Sons, Inc.)

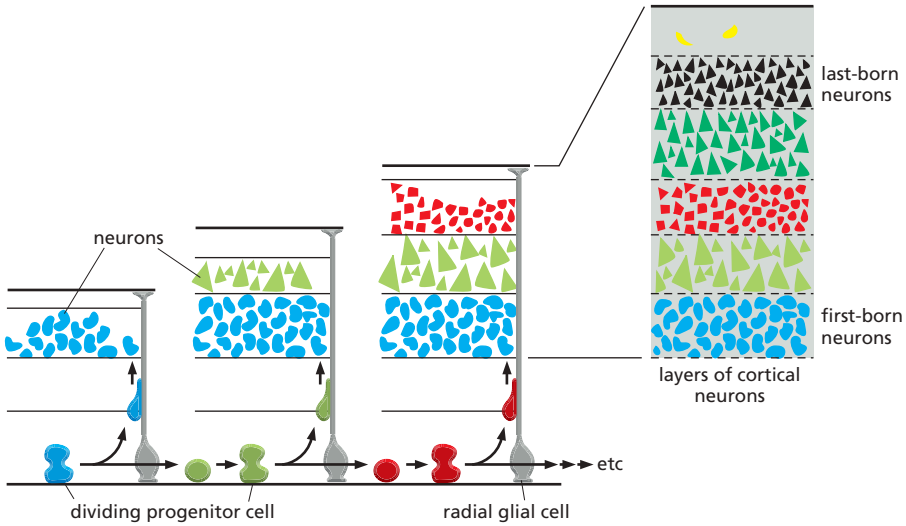


Figure 22–99 Programmed production of different types of neurons at different times from dividing progenitors in the cerebral cortex of the brain of a mammal. Close to one face of the cortical neuroepithelium, progenitor cells divide repeatedly, in stem-cell fashion, to produce neurons. The neurons migrate out toward the opposite face of the epithelium by crawling along the surfaces of radial glial cells, as shown in Figure 22–98. The first-born neurons settle closest to their birthplace, while neurons born later crawl past them to settle farther out. Successive generations of neurons thus occupy different layers in the cortex and have different intrinsic characters according to their birth dates.

recognition of different target cells at the end of the journey. In the dorsal part of the spinal cord lie neurons that receive and relay sensory information from sensory neurons in the periphery of the body. In intermediate positions, there are various other classes of interneurons, connecting specific sets of nerve cells to one another. Some send their axons dorsally, others ventrally; some up toward the head, others down toward the tail, still others across the floor of the neural tube to the other side of the body (**Figure 22–100**). In a timelapse film where the developing neurons are stained with a fluorescent dye, one can watch the movements of the growing tips of the axons as they extend: one is reminded of the lights of rush-hour traffic at night, as the cars streak along a network of highways, turning this way or that at busy junctions, each one making its own choice of route.

How are these complex movements guided? Before attempting an answer, we must examine more closely the structure of the growing neuron.

Each Axon or Dendrite Extends by Means of a Growth Cone at Its Tip

A typical neuron sends out one long axon, projecting toward a distant target to which signals are to be delivered, and several shorter dendrites, on which it mainly receives incoming signals from axon terminals of other neurons. Each

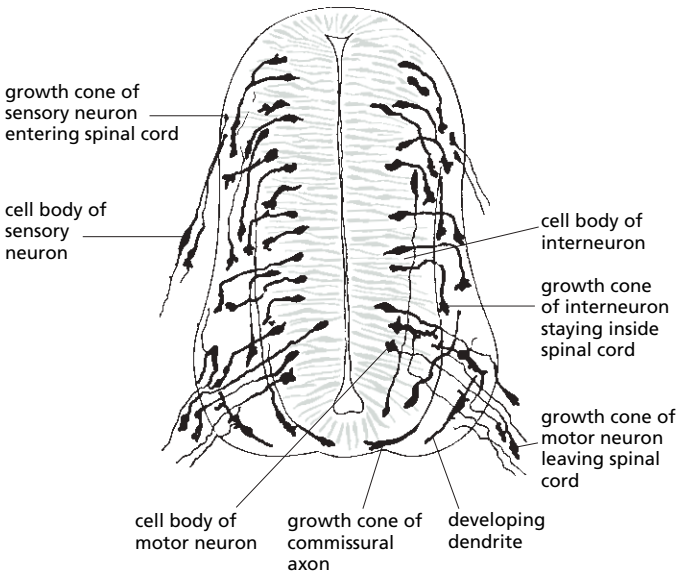


Figure 22–100 Growing axons in the developing spinal cord of a 3-day chick embryo. The drawing shows a cross section stained by the Golgi technique. Most of the neurons, apparently, have as yet only one elongated process—the future axon. An irregularly shaped expansion—a growth cone—is seen at the growing tip of each axon. The growth cones of the motor neurons emerge from the spinal cord (to make their way toward muscles), those of the sensory neurons grow into it from outside (where their cell bodies lie), and those of the interneurons remain inside the spinal cord. Many of the interneurons send their axons down toward the floor plate to cross to the other side of the spinal cord; these axons are called commissural. At this early stage, many of the embryonic spinal-cord cells (in the regions shaded *gray)* are still proliferating and have not yet begun to differentiate as neurons or glial cells. (From S. Ramón y Cajal, Histologie du Système Nerveux de l'Homme et des Vertébrés, 1909–1911. Paris: Maloine; reprinted, Madrid: C.S.I.C., 1972.)

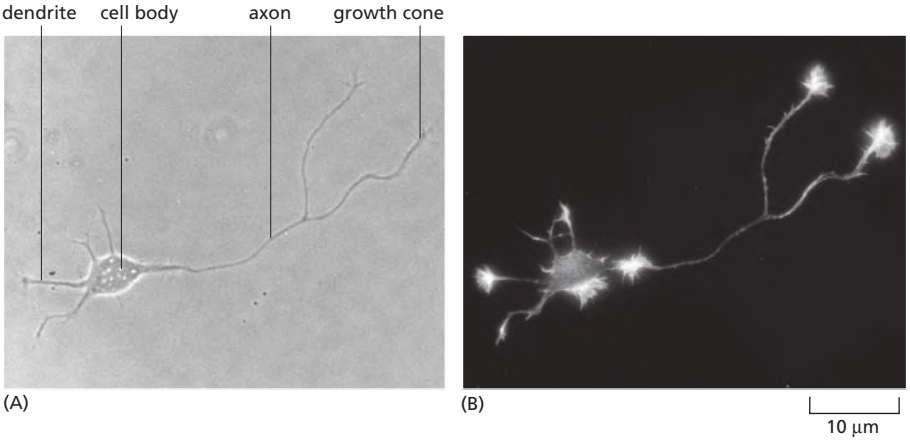


Figure 22–101 Formation of axon and dendrites in culture. A young neuron has been isolated from the brain of a mammal and put to develop in culture, where it sends out processes. One of these processes, the future axon, has begun to grow out faster than the rest (the future dendrites) and has bifurcated. (A) A phase-contrast picture; (B) the pattern of staining with fluorescent phalloidin, which binds to filamentous actin. Actin is concentrated in the growth cones at the tips of the processes that are actively extending and at some other sites of lamellipodial activity. (Courtesy of Kimberly Goslin.)

process extends by growth at its tip, where an irregular, spiky enlargement is seen. This structure, called the **growth cone**, crawls through the surrounding tissue, trailing a slender axon or dendrite behind it (see Figure 22–100). <AAGA> The growth cone comprises both the engine that produces the movement and the steering apparatus that directs the tip of each process along the proper path (see Figure 16–105).

Much of what we know about the properties of growth cones has come from studies in tissue or cell culture. One can watch as a neuron begins to put out its processes, all at first alike, until one of the growth cones puts on a sudden turn of speed, identifying its process as the axon, with its own axon-specific set of proteins (**Figure 22–101**). The contrast between axon and dendrite established at this stage involves polarized intracellular transport of different materials into the two types of process. As a result, they will grow out for different distances, follow different paths, and play different parts in synapse formation.

The growth cone at the end of a typical growing nerve cell process—either axon or dendrite—moves forward at a speed of about 1 mm per day, continually probing the regions that lie ahead and on either side by putting out filopodia and lamellipodia. When such a protrusion contacts an unfavorable surface, it withdraws; when it contacts a more favorable surface, it persists longer, steering the growth cone as a whole to move in that direction. In this way the growth cone can be guided by subtle variations in the surface properties of the substrata over which it moves. At the same time, it is sensitive to diffusible chemotactic factors in the surrounding medium, which can also encourage or hinder its advance. These behaviors depend on the cytoskeletal machinery inside the growth cone, as discussed in Chapter 16. A multitude of receptors in the growth cone membrane detect the external signals and, through the agency of intracellular regulators such as the monomeric GTPases Rho and Rac, control the assembly and disassembly of actin filaments and other components of the machinery of cell movement.

The Growth Cone Pilots the Developing Neurite Along a Precisely Defined Path *In Vivo*

In living animals, growth cones generally travel toward their targets along predictable, stereotyped routes, exploiting a multitude of different cues to find their way, but always requiring a substratum of extracellular matrix or cell surface to crawl over. Often, growth cones take routes that have been pioneered by other neurites, which they follow by contact guidance. As a result, nerve fibers in a mature animal are usually found grouped together in tight parallel bundles (called fascicles or fiber tracts). Such crawling of growth cones along axons is thought to be mediated by homophilic cell–cell adhesion molecules—membrane glycoproteins that help a cell displaying them to stick to any other cell that also displays them. As discussed in Chapter 19, two of the most important

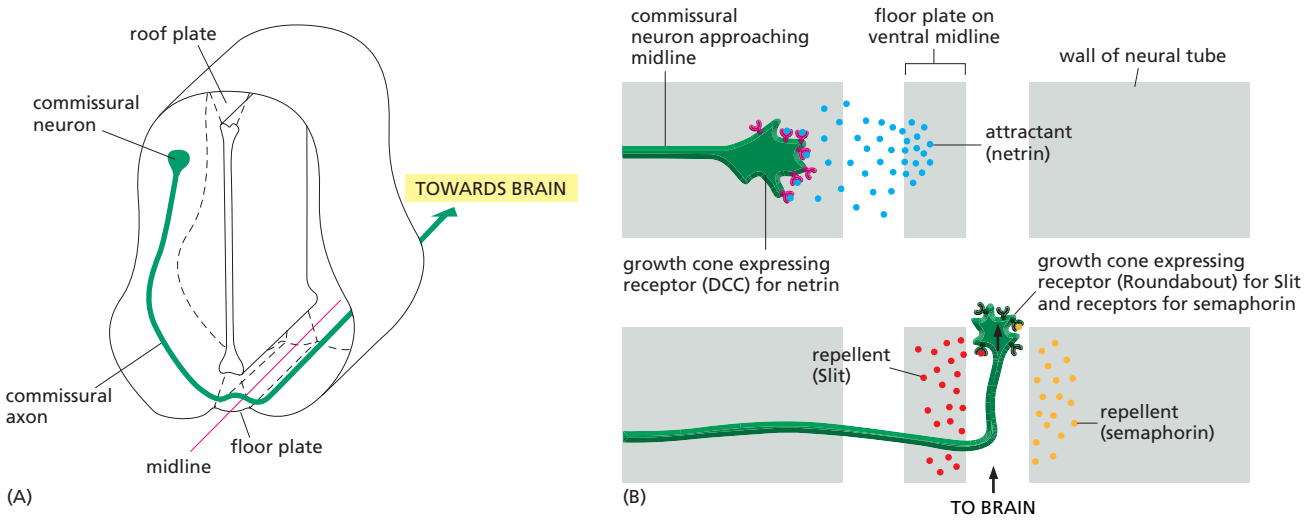


Figure 22–102 The guidance of commissural axons. (A) The pathway taken by commissural axons in the embryonic spinal cord of a vertebrate. (B) The signals that guide them. The growth cones are first attracted to the floor plate by netrin, which is secreted by the floor-plate cells and acts on the receptor DCC in the axonal membrane. As they cross the floor plate, the growth cones upregulate their expression of Roundabout, the receptor for a repellent protein, Slit, that is also secreted by the floor plate. Slit, binding to Roundabout, not only acts as a repellent to keep the cells from re-entering the floor plate, but also blocks responsiveness to the attractant netrin. At the same time, the growth cones switch on expression of receptors for another repellent protein, semaphorin, that is secreted by the cells in the side walls of the neural tube. Trapped between two repellent territories, the growth cones, having crossed the midline, travel in a tight fascicle up toward the brain.

classes of such molecules are those that belong to the immunoglobulin superfamily, such as N-CAM, and those of the Ca^{2+}-dependent cadherin family, such as N-cadherin. Members of both families are generally present on the surfaces of growth cones, of axons, and of various other cell types that growth cones crawl over, including glial cells in the central nervous system and muscle cells in the periphery of the body. The human genome contains more than 100 cadherin genes, for example, and most of them are expressed in the brain (see Figure 19–6). Different sets of cell–cell adhesion molecules, acting in varied combinations, provide a mechanism for selective neuronal guidance and recognition. Growth cones also migrate over components of the extracellular matrix. Some of the matrix molecules, such as *laminin*, favor axon outgrowth, while others, such as chondroitin sulfate proteoglycans, discourage it.

Growth cones are guided by a succession of different cues at different stages of their journey, and the stickiness of the substratum is not the only thing that matters. Another important part is played by chemotactic factors, secreted from cells that act as beacons at strategic points along the path—some attracting, others repelling. The trajectory of *commissural* axons—those that cross from one side of the body to the other—provides a beautiful example of how a combination of guidance signals can specify a complex path. Commissural axons are a general feature of bilaterally symmetrical animals, because the two sides of the body have to be neurally coordinated. Worms, flies and vertebrates use closely related mechanisms to guide their outgrowth.

In the developing spinal cord of a vertebrate, for example, a large number of neurons send their axonal growth cones ventrally toward the floor plate—a specialized band of cells forming the ventral midline of the neural tube (see Figure 22–100). The growth cones cross the floor plate and then turn abruptly through a right angle to follow a longitudinal path up toward the brain, parallel to the floor plate but never again crossing it (**Figure 22–102**A). The first stage of the journey depends on a concentration gradient of the protein *netrin*, secreted by the cells of the floor plate: the commissural growth cones sniff their way toward its source. Netrin was purified from chick embryos, by assaying extracts of neural tissue for an activity that would attract commissural growth cones in a culture dish. Its sequence revealed that it was the vertebrate homolog of a protein already known from *C. elegans*, through genetic screens for mutant worms with misguided axons—called *Unc* mutants because they move in an *unc*oordinated fashion. One of the *Unc* genes, *Unc6*, codes for the homolog of netrin. Another, *Unc40*, codes for its transmembrane receptor; and this too has a vertebrate homolog called DCC that is expressed in the commissural neurons and mediates their response to the netrin gradient.

Localized activation of DCC by netrin leads to opening of a specialized class of ion channels in the plasma membrane. These channels, called TRPC (Transient Receptor Potential C) channels, belong to a large family (the TRP family)

that is responsible for many other sensory transduction processes, from mechanosensation to the perception of heat and cold. When open, the TRPC channels allow Ca^{2+} (and other cations) to enter the cell. The localized rise in Ca^{2+} then activates the machinery for extension of filopodia and movement of the growth cone toward the netrin source.

The receptors on each growth cone determine the route it will take: non-commissural neurons in the neural tube, lacking DCC, are not attracted to the floor plate; and neurons expressing a different netrin receptor—called Unc5H in vertebrates (with a counterpart Unc5 in the worm)—are actively repelled by the floor plate and send their axons instead toward the roof plate.

Growth Cones Can Change Their Sensibilities as They Travel

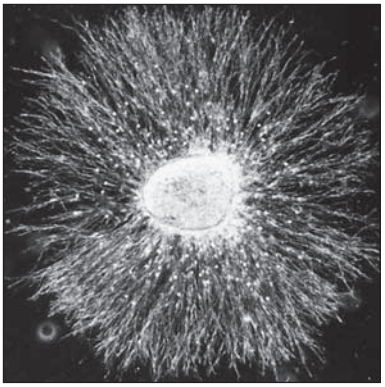
If commissural growth cones are attracted to the floor plate, why do they cross it and emerge on the other side, instead of staying in the attractive territory? And having crossed it, why do they never veer back onto it again? The likely answer lies in another set of molecules, several of which are also conserved between vertebrates and invertebrates. Studies of *Drosophila* mutants with misguided commissural axons first identified three of the key proteins: Slit, Roundabout, and Commissureless.

Slit, like netrin, is produced by midline cells of the developing fly, while its receptor, Roundabout, is expressed in the commissural neurons. Slit, acting on Roundabout, has an effect exactly opposite to that of netrin: it repels the growth cones, blocking entry to the midline territory. Commissureless, however, interferes with the delivery of Roundabout to the cell surface and thereby makes the growth cones initially blind to this "keep-out" signal. Commissural growth cones in this state advance to the midline; as they cross it, they seem, by some mechanism that we do not yet understand, to lose their blindfold of Commissureless protein and begin to be repelled. Emerging on the far side, they now have functional Roundabout on their surfaces and are thereby prohibited from re-entry.

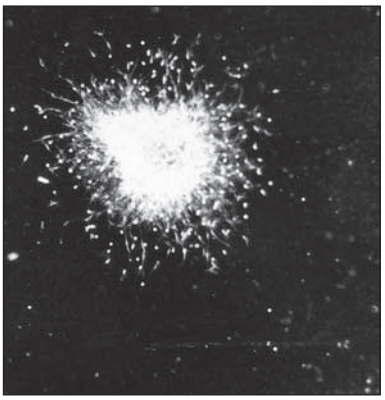
In vertebrates, a similar mechanism operates, involving homologs of Slit and Roundabout. Commissural growth cones are at first attracted to the midline, and then somehow change their surface receptor proteins as they cross; in this way they switch their sensibilities, gaining sensitivity to repulsion by Slit—which is expressed in the floor plate—and losing sensitivity to attraction by netrin. Sensitivity to Slit in the initial approach to the midline is blocked not by any homolog of Commissureless but by a divergent member of the Roundabout receptor family called Rig1, which sits in the plasma membrane and interferes with signal reception by its cousins. The Rig1 block is switched off by some unknown mechanism once the growth cones have crossed the midline. Repulsion from the midline now prevents them from straying back across it. At the same time, the growth cones apparently become sensitive to another set of repulsive signals, in the form of proteins called semaphorins, which prevent them from traveling back up into the dorsal regions of the spinal cord. Trapped between the two sets of repulsive signals, the growth cones have no choice but to travel in a narrow track, running parallel to the floor plate but never re-entering it (Figure 22–102B).

Target Tissues Release Neurotrophic Factors That Control Nerve Cell Growth and Survival

Eventually, axonal growth cones reach the target region where they must halt and make synapses. The neurons that sent out the axons can now begin to communicate with their target cells. Although synapses generally transmit signals in one direction, from axon to either dendrite or muscle, the developmental communications are a two-way affair. Signals from the target tissue not only regulate which growth cones are to synapse where (as we discuss below), but also how many of the innervating neurons are to survive.



NGF



control

Figure 22–103 NGF effects on neurite outgrowth. Dark-field photomicrographs of a sympathetic ganglion cultured for 48 hours with *(above)* or without *(below)* NGF. Neurites grow out from the sympathetic neurons only if NGF is present in the medium. Each culture also contains Schwann (glial) cells that have migrated out of the ganglion; these are not affected by NGF. Neuronal survival and maintenance of growth cones for neurite extension represent two distinct effects of NGF. The effect on growth cones is local, direct, rapid, and independent of communication with the cell body; when NGF is removed, the deprived growth cones halt their movements within a minute or two. The effect of NGF on cell survival is less immediate and is associated with uptake of NGF by endocytosis and its intracellular transport back to the cell body. (Courtesy of Naomi Kleitman.)

Most types of neurons in the vertebrate central and peripheral nervous system are produced in excess; up to 50% or more of them then die soon after they reach their target, even though they appear perfectly normal and healthy up to the time of their death. About half of all the motor neurons that send axons to skeletal muscle, for example, die within a few days after making contact with their target muscle cells. A similar proportion of the sensory neurons that innervate the skin die after their growth cones have arrived there.

This large-scale death of neurons is thought to reflect the outcome of a competition. Each type of target cell releases a limited amount of a specific neurotrophic factor that the neurons innervating that target require to survive. The neurons apparently compete for the factor, and those that do not get enough die by programmed cell death. If the amount of target tissue is increased—for example, by grafting an extra limb bud onto the side of the embryo—more limb-innervating neurons survive; conversely, if the limb bud is cut off, the limb-innervating neurons all die. In this way, although individuals may vary in their bodily proportions, they always retain the right number of motor neurons to innervate all their muscles and the right number of sensory neurons to innervate their whole body surface. The seemingly wasteful strategy of overproduction followed by death of surplus cells operates in almost every region of the nervous system. It provides a simple and effective means to adjust each population of innervating neurons according to the amount of tissue requiring innervation.

The first neurotrophic factor to be identified, and still the best characterized, is known simply as nerve growth factor, or NGF—the founding member of the *neurotrophin* family of signal proteins. It promotes the survival of specific classes of sensory neurons derived from the neural crest and of sympathetic neurons (a subclass of peripheral neurons that control contractions of smooth muscle and secretion from exocrine glands). NGF is produced by the tissues that these neurons innervate. When extra NGF is provided, extra sensory and sympathetic neurons survive, just as if extra target tissue were present. Conversely, in a mouse with a mutation that knocks out the gene for NGF or for its receptor (a transmembrane tyrosine kinase called TrkA), almost all sympathetic neurons and the NGF-dependent sensory neurons are lost. There are many neurotrophic factors, only a few of which belong to the neurotrophin family, and they act in different combinations to promote survival of different classes of neurons.

NGF and its relatives have an additional role: besides acting on the nerve cell as a whole to control its survival, they regulate the outgrowth of axons and dendrites (**Figure 22–103**). These can even act locally on just one part of the tree of nerve cell processes, promoting or pruning the growth of individual branches: a growth cone exposed to NGF shows an immediate increase of motility. Conversely, an axon branch that is deprived of NGF, while the rest of the neuron continues to be bathed in the factor, dies back.

The peripheral action of NGF continues to be important after the phase of neuronal death. In the skin, for example, it controls the branching of sensory nerve fibers, ensuring not only that the whole body surface becomes innervated during development but also that it recovers its innervation after damage.

Neuronal Specificity Guides the Formation of Orderly Neural Maps

In many cases, axons originating from neurons of a similar type but located in different positions come together for the journey and arrive at the target in a tight bundle. There they disperse again, to terminate at different sites in the target territory.

The projection from the eye to the brain provides an important example. <TACC> The neurons in the retina that convey visual information back to the brain are called *retinal ganglion cells*. There are more than a million of them, each one reporting on a different part of the visual field. Their axons converge on the optic nerve head at the back of the eye and travel together along the optic stalk into the brain. Their main site of termination, in most vertebrates other than mammals, is the *optic tectum*—a broad expanse of cells in the midbrain. In connecting with the tectal neurons, the retinal axons distribute themselves in a predictable pattern according to the arrangement of their cell bodies in the retina: ganglion cells that are neighbors in the retina connect with target cells that are neighbors in the tectum. The orderly projection creates a **map** of visual space on the tectum (**Figure 22–104**).

Orderly maps of this sort are found in many brain regions. In the auditory system, for example, neurons project from the ear to the brain in a tonotopic order, creating a map in which brain cells receiving information about sounds of different pitch are ordered along a line, like the keys of a piano. And in the somatosensory system, neurons conveying information about touch map onto the cerebral cortex so as to mark out a "homunculus"—a small, distorted, two-dimensional image of the body surface (**Figure 22–105**).

The retinotopic map of visual space in the optic tectum is the best characterized of all these maps. How does it arise? In principle, the growth cones could be physically channeled to different destinations as a consequence of their different starting positions, like drivers on a multilane highway where it is forbidden to change lanes. This possibility was tested in the visual system by a famous experiment in the 1940s. If the optic nerve of a frog is cut, it will regenerate. The retinal axons grow back to the optic tectum, restoring normal vision. If, in addition, the eye is rotated in its socket at the time of cutting of the nerve, so as to put originally ventral retinal cells in the position of dorsal retinal cells, vision is still restored, but with an awkward flaw: the animal behaves as though it sees the world upside down and left–right inverted. This is because the misplaced retinal

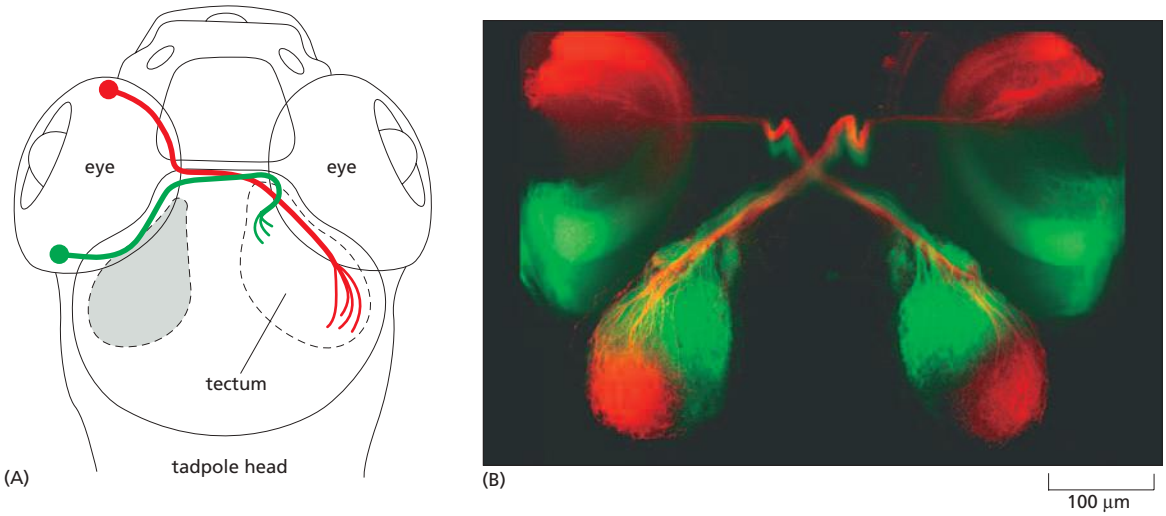


Figure 22–104 The neural map from eye to brain in a young zebrafish. (A) Diagrammatic view, looking down on the top of the head. (B) Fluorescence micrograph. Fluorescent tracer dyes have been injected into each eye—*red* into the anterior part, *green* into the posterior part. The tracer molecules have been taken up by the neurons in the retina and carried along their axons, revealing the paths they take to the optic tectum in the brain and the map that they form there. (Courtesy of Chi-Bin Chien, from D.H. Sanes, T.A. Reh and W.A. Harris, Development of the Nervous System. San Diego, CA: Academic Press, 2000.)

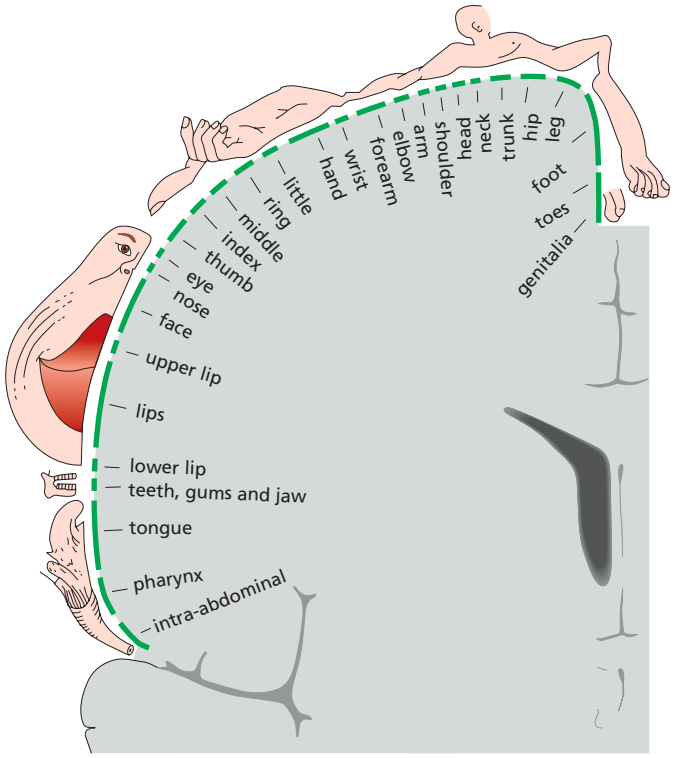


Figure 22–105 A map of the body surface in the human brain. The surface of the body is mapped onto the somatosensory region of the cerebral cortex by an orderly system of nerve cell connections, such that sensory information from neighboring body sites is delivered to neighboring sites in the brain. This means that the map in the brain is largely faithful to the topology of the body surface, even though different body regions are represented at different magnifications according to their density of innervation. The homunculus (the "little man" in the brain) has big lips, for example, because the lips are a particularly large and important source of sensory information. The map was determined by stimulating different points in the cortex of conscious patients during brain surgery and recording what they said they felt. (After W. Penfield and T. Rasmussen, The Cerebral Cortex of Man. New York: Macmillan, 1950.)

cells make the connections appropriate to their original, not their actual, positions. It seems that the cells have positional values—position-specific biochemical properties representing records of their original location. As a result, cells on opposite sides of the retina are intrinsically different, just as the motor neurons in the spinal cord that project to different muscles are intrinsically different.

Such nonequivalence among neurons is referred to as *neuronal specificity*. It is this intrinsic characteristic that guides the retinal axons to their appropriate target sites in the tectum. Those target sites themselves are distinguishable by the retinal axons because the tectal cells also carry positional labels. Thus, the neuronal map depends on a correspondence between two systems of positional markers, one in the retina and the other in the tectum.

Axons From Different Regions of the Retina Respond Differently to a Gradient of Repulsive Molecules in the Tectum

Axons from the nasal retina (the side closest to the nose) project to the posterior tectum, and axons from the temporal retina (the side farthest from the nose) project to the anterior tectum, with intermediate regions of retina projecting to intermediate regions of tectum. When nasal and temporal axons are allowed to grow out over a carpet of anterior or posterior tectal membranes in a culture dish, they also show selectivity (**Figure 22–106**). Temporal axons strongly prefer the anterior tectal membranes, as *in vivo*, whereas nasal axons either prefer posterior tectal membranes, or show no preference (depending on the species of animal). The key difference between anterior and posterior tectum appears to be a repulsive factor on the posterior tectum, to which temporal retinal axons are sensitive but nasal retinal axons are not: if a temporal retinal growth cone touches posterior tectal membrane, it collapses its filopodia and withdraws.

Assays based on these phenomena *in vitro* have identified some of the molecules responsible. The repulsive factor on posterior tectal membrane seems to be partly or entirely comprised of *ephrinA* proteins, a subset of the family of GPI-linked proteins that act as ligands for the *EphA* family of tyrosine kinase receptors. In the mouse, two different ephrins are expressed to form an anterior-to-posterior gradient on the tectal cells. Anterior cells have little or no

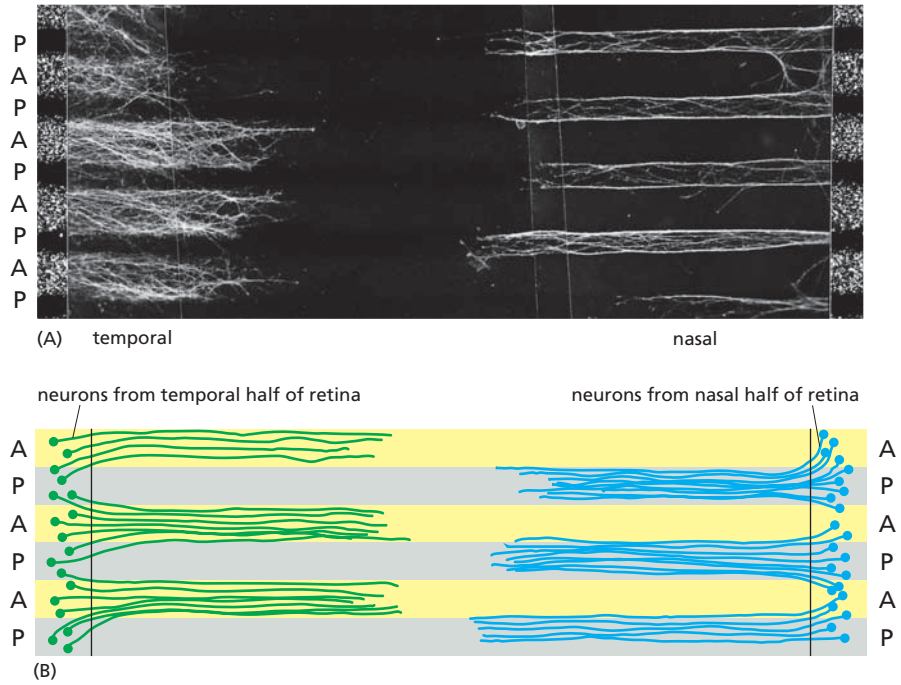


Figure 22–106 Selectivity of retinal axons growing over tectal membranes. (A) A photograph of the experimental observation. (B) A diagram of what is happening. The culture substratum has been coated with alternating stripes of membrane prepared either from posterior tectum (P) or from anterior tectum (A). In the photograph, the anterior tectal stripes are made visible by staining them with a fluorescent marker in the vertical strips at the sides of the picture. Axons of neurons from the temporal half of the retina (growing in from the left) follow the stripes of anterior tectal membrane but avoid the posterior tectal membrane, while axons of neurons from the nasal half of the retina (growing in from the right) do the converse. Thus anterior tectum differs from posterior tectum and nasal retina from temporal retina, and the differences guide selective axon outgrowth. These experiments were performed with cells from the chick embryo. (From Y. von Boxberg, S. Deiss and U. Schwarz, *Neuron* 10:345–357, 1993. With permission from Elsevier.)

ephrin, cells in the center of the tectum express ephrin A2, and cells at the posterior edge of the tectum express ephrin A2 and ephrin A5. Thus, there is a gradient of ephrin expression across the tectum. Meanwhile, the incoming axons express Eph receptors, also in a gradient: temporal axons express high Eph levels, making them sensitive to repulsion by ephrinA, whereas nasal axons express low Eph levels. In a similar way, distributed across the other main axis of the tectum, from medial to lateral, there is graded expression of ephrinB protein and also of another type of signal molecule, Wnt3, with correspondingly graded expression of EphB receptors and Wnt receptors along the dorso-ventral axis of the retina.

This system of signals and receptors is enough to produce an orderly two-dimensional map, if we make one further assumption—an assumption supported by experiments *in vivo*: that the retinal axons somehow interact with one another and compete for tectal territory. Thus, temporal axons are restricted to anterior tectum, and drive nasal axons off it; nasal axons, consequently, are restricted to posterior tectum. Between the extremes, a balance is struck, creating a smooth map of the temporo-nasal axis of the retina onto the anteroposterior axis of the tectum.

Diffuse Patterns of Synaptic Connections Are Sharpened by Activity-Dependent Remodeling

In a normal animal the retinotectal map is initially fuzzy and imprecise: the system of matching markers we have just described is enough to define the broad layout of the map, but not sufficient to specify its fine details. Studies in frogs and fish show that each retinal axon at first branches widely in the tectum and makes a profusion of synapses, distributed over a large area of tectum that overlaps with the territories innervated by other axons. These territories are subsequently trimmed back by selective elimination of synapses and retraction of axon branches. This is accompanied by the formation of new sprouts, through which each axon develops a denser distribution of synapses in the territory that it retains.

A central part in this remodeling and refinement of the map is played by two competition rules that jointly help to create spatial order: (1) axons from separate regions of retina, which tend to be excited at different times, compete to dominate the available tectal territory, but (2) axons from neighboring sites in

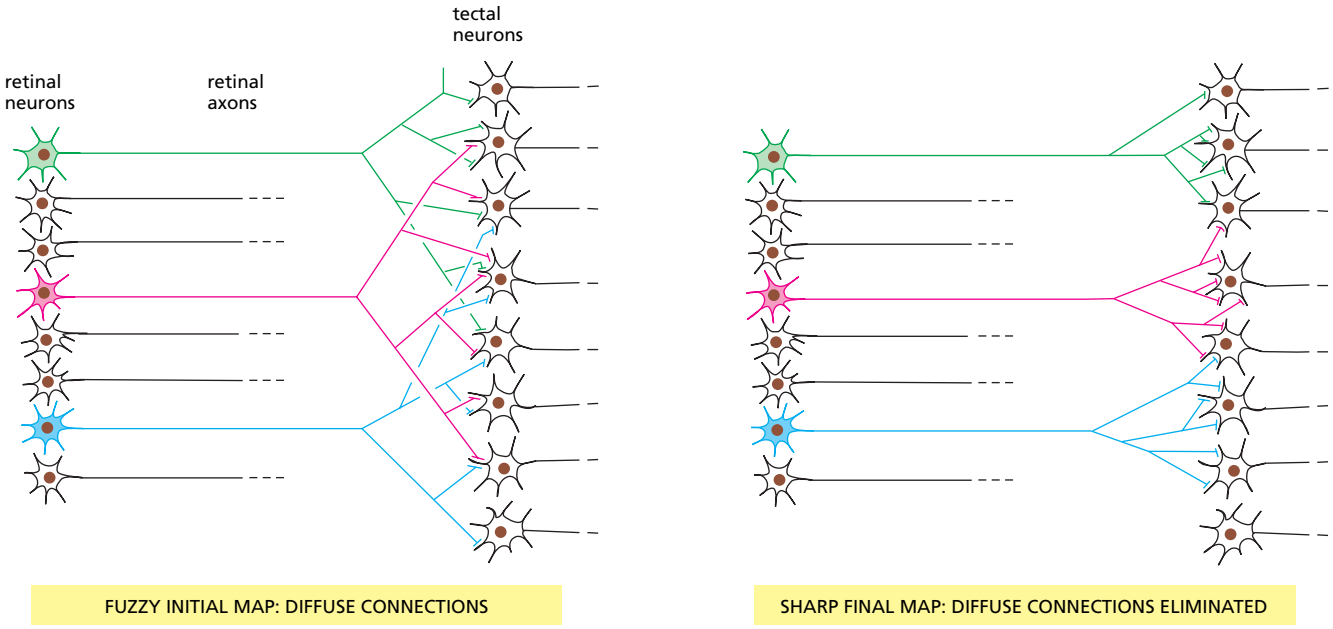


Figure 22–107 Sharpening of the retinotectal map by synapse elimination. At first the map is fuzzy because each retinal axon branches widely to innervate a broad region of tectum overlapping the regions innervated by other retinal axons. The map is then refined by synapse elimination. Where axons from separate parts of the retina synapse on the same tectal cell, competition occurs, eliminating the connections made by one of the axons. But axons from cells that are close neighbors in the retina cooperate, maintaining their synapses on shared tectal cells. Thus each retinal axon ends up innervating a small tectal territory, adjacent to and partly overlapping the territory innervated by axons from neighboring sites in the retina.

the retina, which tend to be excited at the same time, innervate neighboring territories in the tectum because they collaborate to retain and strengthen their synapses on shared tectal cells (**Figure 22–107**). The mechanism underlying both these rules depends on electrical activity and signaling at the synapses that are formed. If all action potentials are blocked by a toxin that binds to voltage-gated Na^+ channels, synapse remodeling is inhibited and the map remains fuzzy.

The phenomenon of activity-dependent synapse elimination is encountered in almost every part of the developing vertebrate nervous system. Synapses are first formed in abundance and distributed over a broad target field; then the system of connections is pruned back and remodeled by competitive processes that depend on electrical activity and synaptic signaling. The elimination of synapses in this way is distinct from the elimination of surplus neurons by cell death, and it occurs after the period of normal neuronal death is over.

Much of what we know about the cellular mechanisms of synapse formation and elimination comes from experiments on the innervation of skeletal muscle in vertebrate embryos. A two-way exchange of signals between the nerve axon terminals and the muscle cells controls the initial formation of synapses. At sites of contact, acetylcholine receptors are clustered in the muscle cell membrane and the apparatus for secretion of this neurotransmitter becomes organized in the axon terminals (discussed in Chapter 11). Each muscle cell at first receives synapses from several neurons; but in the end, through a process that typically takes a couple of weeks, it is left innervated by only one. The synapse retraction again depends on synaptic communication: if synaptic transmission is blocked by a toxin that binds to the acetylcholine receptors in the muscle cell membrane, the muscle cell retains its multiple innervation beyond the normal time of elimination.

Experiments on the musculoskeletal system, as well as in the retinotectal system, suggest that it is not only the amount of electrical activity at a synapse that is important for its maintenance, but also its temporal coordination. Whether a synapse is strengthened or weakened seems to depend critically on whether or not activity in the presynaptic cell is synchronized with activity of the other presynaptic cells synapsing on the same target (and thus also synchronized with activity of the target cell itself).

These and many other findings have suggested a simple interpretation of the competition rules for synapse elimination in the retinotectal system (**Figure 22–108**). Axons from different parts of the retina fire at different times and so compete. Each time one of them fires, the synapse(s) made by the other on a shared tectal target cell are weakened, until one of the axons is left in sole command of

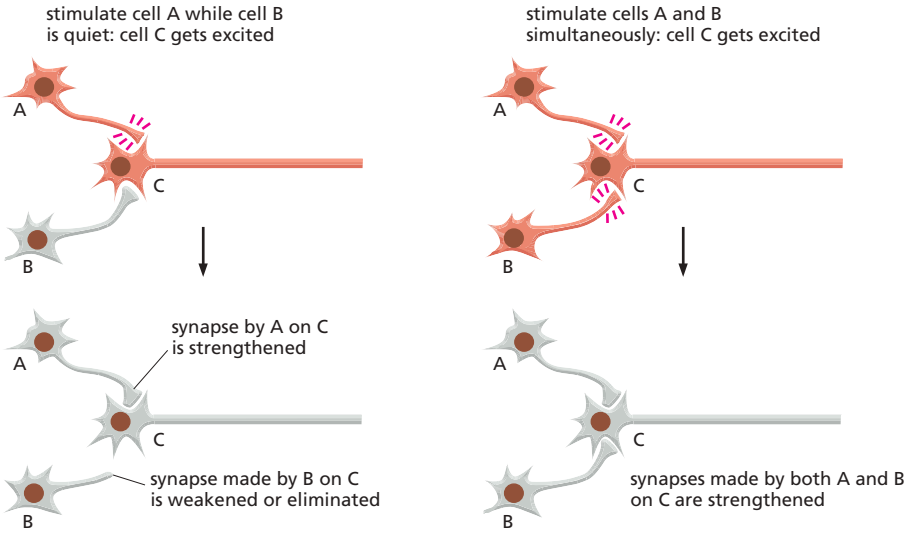


Figure 22–108 Synapse modification and its dependence on electrical activity. Experiments in several systems indicate that synapses are strengthened or weakened by electrical activity according to the rule shown in the diagram. The underlying principle appears to be that each excitation of a target cell tends to weaken any synapse where the presynaptic axon terminal has just been quiet but to strengthen any synapse where the presynaptic axon terminal has just been active. As a result, "neurons that fire together, wire together." A synapse that is repeatedly weakened and rarely strengthened is eventually eliminated altogether.

that cell. Axons from neighboring retinal cells, on the other hand, tend to fire in synchrony with one another: they therefore do not compete but instead maintain synapses on shared tectal cells, creating a precisely ordered map in which neighboring cells of the retina project to neighboring sites in the tectum.

Experience Molds the Pattern of Synaptic Connections in the Brain

The phenomenon that we have just described is summed up in the catch-phrase "neurons that fire together, wire together". The same firing rule relating synapse maintenance to neural activity helps to organize our developing brains in the light of experience.

In the brain of a mammal, axons relaying inputs from the two eyes are brought together in a specific cell layer in the visual region of the cerebral cortex. Here, they form two overlapping maps of the external visual field, one as perceived through the right eye, the other as perceived through the left. Although there is some evidence of a tendency for right- and left-eye inputs to be segregated even before synaptic communication begins, a large proportion of the axons carrying information from the two eyes at early stages synapse together on shared cortical target cells. A period of early signaling activity, however, occurring spontaneously and independently in each retina even before vision begins, leads to a clean segregation of inputs, creating stripes of cells in the cortex that are driven by inputs from the right eye alternating with stripes that are driven by inputs from the left eye (**Figure 22–109**). The firing rule suggests a simple interpretation: a pair of axons bringing information from neighboring sites in the left eye will frequently fire together, and therefore wire

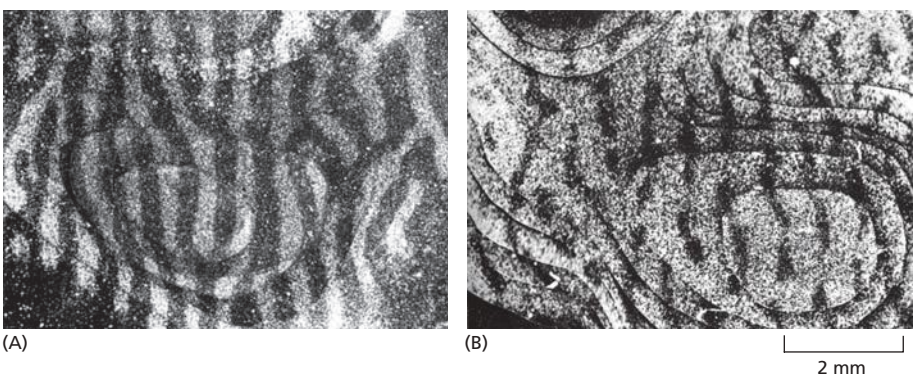


Figure 22–109 Ocular dominance columns in the visual cortex of a monkey's brain, and their sensitivity to visual experience. (A) Normally, stripes of cortical cells driven by the right eye alternate with stripes, of equal width, driven by the left eye. The stripes are revealed here by injecting a radioactive tracer molecule into one eye, allowing time for this tracer to be transported to the visual cortex, and detecting radioactivity there by autoradiography, in sections cut parallel to the cortical surface. (B) If one eye is kept covered during the critical period of development, and thus deprived of visual experience, its stripes shrink and those of the active eye expand. In this way, the deprived eye may lose the power of vision almost entirely. (From D.H. Hubel, T.N. Wiesel and S. Le Vay, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 278:377–409, 1977. With permission from The Royal Society.)

together, as will a pair of axons from neighboring sites in the right eye; but a right-eye axon and a left-eye axon will rarely fire together, and will instead compete. Indeed, if activity from both eyes is silenced using drugs that block action potentials or synaptic transmission, the inputs fail to segregate correctly.

Maintenance of the pattern of connections is extraordinarily sensitive to experience early in life. If, during a certain *critical period* (ending at about the age of 5 years in humans), one eye is kept covered for a time so as to deprive it of visual stimulation, while the other eye is allowed normal stimulation, the deprived eye loses its synaptic connections to the cortex and becomes almost entirely, and irreversibly, blind. In accordance with what the firing rule would predict, a competition has occurred in which synapses in the visual cortex made by inactive axons are eliminated while synapses made by active axons are consolidated. In this way cortical territory is allocated to axons that carry information and is not wasted on those that are silent.

In establishing the nerve connections that enable us to see, it is not only the quantity of visual stimulation that is important, but also its temporal coordination. For example, the ability to see depth—stereo vision—depends on cells in other layers of the visual cortex that receive inputs relayed from both eyes at once, conveying information about the same part of the visual field as seen from two slightly different angles. These binocularly driven cells allow us to compare the view through the right eye with that through the left so as to derive information about the relative distances of objects from us. If, however, the two eyes are prevented during the critical period from ever seeing the same scene at the same time—for example, by covering first one eye and then the other on alternate days, or simply as a consequence of a childhood squint—almost no binocularly driven cells are retained in the cortex, and the capacity for stereo perception is irretrievably lost. Evidently, in accordance with the firing rule, the inputs from each eye to a binocularly driven neuron are maintained only if the two inputs are frequently triggered to fire in synchrony, as occurs when the two eyes look together at the same scene.

Adult Memory and Developmental Synapse Remodeling May Depend on Similar Mechanisms

We saw in Chapter 11 that synaptic changes underlying memory in at least some parts of the adult brain, notably the hippocampus, hinge on the behavior of a particular type of receptor for the neurotransmitter glutamate—the NMDA receptor. Ca^{2+} flooding into the postsynaptic cell through the channels opened by this receptor triggers lasting changes in the strengths of the synapses on that cell, affecting the presynaptic as well as the postsynaptic structures. The changes that are induced by the NMDA-dependent mechanism in the adult brain obey rules closely akin to the developmental firing rule: events in the external world that cause two neurons to be active at the same time, or in quick succession, favor the making or strengthening of synapses between them. This condition, called the *Hebb rule*, has been suggested to be the fundamental principle underlying associative learning.

Is it possible, then, that adult learning and the more drastic forms of synaptic plasticity seen during development both depend on the same basic machinery of synapse adjustment? There are many hints that it may be so. For example, inhibitors that specifically block activation of the NMDA receptor interfere with the refinement and remodeling of synaptic connections in the developing visual system. Both in the developing animal and in the adult, the alterations in the strength of the synaptic connections correspond to changes in physical structure. The scale of these physical changes is, however, very different. In the developing organism, electrical activity often regulates the extension and regression of large branches of the axonal and dendritic trees. But in the adult brain, the structural adjustments occurring in response to activity seem typically to be much more finely localized, affecting the sizes of individual dendritic spines—the tiny knob-shaped protrusions, no more than a few micrometers long, on which dendrites receive individual synapses (**Figure 22–110**). It seems that Ca^{2+}

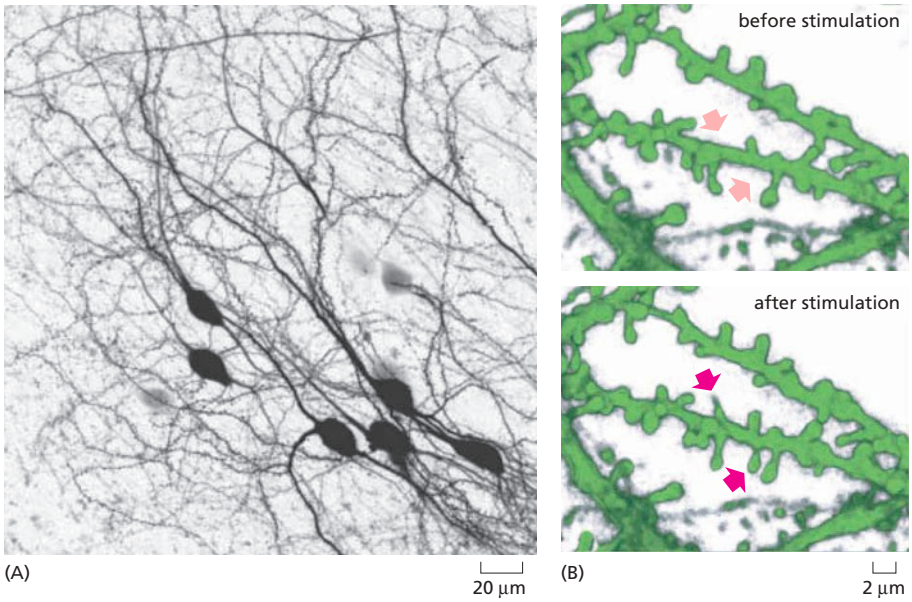


Figure 22–110 Growth of dendritic spines in response to synaptic stimulation. (A) Neurons in a slice of living tissue from the hippocampus of a young mouse. The cells are labeled by expression of Green Fluorescent Protein and observed with a two-photon laser scanning microscope, which allows individual dendrites to be seen at high resolution. The insert shows a processed image of a small part of some of the dendrites. These are covered with tiny dendritic spines, which are the sites of synapses. (B) Repeated intense bursts of synaptic stimulation, triggered by a nearby microelectrode, cause new spines to form within 30 minutes. Low-frequency stimulation has an opposite effect, causing a subset of spines to regress. (From U.V. Nägerl, N. Eberhorn, S.B. Cambridge and T. Bonhoeffer, *Neuron* 44:759–767, 2004. With permission from Elsevier.)

entering a spine through NMDA channels in response to excitation of the synapse on that particular spine can cause just that spine to remodel its actin cytoskeleton. But we still have a lot to learn about the mechanism of such changes and their relationship to learning and memory. The molecular basis of the processes of synapse remodeling through which experience molds our brains remains one of the central challenges that the nervous system presents to cell biology.

Summary

The development of the nervous system proceeds in three phases: first, nerve cells are generated through cell division; then, having ceased dividing, they send out axons and dendrites to form profuse synapses with other, remote cells so that communication can begin; last, the system of synaptic connections is refined and remodeled according to the pattern of electrical activity in the neural network.

The neurons, and the glial cells that always accompany them, are generated from ectodermal precursors, and those born at different times and places express different sets of genes, which help to determine the connections they will form. Axons and dendrites grow out from the neurons by means of growth cones, which follow specific pathways delineated by signals along the way. Structures such as the floor plate of the embryonic spinal cord secrete both chemoattractants and chemorepellents, to which growth cones from different classes of neurons respond differently. On reaching their target area, the axons terminate selectively on a subset of the accessible cells, and in many parts of the nervous system neural maps are set up—orderly projections of one array of neurons onto another. In the retinotectal system, the map is based on the matching of complementary systems of position-specific cell-surface markers—ephrins and Eph receptors—possessed by the two sets of cells.

After the growth cones have reached their targets and initial connections have formed, two major sorts of adjustment occur. First, many of the innervating neurons die as a result of a competition for survival factors such as NGF (nerve growth factor) secreted by the target tissue. This cell death adjusts the quantity of innervation according to the size of the target. Second, individual synapses are pruned away in some places and reinforced in others, so as to create a more precisely ordered pattern of connections. This latter process depends on electrical activity: synapses that are frequently active are reinforced, and different neurons contacting the same target cell tend to maintain their synapses on the shared target only if they are both frequently active at the same time. In this way the structure of the brain can be adjusted to reflect the connections between events in the external world. The underlying molecular mechanism of this synaptic plasticity may be similar to that responsible for the formation of memories in adult life.

PLANT DEVELOPMENT

Plants and animals are separated by about 1.5 billion years of evolutionary history. They have evolved their multicellular organization independently but using the same initial tool kit—the set of genes inherited from their common unicellular eucaryotic ancestor. Most of the contrasts in their developmental strategies spring from two basic peculiarities of plants. First, they get their energy from sunlight, not by ingesting other organisms. This dictates a body plan different from that of animals. Second, their cells are encased in semirigid cell walls and cemented together, preventing them from moving as animal cells do. This dictates a different set of mechanisms for shaping the body and different developmental processes to cope with a changeable environment.

Animal development is largely buffered against environmental changes, and the embryo generates the same genetically determined body structure unaffected by external conditions. The development of most plants, by contrast, is dramatically influenced by the environment. Because they cannot match themselves to their environment by moving from place to place, plants adapt instead by altering the course of their development. Their strategy is opportunistic. A given type of organ—a leaf, a flower, or a root, say—can be produced from the fertilized egg by many different paths according to environmental cues. A begonia leaf pegged to the ground may sprout a root; the root may throw up a shoot; the shoot, given sunlight, may grow leaves and flowers.

The mature plant is typically made of many copies of a small set of standardized modules, as described in **Figure 22–111**. The positions and times at which those modules are generated are strongly influenced by the environment, causing the overall structure of the plant to vary. The choices between alternative modules and their organization into a whole plant depend on external cues and long-range hormonal signals that play a much smaller part in the control of animal development.

But although the global structure of a plant—its pattern of roots or branches, its numbers of leaves or flowers—can be highly variable, its detailed organization on a small scale is not. A leaf, a flower, or indeed an early plant embryo, is as precisely specified as any organ of an animal, possessing a *determinate* structure, in contrast with the *indeterminate* pattern of branching and sprouting of the plant as a whole. The internal organization of a plant module raises essentially the same problems in the genetic control of pattern formation as does animal development, and they are solved in analogous ways. In this section we focus on the cellular mechanisms of development in flowering plants. We examine both the contrasts and the similarities with animals.

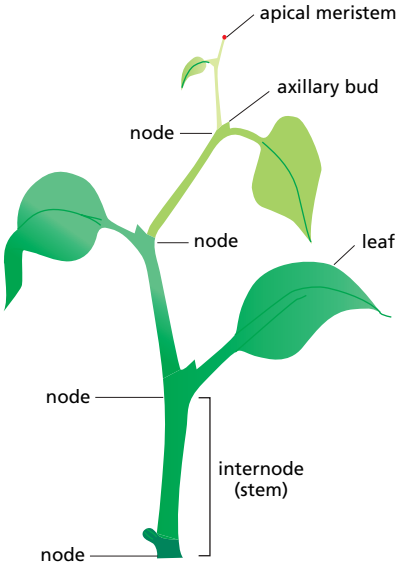


Figure 22–111 A simple example of the modular construction of plants. Each module (shown in different shades of *green)* consists of a stem, a leaf, and a bud containing a potential growth center, or meristem. The bud forms at the branch point, or node, where the leaf diverges from the stem. Modules arise sequentially from the continuing activity of the apical meristem.

Arabidopsis Serves as a Model Organism for Plant Molecular Genetics

Flowering plants, despite their amazing variety, are of relatively recent origin. The earliest known fossil examples are 130 million years old, as against 350 million years or more for vertebrate animals. Underlying the diversity of form, therefore, there is a high degree of similarity in molecular mechanisms. As we shall see, a small genetic change can transform a plant's large-scale structure; and just as plant physiology allows survival in many different environments, so also it allows survival of many differently structured forms. A mutation that gives an animal two heads is generally lethal; one that doubles the number of flowers or branches on a plant is generally not.

To identify the genes that govern plant development and to discover how they function, plant biologists have selected a small weed, the common wall cress *Arabidopsis thaliana* (**Figure 22–112**) as their primary model organism.



Figure 22–112 *Arabidopsis thaliana*. This small plant is a member of the mustard (or crucifer) family (see also Figure 1–46). It is a weed of no economic use but of great value for genetic studies of plant development. (From M.A. Estelle and C.R. Somerville, *Trends Genet.* 12:89–93, 1986. With permission from Elsevier.)

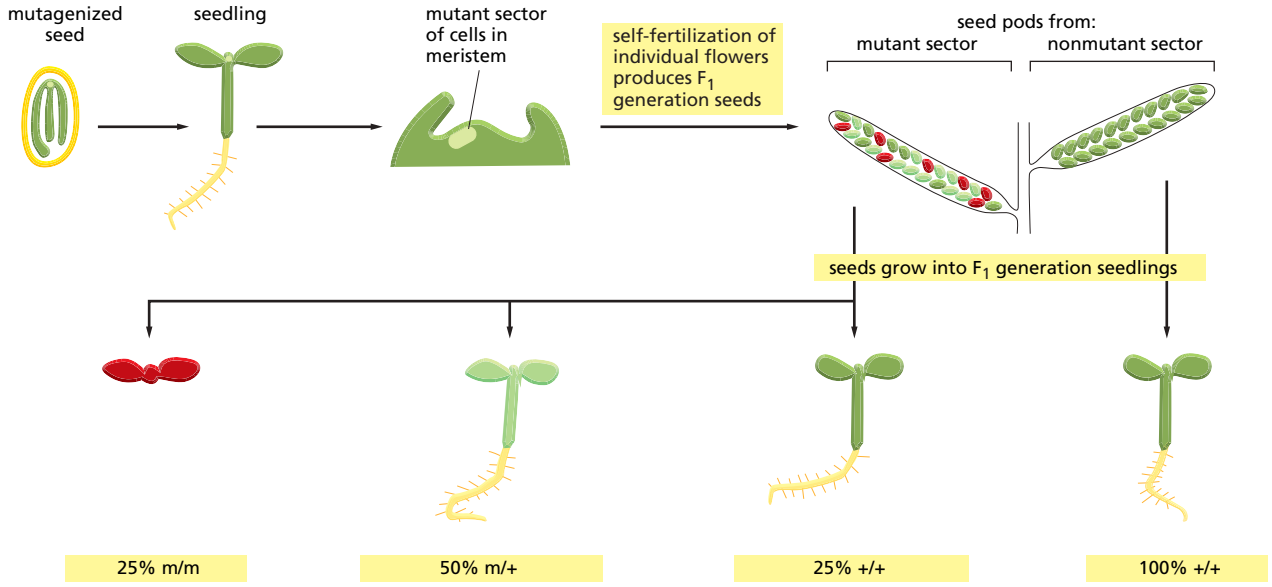


Figure 22–113 Production of mutants in *Arabidopsis*. A seed, containing a multicellular embryo, is treated with a chemical mutagen and left to grow into a plant. In general, this plant will be a mosaic of clones of cells carrying different induced mutations. An individual flower produced by this plant will usually be composed of cells belonging to the same clone, all carrying the same mutation, m, in heterozygous form (m/+). Self-fertilization of individual flowers by their own pollen results in seed pods, each of which contains a family of embryos of whose members half, on average, will be heterozygous (m/+), one-quarter will be homozygous mutant (m/m), and one-quarter will be homozygous wild-type (+/+). Often, the mutation will have a recessive lethal effect, as indicated here by the lack of a root in the m/m seedling. The mutant stock is then maintained by breeding from the heterozygote: it will produce seed pods (F_2 generation) that all contain a mixture of +/+, m/+, and m/m seeds.

Like *Drosophila* or *Caenorhabditis elegans*, it is small, quick to reproduce, and convenient for genetics. It can be grown indoors in Petri dishes or tiny plant pots in large numbers and produces hundreds of seeds per plant after 8–10 weeks. It has, in common with *C. elegans*, a significant advantage over *Drosophila* or vertebrate animals for genetics: like many flowering plants, it can reproduce as a hermaphrodite because a single flower produces both eggs and the male gametes that can fertilize them. Therefore, when a flower that is heterozygous for a recessive lethal mutation is self-fertilized, one-fourth of its seeds will display the homozygous embryonic phenotype. This makes it easy to perform genetic screens (**Figure 22–113**) and so to obtain a catalog of the genes required for specific developmental processes.

The *Arabidopsis* Genome Is Rich in Developmental Control Genes

Arabidopsis has one of the smallest plant genomes—125 million nucleotide pairs, on a par with *C. elegans* and *Drosophila*—and the complete DNA sequence is now known. It contains approximately 26,000 genes. This total includes many recently generated duplicates, however, so that the number of functionally distinct types of protein represented may be considerably less. Cell culture and genetic transformation methods have been established, as well as vast libraries of seeds carrying mutations produced by random insertions of mobile genetic elements, so that plants with mutations in any chosen gene can be obtained to order. Powerful tools are thus available to analyze gene functions. Although only a small fraction of the total gene set has been characterized experimentally as yet, functions can be tentatively assigned to many genes—about 18,000—on the basis of their sequence similarities to well-characterized genes in *Arabidopsis* and other organisms.

Even more than the genomes of multicellular animals, the *Arabidopsis* genome is rich in genes that code for gene regulatory proteins (**Table 22–2**). Some major families of animal gene regulatory proteins (such as the Myb family of DNA-binding proteins) are greatly expanded, while others (such as nuclear hormone receptors) seem to be entirely absent, and there are large families of gene regulatory proteins in the plant that have no animal homologs.

Where homologous gene regulatory proteins (such as homeodomain proteins) can be recognized in both plants and animals, they have little in common with regard to the genes they regulate or the types of developmental decisions that they control, and there is very little conservation of protein sequence outside the DNA-binding domains.

Arabidopsis is like multicellular animals in possessing many genes for cell communication and signal transduction (1900 genes out of 18,000 classified),

Table 22–2 Some Major Families of Gene Regulatory Proteins in *Arabidopsis, Drosophila, C. elegans,* and the Yeast *Saccharomyces cerevisiae*

FAMILY	NUMBER OF FAMILY MEMBERS PREDICTED FROM GENOME ANALYSIS			
	Arabidopsis	*Drosophila*	*C. elegans*	YEAST
Myb	190	6	3	10
AP2/EREBP (Apetala2/ethylene-responsive-element binding protein)	144	0	0	0
bHLH (basic helix–loop–helix)	139	46	25	8
NAC	109	0	0	0
C2H2 (Zn finger)	105	291	139	53
Homeobox	89	103	84	9
MADS box	82	2	2	4
bZIP	81	21	25	21
WRKY (Zn finger)	72	0	0	0
GARP	56	0	0	0
C2C2 (Zn finger)/GATA	104	6	9	10
Nuclear hormone receptor	0	21	25	0
C6 (Zn finger)	0	0	0	52
Estimated total (including many not listed above)	1533	635	669	209
% of genes in genome	5.9	4.5	3.5	3.5

The Table lists only those families that have at least 50 members in at least one organism. (Data from J.L. Riechmann et al., *Science* 290:2105–2110, 2000. With permission from AAAS.)

but the specific details of these gene sets are very different, as discussed in Chapter 15. The Wnt, Hedgehog, Notch, and TGFβ signaling mechanisms are all absent in *Arabidopsis*. In compensation, other signaling pathways peculiar to plants are highly developed. Cell-surface receptors of the tyrosine kinase class seem to be entirely absent, although many of the signaling components downstream of these receptors in animals are present. Conversely, receptors of the serine/threonine kinase class are very plentiful, but they do not act through the same system of intracellular messengers as the receptor serine/threonine kinases in animals. Substantial sets of genes are devoted to developmental processes of special importance in plants: more than 1000 for synthesis and remodeling of the plant cell wall, for example, and more than 100 for detecting and responding to light.

We must now examine how the genes of the plant are used to control plant development.

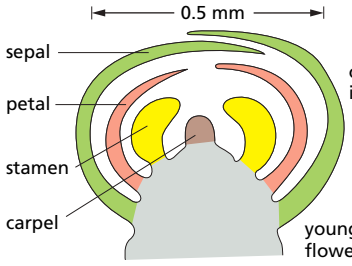
Embryonic Development Starts by Establishing a Root–Shoot Axis and Then Halts Inside the Seed

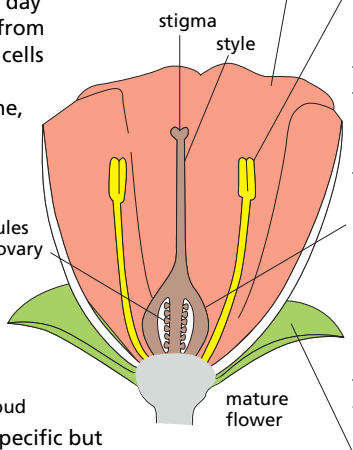
The basic strategy of sexual reproduction in flowering plants is briefly summarized in **Panel 22–1**. The fertilized egg, or zygote, of a higher plant begins by dividing asymmetrically to establish the polarity of the future embryo. One product of this division is a small cell with dense cytoplasm, which will become the embryo proper. The other is a large vacuolated cell that divides further and forms a structure called the *suspensor*, which in some ways is comparable to the umbilical cord in mammals. The suspensor attaches the embryo to the adjacent nutritive tissue and provides a pathway for the transport of nutrients.

During the next step in development the diploid embryo cell proliferates to form a ball of cells that quickly acquires a polarized structure. This comprises two key groups of proliferating cells—one at the suspensor end of the embryo that will collaborate with the uppermost suspensor cell to generate a root, and one at the opposite end that will generate a shoot (**Figure 22–114**). The main root–shoot axis established in this way is analogous to the head-to-tail axis of an animal. At the same time it begins to be possible to distinguish the future *epidermal cells*, forming the outermost layer of the embryo, the future *ground*

THE FLOWER

Flowers, which contain the reproductive cells of higher plants, arise from vegetative shoot apical meristems, where they terminate further vegetative growth. Environmental factors, often the rhythms of day length and temperature, trigger the switch from vegetative to floral development. The germ cells thus arise late in plant development from somatic cells rather than from a germ-cell line, as in animals.

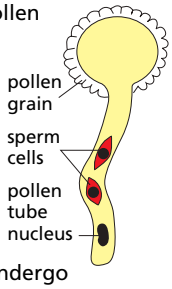




Flower structure is both varied and species-specific but generally comprises four concentrically arranged sets of structures that may each be regarded as modified leaves.

Petal: distinctive leaflike structures, usually brightly colored, facilitate pollination via, for example, attracted insects.

Stamen: an organ containing cells that undergo meiosis and form haploid pollen grains, each of which contains two male sperm cells. Pollen transferred to a stigma germinates, and the pollen tube delivers the two nonmotile sperm to the ovary.

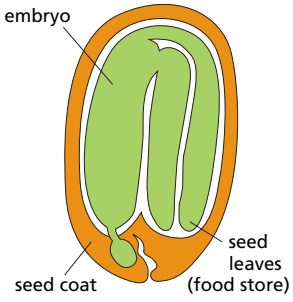


Carpel: an organ containing one or more ovaries, each of which contains ovules. Each ovule houses cells that undergo meiosis and form an embryo sac containing the female egg cell. At fertilization, one sperm cell fuses with the egg cell and will form the future diploid embryo, while the other fuses with two cells in the embryo sac to form the triploid endosperm tissue.

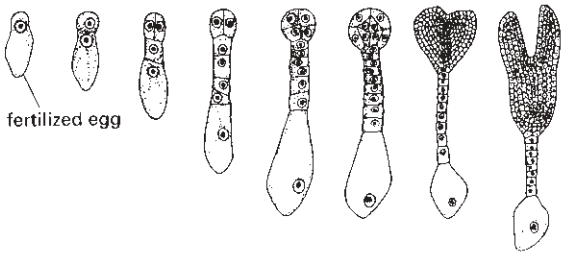
Sepals: leaflike structures that form a protective covering during early flower development.

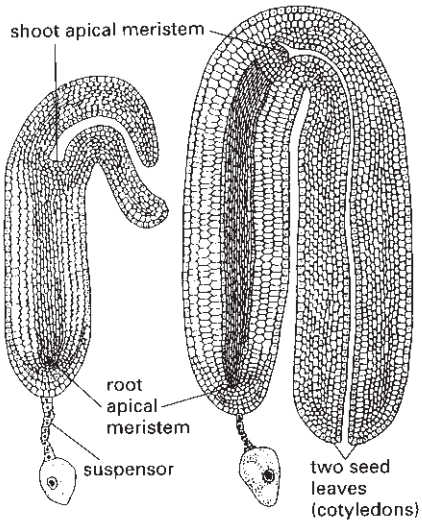
THE SEED

A seed contains a dormant embryo, a food store, and a seed coat. By the end of its development a seed's water content can drop from 90% to 5%. The seed is usually protected in a *fruit* whose tissues are of maternal origin.



THE EMBRYO





The fertilized egg within the ovule will grow to form an embryo using nutrients transported from the endosperm by the suspensor. A complex series of cell divisions, illustrated here for the common weed called shepherd's purse, produces an embryo with a root apical meristem, a shoot apical meristem, and either one (monocots) or two (dicots) seed leaves, called cotyledons.

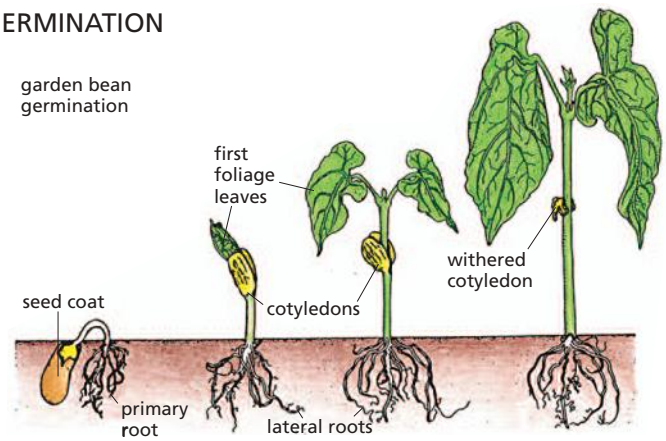
Development is arrested at this stage, and the ovule, containing the embryo, now becomes a seed, adapted for dispersal and survival.

For the embryo to resume its growth the seed must germinate, a process dependent upon both internal factors (dormancy) and environmental factors including water, temperature, and oxygen. The food reserves for the early phase of germination may either be the endosperm (maize) or the cotyledons (pea and bean).

The primary root usually emerges first from the seed to ensure an early water supply for the seedling. The cotyledon(s) may appear above the ground, as in the garden bean shown here, or they may remain in the soil, as in peas. In both cases the cotyledons eventually wither away.

The apical meristem can now show its capacity for continuous growth, producing a typical pattern of nodes, internodes, and buds (see Figure 22–106).

GERMINATION



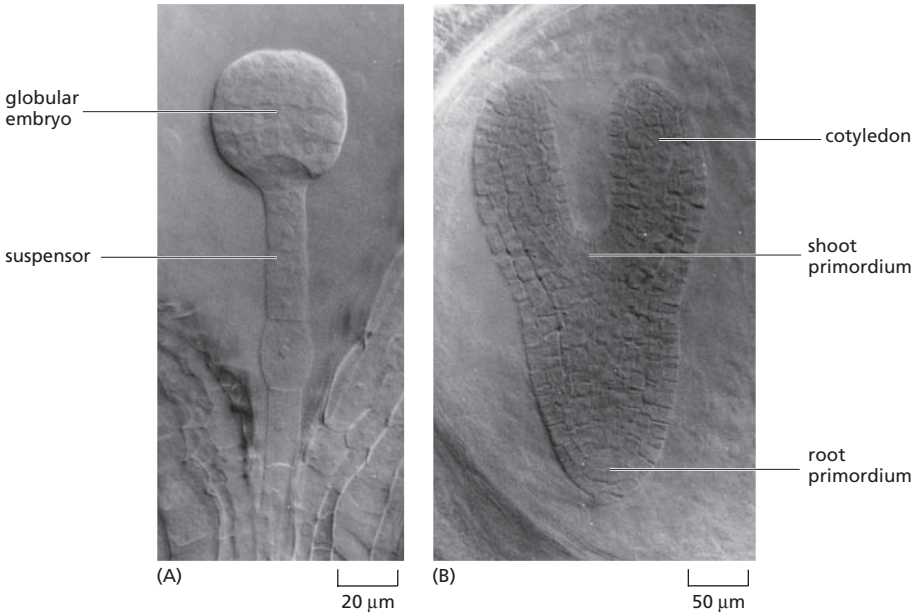


Figure 22–114 Two stages of embryogenesis in *Arabidopsis thaliana*. (From G. Jürgens et al., *Development [Suppl.]* 1:27–38, 1991. With permission from The Company of Biologists.)

tissue cells, occupying most of the interior, and the future *vascular tissue cells*, forming the central core (**Panel 22–2**). These three sets of cells can be compared to the three germ layers of an animal embryo. Slightly later in development, the rudiment of the shoot begins to produce the embryonic seed leaves, or *cotyledons*—one in monocots and two in dicots. Soon after this stage, development usually halts and the embryo becomes packaged in a **seed** (a case formed by tissues of the mother plant), specialized for dispersal and for survival in harsh conditions. The embryo in a seed is stabilized by dehydration, and it can remain dormant for a very long time—even hundreds of years. When rehydrated, the seeds germinate and embryonic development resumes.

Genetic screens can be used in *Arabidopsis*, just as in *Drosophila* or *C. elegans*, to identify the genes that govern the organization of the embryo and to group these into categories according to their homozygous mutant phenotypes. Some are required for formation of the seedling root, some for the seedling stem, and some for the seedling apex with its cotyledons. Another class is required for formation of the three major tissue types—epidermis, ground tissue, and vascular tissue—and yet another class for the organized changes of cell shape that give the embryo and seedling their elongated form (**Figure 22–115**).

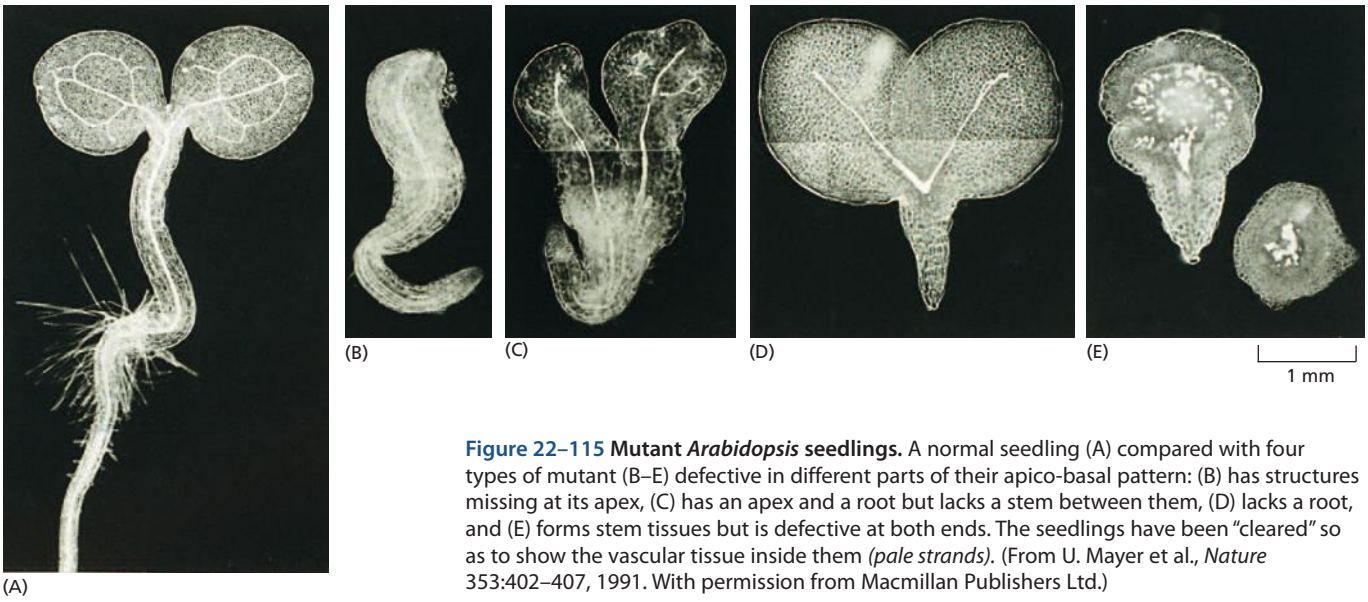


Figure 22–115 Mutant *Arabidopsis* seedlings. A normal seedling (A) compared with four types of mutant (B–E) defective in different parts of their apico-basal pattern: (B) has structures missing at its apex, (C) has an apex and a root but lacks a stem between them, (D) lacks a root, and (E) forms stem tissues but is defective at both ends. The seedlings have been "cleared" so as to show the vascular tissue inside them *(pale strands)*. (From U. Mayer et al., *Nature* 353:402–407, 1991. With permission from Macmillan Publishers Ltd.)

The Parts of a Plant Are Generated Sequentially by Meristems

Roughly speaking, the embryo of an insect or a vertebrate animal is a rudimentary miniature scale model of the later organism, and the details of body structure are filled in progressively as it enlarges. The plant embryo grows into an adult in a quite different way: the parts of the adult plant are created sequentially by groups of cells that proliferate to lay down additional structures at the plant's periphery. These all-important groups of cells are called **apical meristems** (see Figure 22–111). Each meristem consists of a self-renewing population of stem cells. As these divide, they leave behind a trail of progeny that become displaced from the meristem region, enlarge, and finally differentiate. Although the shoot and root apical meristems generate all the basic varieties of cells that are needed to build leaves, roots, and stems, many cells outside the apical meristems also keep a capacity for further proliferation and retain meristem potential. In this way trees and other perennial plants, for example, are able to increase the girth of their stems and roots as the years go by and can sprout new shoots from dormant regions if the plant is damaged.

The rudiments of the apical meristems of root and shoot are already determined in the embryo. As soon as the seed coat ruptures during germination, a dramatic enlargement of nonmeristematic cells occurs, driving the emergence first of a root, to establish an immediate foothold in the soil, and then of a shoot (**Figure 22–116**). This is accompanied by rapid and continual cell divisions in the apical meristems: in the apical meristem of a maize root, for example, cells divide every 12 hours, producing 5×10^5 cells per day. The rapidly growing root and shoot probe the environment—the root increasing the plant's capacity for taking up water and minerals from the soil, the shoot increasing its capacity for photosynthesis (see Panel 22–1).



Figure 22–116 A seedling of *Arabidopsis*. The *brown* objects to the right of the young seedling are the two halves of the discarded seed coat. (Courtesy of Catherine Duckett.)

Development of the Seedling Depends on Environmental Signals

From germination onward, the course of plant development is powerfully influenced by signals from the environment. The shoot has to push its way rapidly up through the soil, and must open its cotyledons and begin photosynthesis only after it has reached the light. The timing of this transition from rapid subterranean sprouting to illuminated growth cannot be genetically programmed, because the depth at which the seed is buried is unpredictable. The developmental switch is controlled instead by light, which, among other effects, acts on the seedling by inhibiting production of a class of plant growth regulators called *brassinosteroids*, discussed in Chapter 15. Mutations in genes required for production or reception of the brassinosteroid signal cause the stem of the seedling to go green, slow its elongation, and open its cotyledons prematurely, while it is still in the dark.

Long-Range Hormonal Signals Coordinate Developmental Events in Separate Parts of the Plant

Separate parts of a plant experience different environments and react to them individually by changes in their mode of development. The plant, however, must continue to function as a whole. This demands that developmental choices and events in one part of the plant affect developmental choices elsewhere. There must be long-range signals to bring about such coordination.

As gardeners know, for example, by pinching off the tip of a branch one can stimulate side growth: removal of the apical meristem relieves the quiescent axillary meristems of an inhibition and allows them to form new shoots. In this case the long-range signal from the apical meristem, or at least a key component has been identified. It is an auxin, a member of one of several classes of **plant growth regulators** (sometimes called *plant hormones*), all of which have powerful influences on plant development. Other known classes include the *gibberellins*, the *cytokinins*, *abscisic acid*, the gas *ethylene*, and the *brassinosteroids*. As shown in **Figure 22–117**, all are small molecules that readily penetrate cell walls. They are

THE THREE TISSUE SYSTEMS

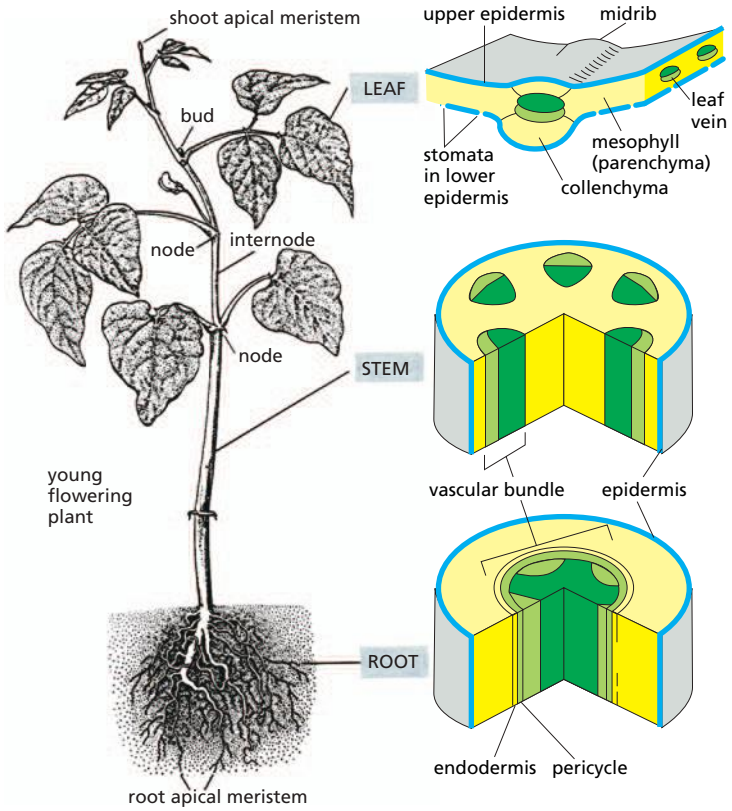
Cell division, growth, and differentiation give rise to tissue systems with specialized functions.

DERMAL TISSUE (▭): This is the plant's protective outer covering in contact with the environment. It facilitates water and ion uptake in roots and regulates gas exchange in leaves and stems.

VASCULAR TISSUE: Together the phloem (▭) and the xylem (▭) form a continuous vascular system throughout the plant. This tissue conducts water and solutes between organs and also provides mechanical support.

GROUND TISSUE (▭): This packing and supportive tissue accounts for much of the bulk of the young plant. It also functions in food manufacture and storage.

THE PLANT



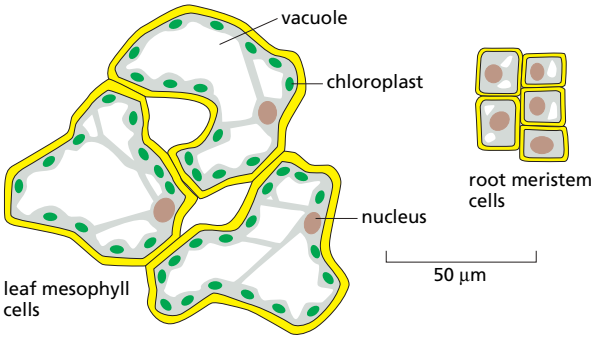
The young flowering plant shown on the *right* is constructed from three main types of organs: leaves, stems, and roots. Each plant organ in turn is made from three tissue systems: ground (▭), dermal (▭), and vascular (▭).

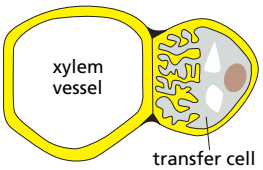
All three tissue systems derive ultimately from the cell proliferative activity of the shoot or root apical meristems, and each contains a relatively small number of specialized cell types. These three common tissue systems, and the cells that comprise them, are described in this panel.

GROUND TISSUE

The ground tissue system contains three main cell types called parenchyma, collenchyma, and sclerenchyma.

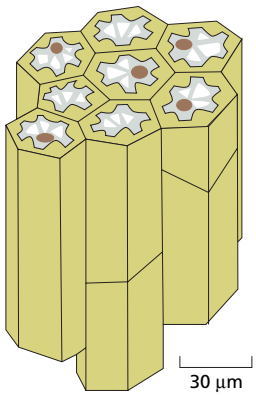
Parenchyma cells are found in all tissue systems. They are living cells, generally capable of further division, and have a thin primary cell wall. These cells have a variety of functions. The apical and lateral meristematic cells of shoots and roots provide the new cells required for growth. Food production and storage occur in the photosynthetic cells of the leaf and stem (called mesophyll cells); storage parenchyma cells form the bulk of most fruits and vegetables. Because of their proliferative capacity, parenchyma cells also serve as stem cells for wound healing and regeneration.

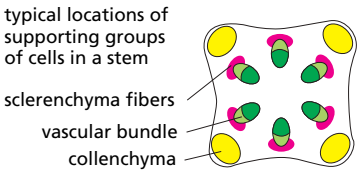




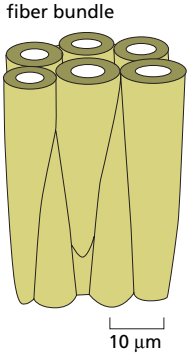
A transfer cell, a specialized form of the parenchyma cell, is readily identified by elaborate ingrowths of the primary cell wall. The increase in the area of the plasma membrane beneath these walls facilitates the rapid transport of solutes to and from cells of the vascular system.

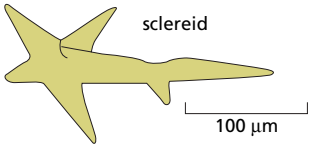
Collenchyma are living cells similar to parenchyma cells except that they have much thicker cell walls and are usually elongated and packed into long ropelike fibers. They are capable of stretching and provide mechanical support in the ground tissue system of the elongating regions of the plant. Collenchyma cells are especially common in subepidermal regions of stems.





Sclerenchyma, like collenchyma, have strengthening and supporting functions. However, they are usually dead cells with thick, lignified secondary cell walls that prevent them from stretching as the plant grows. Two common types are fibers, which often form long bundles, and sclereids, which are shorter branched cells found in seed coats and fruit.

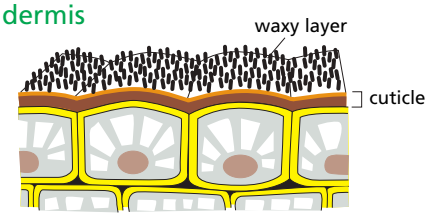




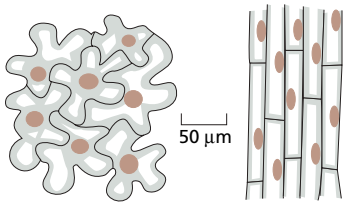
DERMAL TISSUE

The epidermis is the primary outer protective covering of the plant body. Cells of the epidermis are also modified to form stomata and hairs of various kinds.

Epidermis



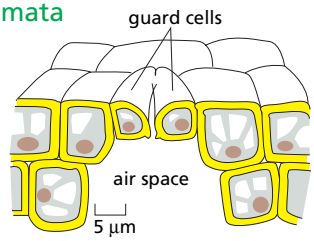
The epidermis (usually one layer of cells deep) covers the entire stem, leaf, and root of the young plant. The cells are living, have thick primary cell walls, and are covered on their outer surface by a special cuticle with an outer waxy layer. The cells are tightly interlocked in different patterns.



upper epidermis of a leaf

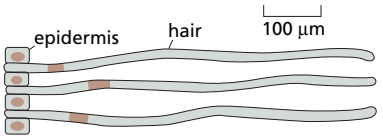
epidermis of a stem

Stomata

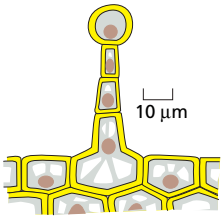


Stomata are openings in the epidermis, mainly on the lower surface of the leaf, that regulate gas exchange in the plant. They are formed by two specialized epidermal cells called *guard cells,* which regulate the diameter of the pore. Stomata are distributed in a distinct species-specific pattern within each epidermis.

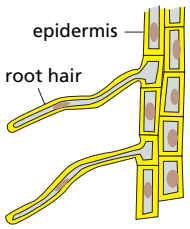
Hairs (or trichomes) are appendages derived from epidermal cells. They exist in a variety of forms and are commonly found in all plant parts. Hairs function in protection, absorption, and secretion; for example,



young, single-celled hairs in the epidermis of the cotton seed. When these grow, the walls will be secondarily thickened with cellulose to form cotton fibers.



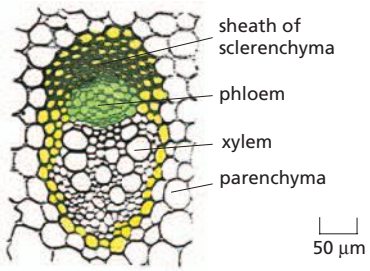
a multicellular secretory hair from a geranium leaf



Single-celled root hairs have an important function in water and ion uptake.

Vascular bundles

Roots usually have a single vascular bundle, but stems have several bundles. These are arranged with strict radial symmetry in dicots, but they are more irregularly dispersed in monocots.

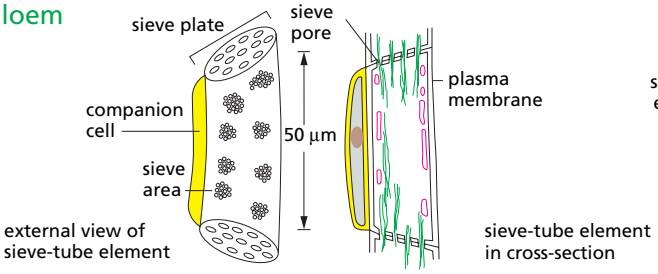


a typical vascular bundle from the young stem of a buttercup

VASCULAR TISSUE

The phloem and the xylem together form a continuous vascular system throughout the plant. In young plants they are usually associated with a variety of other cell types in *vascular bundles.* Both phloem and xylem are complex tissues. Their conducting elements are associated with parenchyma cells that maintain the elements and exchange materials with them. In addition, groups of collenchyma and sclerenchyma cells provide mechanical support.

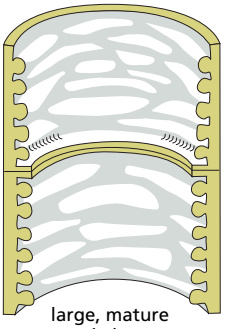
Phloem



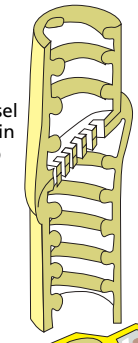
Phloem is involved in the transport of organic solutes in the plant. The main conducting cells (elements) are aligned to form tubes called *sieve tubes*. The sieve-tube elements at maturity are living cells, interconnected by perforations in their end walls formed from enlarged and modified plasmodesmata (sieve plates). These cells retain their plasma membrane, but they have lost their nuclei and much of their cytoplasm; they therefore rely on associated *companion cells* for their maintenance. These companion cells have the additional function of actively transporting soluble food molecules into and out of sieve-tube elements through porous sieve areas in the wall.

Xylem

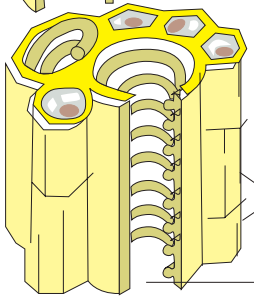
Xylem carries water and dissolved ions in the plant. The main conducting cells are the vessel elements shown here, which are dead cells at maturity that lack a plasma membrane. The cell wall has been secondarily thickened and heavily lignified. As shown below, its end wall is largely removed, enabling very long, continuous tubes to be formed.



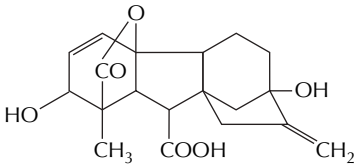
large, mature vessel element



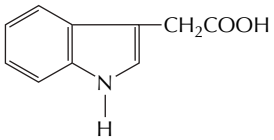
small vessel element in root tip



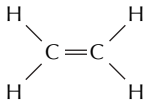
The vessel elements are closely associated with xylem parenchyma cells, which actively transport selected solutes into and out of the elements across the parenchyma cell plasma membrane.



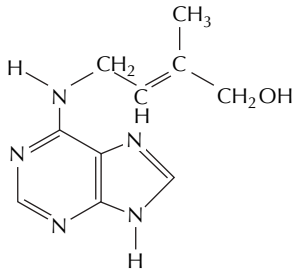
gibberellic acid (GA3) [a gibberellin]



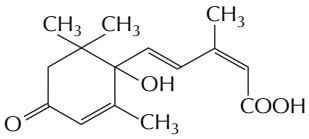
indole-3-acetic acid (IAA) [an auxin]



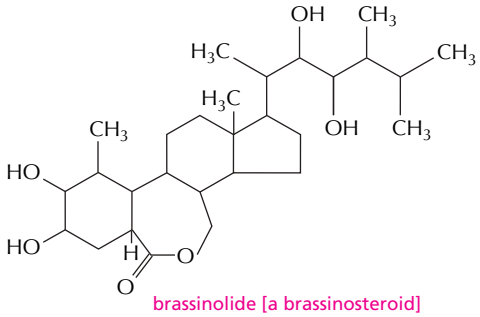
ethylene



zeatin [a cytokinin]



abscisic acid (ABA)



brassinolide [a brassinosteroid]

Figure 22–117 Plant growth regulators. The formula of one naturally occurring representative molecule from each of six groups of plant growth regulatory molecules is shown.

all synthesized by most plant cells and can either act locally or be transported to influence target cells at a distance. Auxin, for example, is transported from cell to cell at a rate of about 1 cm per hour from the tip of a shoot toward its base. Each growth regulator has multiple effects, and these are modulated by the other growth regulators, as well as by environmental cues and nutritional status. Thus, auxin alone can promote root formation, but in conjunction with gibberellin it can promote stem elongation, with cytokinin, it can suppress lateral shoot outgrowth, and with ethylene it can stimulate lateral root growth. Remarkably, as we shall see below, auxin also controls the detailed patterns of cell specialization on a microscopic scale in the apical meristem. The receptors that recognize some of these growth regulators are discussed in Chapter 15.

The Shaping of Each New Structure Depends on Oriented Cell Division and Expansion

Plant cells, imprisoned within their cell walls, cannot crawl about and cannot be shuffled as the plant grows; but they can divide, and they can swell, stretch, and bend. The morphogenesis of a developing plant therefore depends on orderly cell divisions followed by strictly oriented cell expansions. Most cells produced in the root-tip meristem, for example, go through three distinct phases of development—division, growth (elongation), and differentiation. These three steps, which overlap in both space and time, give rise to the characteristic architecture of a root tip. Although the process of cell differentiation often begins while a cell is still enlarging, it is comparatively easy to distinguish in a root tip a zone of cell division, a zone of oriented cell elongation (which accounts for the growth in length of the root), and a zone of cell differentiation (**Figure 22–118**).

In the phase of controlled expansion that generally follows cell division, the daughter cells may often increase in volume by a factor of 50 or more. This expansion is driven by an osmotically based turgor pressure that presses outward on the plant cell wall, and its direction is determined by the orientation of the cellulose fibrils in the cell wall, which constrain expansion along one axis (see Figure 19–73). The orientation of the cellulose in turn is apparently controlled by the orientation of arrays of microtubules just inside the plasma membrane, which are thought to guide cellulose deposition (discussed in Chapter 19). These orientations can be rapidly changed by plant growth regulators, such as ethylene and gibberellic acid (**Figure 22–119**), but the molecular mechanisms underlying these dramatic cytoskeletal rearrangements are still unknown.

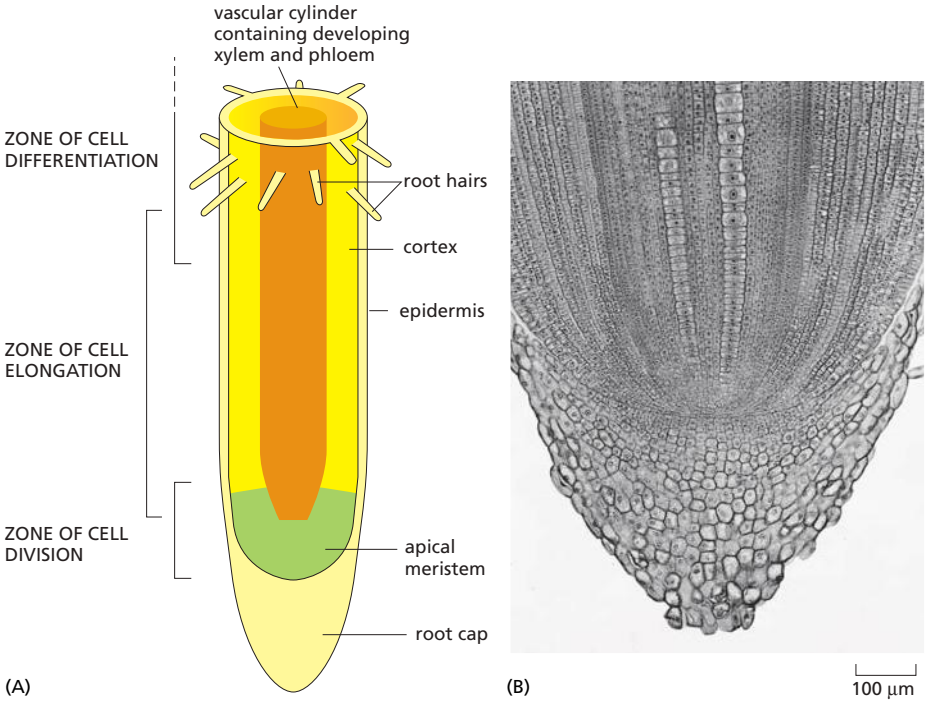


Figure 22–118 A growing root tip. (A) The organization of the final 2 mm of a growing root tip. The approximate zones in which cells can be found dividing, elongating, and differentiating are indicated. (B) The apical meristem and root cap of a corn root tip, showing the orderly files of cells produced. (B, from R.F. Evert, Biology of Plants, 4th ed. New York: Worth, 1986.)

Each Plant Module Grows From a Microscopic Set of Primordia in a Meristem

The apical meristems are self-perpetuating: in a perennial plant, they carry on with their functions indefinitely, as long as the plant survives, and they are responsible for its continuous growth and development. But apical meristems also give rise to a second type of outgrowth, whose development is strictly limited and culminates in the formation of a structure such as a leaf or a flower, with a determinate size and shape and a short lifespan. Thus, as a vegetative (nonflowering) shoot elongates, its apical meristem lays down behind itself an orderly sequence of *nodes*, where leaves have grown out, and *internodes* (segments of stem). In this way the continuous activity of the meristem produces an ever increasing number of similar modules, each consisting of a stem, a leaf, and a bud (see Figure 22–111). The modules are connected to one another by supportive and transport tissue, and successive modules are precisely located relative to each other, giving rise to a repetitively patterned structure. This iterative mode of development is characteristic of plants and is seen in many other structures besides the stem–leaf system (**Figure 22–120**).

Although the final module may be large, its organization, like that of an animal embryo, is mapped out at first on a microscopic scale. At the apex of the shoot, within a space of a millimeter or less, one finds a small, low central dome surrounded by a set of distinctive swellings in various stages of enlargement (**Figure 22–121**). The central dome is the apical meristem itself; each of the surrounding swellings is the primordium of a leaf. This small region, therefore, contains the already distinct rudiments of several entire modules. Through a well-

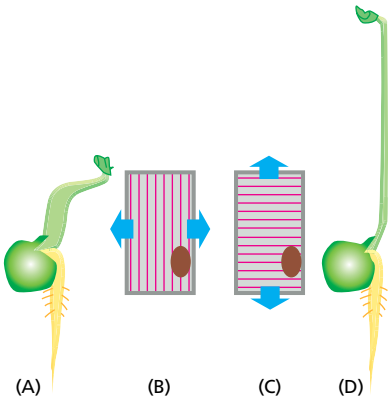


Figure 22–119 The different effects of the plant growth regulators ethylene and gibberellic acid. These regulators exert rapid and opposing effects on the orientation of the cortical microtubule array in cells of young pea shoots. A typical cell in an ethylene-treated plant (B) shows a net longitudinal orientation of microtubules, while a typical cell in a gibberellic-acid-treated plant (C) shows a net transverse orientation. New cellulose microfibrils are deposited parallel to the microtubules. Since this influences the direction of cell expansion, gibberellic acid and ethylene encourage growth in opposing directions: ethylene-treated seedlings will develop short, fat shoots (A), while gibberellic-acid-treated seedlings will develop long, thin shoots (D).

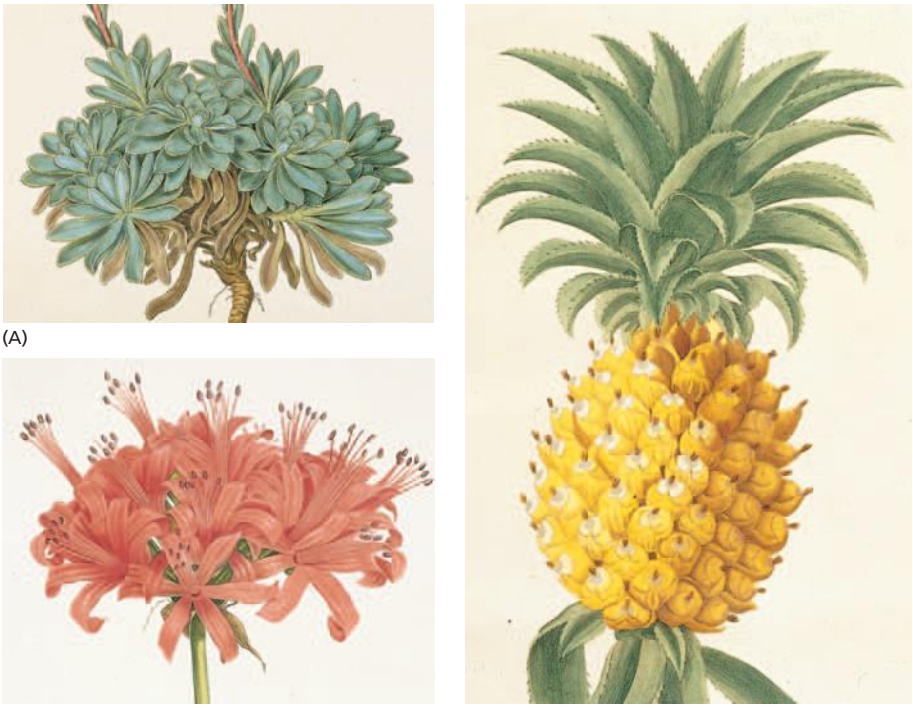


Figure 22–120 Repetitive patterning in plants. Accurate placing of successive modules from a single apical meristem produces these elaborate but regular patterns in leaves (A), flowers (B), and fruits (C). (A, from John Sibthorp, Flora Graeca. London: R. Taylor, 1806–1840; B, from Pierre Joseph Redouté, Les Liliacées. Paris: chez l'Auteur, 1807; C, from Christopher Jacob Trew, Uitgezochte planten. Amsterdam: Jan Christiaan Sepp, 1771—all courtesy of the John Innes Foundation.)

defined program of cell proliferation and cell enlargement, each leaf primordium and its adjacent cells will grow to form a leaf, a node, and an internode. Meanwhile, the apical meristem itself will give rise to new leaf primordia, so as to generate more and more modules in a potentially unending succession. The serial organization of the modules of the plant is thus controlled by events at the shoot apex.

Polarized Auxin Transport Controls the Pattern of Primordia in the Meristem

What are the signals that operate in the tiny apical region to determine the arrangement of primordia, and how are these signals generated in the appropriate pattern? A clue comes from mutation of a gene called *Pin1*, whose loss prevents formation of leaf primordia but allows the main stem to continue growing, producing a long thin bare structure shaped like a pin, with the apical meristem

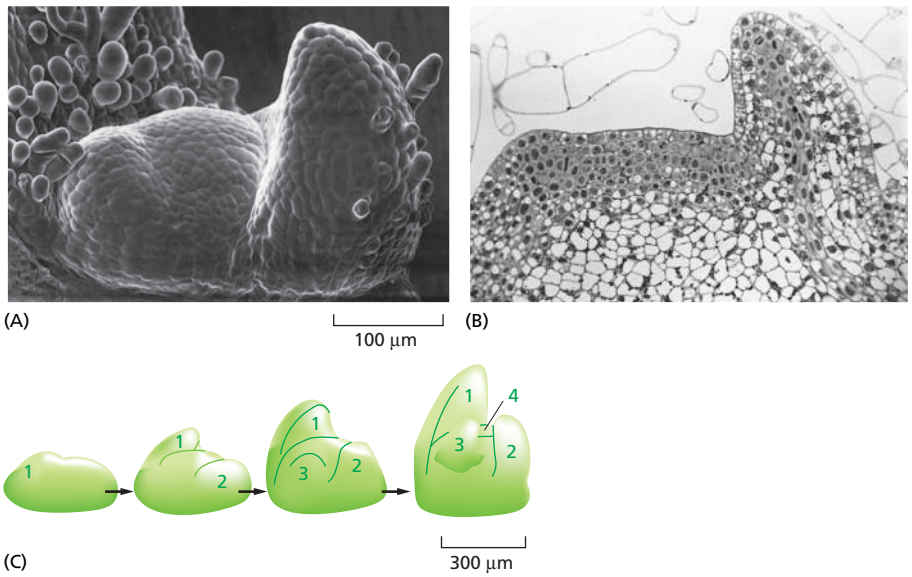
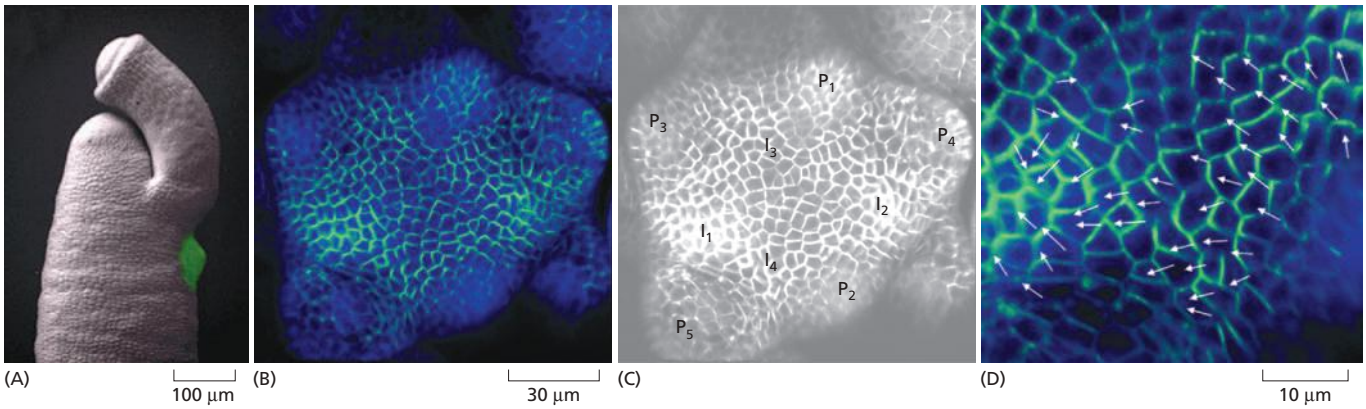


Figure 22–121 A shoot apex from a young tobacco plant. (A) A scanning electron micrograph shows the shoot apex with two sequentially emerging leaf primordia, seen here as lateral swellings on either side of the domed apical meristem. (B) A thin section of a similar apex shows that the youngest leaf primordium arises from a small group of cells (about 100) in the outer four or five layers of cells. (C) A very schematic drawing showing that the sequential appearance of leaf primordia takes place over a small distance and very early in shoot development. Growth of the apex will eventually form internodes that will separate the leaves in order along the stem (see Figure 22–111). (A and B, from R.S. Poethig and I.M. Sussex, *Planta* 165:158–169, 1985. With permission from Springer-Verlag.)



at its head. The Pin1 protein is an auxin transporter, driving efflux across the plasma membrane into the extracellular space. This suggests that the leaf primordia are missing in the mutant because the auxin distribution is wrong. Indeed, a microdroplet of auxin applied to one side of a *Pin1* or similar type of mutant apical meristem, on the side of the head of the "pin", will induce a leaf or flower primordium to form at the site of the auxin application (**Figure 22–122**A).

One can observe the distribution of Pin1 transporter protein in the living tissue by creating a transgenic (but otherwise normal) plant that expresses a form of Pin1 tagged with Green Fluorescent Protein (Figure 22–122B–D). In the outermost layer of meristem cells, the amount of Pin1 varies from region to region in a pattern that correlates with the pattern of developing primordia because the *Pin1* gene is upregulated by auxin. Moreover, the Pin1 protein is asymmetrically distributed in the membranes of the individual cells, so that they pump out more auxin on one side than the other, creating local maxima that specify where primordia will begin to form. The pumps appear to be concentrated on the side facing the neighbors whose own auxin concentration is the highest, suggesting that there is a positive feedback in the accumulation of auxin. Computer models show that positive feedback of this type can amplify asymmetry and generate a pattern of peaks and troughs of auxin concentration of the sort observed. Localized transport of auxin in a perpendicular direction, between the outer sheet of meristematic cells and the developing strands of vascular tissue below, contributes to the asymmetry. As the cells proliferate and the tissue grows, the distributions of Pin1 protein and of auxin adjust, producing new peaks and new lateral primordia in regular succession.

Variations on this basic repetitive theme can give rise to more complex architectures, including structures such as tendrils, leaves, branches, and flowers. Thus, by switching on different sets of genes at the shoot apex, the plant can produce different types of primordia, in different spatial patterns.

Cell Signaling Maintains the Meristem

Central to all these phenomena is the question of how the apical meristem maintains itself. The meristem cells must continue to proliferate for weeks, years, or even centuries as a plant grows, replacing themselves while continually generating progeny cells that differentiate. Through all this, the size of the cluster of cells that constitute the meristem remains practically constant (about 100 cells in *Arabidopsis*, for example). New meristems may arise as the plant branches, but they too preserve the same size.

Genetic screens have identified genes required for meristem maintenance. For example, mutations that disrupt the *Wuschel* gene, which codes for a homeodomain protein, convert the apical meristem into non-meristematic tissue, so that the seedling fails to sprout. Conversely, mutations in the *Clavata* group of genes, coding for components of a cell–cell signaling pathway (see Figure 15–83), make the meristem abnormally big. These genes are expressed in different layers of cells in the meristem region (**Figure 22–123**A). The two most

Figure 22–122 Control of patterning in a meristem by auxin and Pin1. (A) A microdroplet containing auxin *(green spot)* has been applied on one side of a mutant meristem, phenotypically similar to a *Pin1* mutant because it lacks a protein required for control of auxin transport. The auxin has induced formation of a lateral flower primordium. (B) Distribution of the Pin1 auxin transporter at a meristem. (B) An apical meristem of *Arabidopsis* is viewed from above by fluorescence microscopy, revealing the distribution of GFP-tagged Pin1 protein in the surface layer of cells. (C) The same image labeled to show the locations of the established primordia (P_1 being the most recently formed, P_4 being the most mature) and the predicted incipient primordia (I_1 being the next to form, I_4 being the one that lies farthest in the future). (D) Magnified part of (B), showing the asymmetric distribution of Pin1 in the membranes of the individual cells, driving auxin toward the site of an incipient primordium. *Arrows* indicate the direction of transport. As primordia become established, the amount of Pin1 in their surface layer declines, in part because further changes in the distribution of the transport proteins cause auxin to be pumped downward into the developing vascular tissue below. Complex patterns of auxin transport control the detailed structure of many other developing plant tissues also. (A, from D. Reinhardt et al., *Nature* 426:255–260, 2003. With permission from Macmillan Publishers Ltd; B–D, from M.G. Heisler et al., *Curr. Biol.* 15:1899–1911, 2005. With permission from Elsevier.)

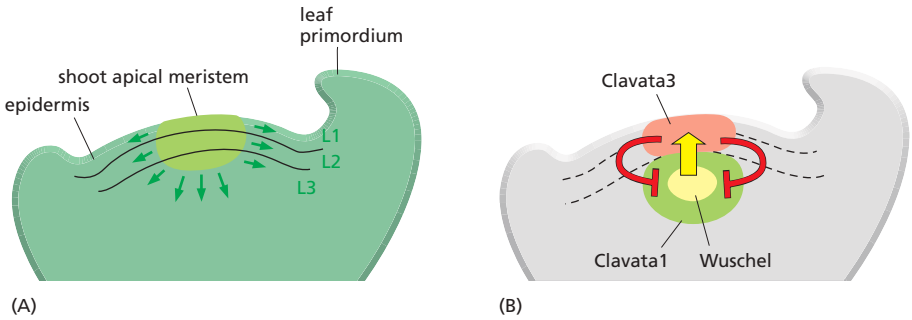


Figure 22–123 The feedback loops that are thought to maintain the shoot apical meristem. (A) The arrangement of cell layers constituting a shoot apical meristem. (B) The pattern of cell–cell communication that maintains the meristem. Artificial overexpression of *Wuschel* in the L3 region causes an increase in the number of cells in the L1 and L2 layers that behave as meristem cells and express *Clavata3*; artificial overexpression of *Clavata3* in the L1 and L2 layers causes a reduction of *Wuschel* expression in the L3 region below and a decrease in the number of meristem cells. *Clavata3* codes for a small signal protein, while *Clavata1* codes for its receptor, a transmembrane protein kinase. *Wuschel,* which is expressed in the central part of the region that expresses the receptor Clavata1, codes for a gene regulatory protein of the homeodomain class. The size of the meristem is thought to be controlled by a self-regulating balance between a short-range stimulatory signal produced by cells expressing Wuschel *(yellow arrow),* and a longer-range inhibitory signal delivered by Clavata3 *(red bars).*

superficial cell layers, called the L1 and L2 layers, together with the uppermost part of the L3 layer, contain the cells of the meristem proper, that is, the stem cells, capable of dividing indefinitely to give rise to future parts of the plant. The meristematic cells of the L1 and L2 layers express Clavata3, a small secreted signal protein. Just beneath, in the L3 layer, lies a cluster of cells expressing Clavata1 (the receptor for Clavata3). In the center of this Clavata1 patch are cells that express the Wuschel gene regulatory protein.

The pattern of cell divisions implies that the cells expressing Wuschel are not themselves part of the meristem proper; new Wuschel-expressing cells are apparently continually recruited from the meristematic (stem-cell) part of the L3 population, just above the Wuschel domain. Nevertheless, the Wuschel-expressing cells are at the heart of the mechanism that maintains the meristem. A signal that they produce maintains meristematic behavior in the cells above, stimulates expression of the *Clavata* genes, and, presumably, causes new cells recruited into the Wuschel domain to switch on Wuschel. Negative feedback from the upper meristematic cells, delivered by the Clavata signaling pathway, acts back on the regions below to limit the size of the Wuschel domain, thereby preventing the meristem from becoming too big (Figure 22–123B).

This account of the plant meristem, though uncertain in some details and certainly oversimplified, provides one of the clearest examples of an important general developmental strategy: it shows how a feedback loop involving a short-range activating signal (such as that produced by the Wuschel-expressing cells) and a long-range inhibitory signal (such as Clavata3) can stably maintain a signaling center of a well-defined size even when there is continual proliferation and turnover of the cells that form that center. As we pointed out at the beginning of this chapter, analogous systems of signals are thought to operate in animal development to maintain localized signaling centers—such as the Organizer of the amphibian gastrula, or the zone of polarizing activity in a limb bud. And just as this strategy serves in the mature plant to maintain its meristems, it may also serve in adult animal tissues such as the gut lining (discussed in Chapter 23) to maintain the all-important clusters of adult animal stem cells.

Regulatory Mutations Can Transform Plant Topology by Altering Cell Behavior in the Meristem

If a plant stem is to branch, new shoot apical meristems must be created, and this too depends on events in the neighborhood of the shoot apex. At each developing node, in the acute angle (the axil) between the leaf primordium and the stem, a bud is formed (**Figure 22–124**). This contains a nest of cells, derived from the apical meristem, that keep a meristematic character. They have the capacity to become the apical meristem of a new branch or the primordium of a structure such as a flower; but they also have the alternative option of remaining quiescent as *axillary buds*. The plant's pattern of branching is regulated through this choice of fate, and mutations that affect it can transform the structure of the plant. Maize provides a beautiful example.

Maize represents one of mankind's most remarkable feats of genetic engineering. Native Americans created it by selective breeding, over a period of several centuries or perhaps millennia between 5000 and 10,000 years ago. They started from a wild grass known as teosinte, with highly branched leafy stems

and tiny ears bearing hard, inedible kernels. Detailed genetic analysis has identified a handful of genetic loci—about five—as the sites of the mutations that account for most of the difference between this unpromising ancestor and modern corn. One of these loci, with a particularly dramatic effect, corresponds to a gene called *Teosinte branched-1 (Tb1)*. In maize with loss-of-function mutations in *Tb1*, the usual simple unbranched stem, with a few large leaves at intervals along it, is transformed into a dense, branching, leafy mass reminiscent of teosinte (**Figure 22–125**A). The pattern of branching in the mutant implies that axillary buds, originating in normal positions, have escaped from an inhibition that prevents them, in normal maize, from growing into branches.

In normal maize, the single stem is crowned with a tassel—a male flower—while a few of the axillary buds along the stem develop into female flowers and, upon fertilization, form the ears of corn that we eat. In the mutant maize with a defective *Tb1* gene, these fruitful axillary buds are transformed into branches bearing tassels. The wild teosinte plant is like the *Tb1*–defective maize in its leafy, highly branched appearance, but unlike this mutant it makes ears on many of its side branches, as though *Tb1* were active. DNA analysis reveals the explanation. Both teosinte and normal maize possess a functional *Tb1* gene, with an almost identical coding sequence, but in maize the regulatory region has undergone a mutation that boosts the level of gene expression. Thus, in normal maize the gene is expressed at a high level in every axillary bud, inhibiting branch formation, while in teosinte the expression in many axillary buds is low, so that branches are permitted to form (Figure 22–125B).

This example shows how simple mutations, by switching the behavior of meristem cells, can transform plant structure—a principle of enormous

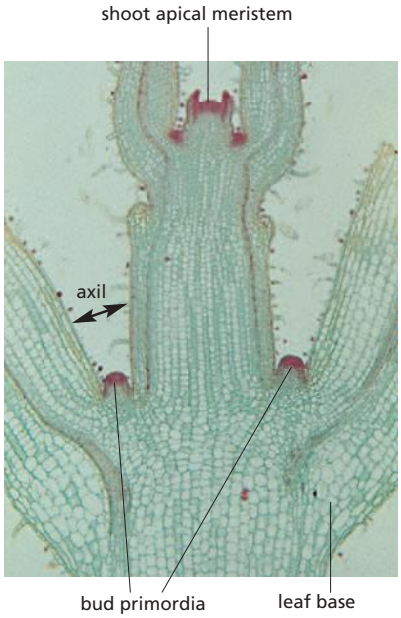


Figure 22–124 Axillary buds in the neighborhood of a shoot apex. The photograph shows a longitudinal section of *Coleus blumei,* a common houseplant. (From P.H. Raven, R.F. Evert and S.E. Eichhorn, Biology of Plants, 6th ed. New York: Freeman/Worth, 1999, used with permission.)



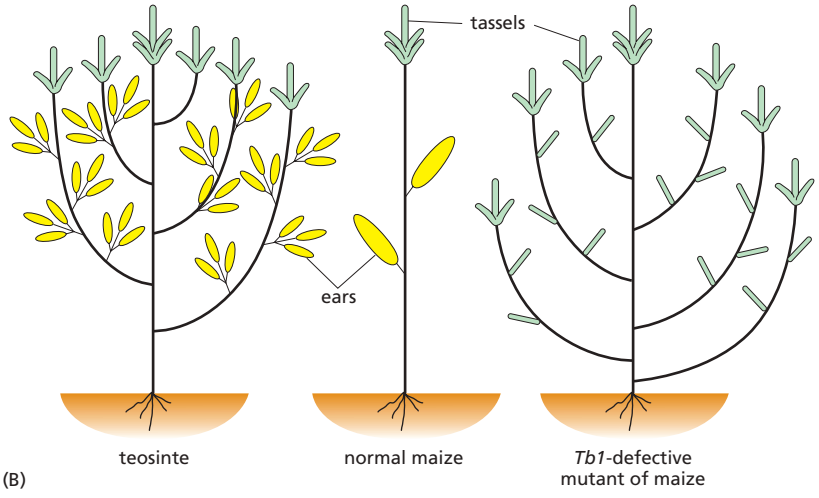


Figure 22–125 Transformation of plant architecture by mutation: a comparison of teosinte, normal maize, and *Tb1*-defective maize. (A) Photographs of the three types of plants. (B) The architecture of teosinte, normal maize and the *Tb1*-defective maize compared schematically. The *Tb1* gene product is needed for development of ears. It is absent in the *Tb1* mutant; it is present in both teosinte and normal maize, but these two plants differ because the gene is differently regulated. (A. left image, from J. Doebley and R.L. Wang, *Cold Spring Harbor Symp. Quant Biol.* 62:361–367, 1997. With permission from Cold Spring Harbor Laboratory Press; A, middle and right images. from J. Doebley, A. Stec and L. Hubbard, *Nature* 386:485-488, 1997. With permission from Macmillan Publishers Ltd.)

importance in the breeding of plants for food. More generally, the case of *Tb1* illustrates how new body plans, whether of plant or animal, can evolve through changes in regulatory DNA without change in the characters of the proteins made.

The Switch to Flowering Depends on Past and Present Environmental Cues

Meristems face other developmental choices besides that between quiescence and growth, as we have already seen in our discussion of maize, and these also are frequently regulated by the environment. The most important is the decision to form a flower (**Figure 22–126**).

The switch from meristematic growth to flower formation is triggered by a combination of cues. The plant does not merely take account of the current temperature, light intensity, and nutritional conditions; it bases its decision to flower on past conditions as well. One important cue, for many plants, is day length. To sense this, the plant uses its circadian clock—an endogenous 24-hour rhythm of gene expression—to generate a signal for flowering only when there is light for the appropriate part of the day. The clock itself is influenced by light, and the plant in effect uses the clock to compare past to present lighting conditions. Important parts of the genetic circuitry underlying these phenomena have been identified, from the phytochromes and cryptochromes that act as light receptors (discussed in Chapter 15) to the *Constans* gene, whose expression in the leaves of the plant represents a signal for flowering. The signal is thought to be relayed from the leaves to the meristem via the vasculature by the product of another gene, *Flowering locus T* (*Ft*), that is regulated by Constans.

But this signal itself will reach the meristem and trigger flowering only if the plant is in a receptive condition, typically depending on its history over a much longer period. Many plants will flower only if they have previously spent a long time in the cold: they must pass through winter before they will behave as though it is spring—a process called *vernalization*. The prolonged cold brings about changes in chromatin structure, dependent on another large collection of genes, including homologs of members of the *Polycomb* group that we mentioned earlier for their role in perpetuating patterns of gene expression in *Drosophila*. These epigenetic changes (discussed in Chapters 4 and 7) result in the gradual silencing of the *Flowering locus C* (*Flc*) gene. The effect is long-lasting, persisting through many rounds of cell division even as the weather grows warmer. *Flc* codes for an inhibitor of flowering, antagonizing the expression and action of *Ft*. Thus vernalization, by blocking production of the inhibitor, enables the meristem to receive the Ft signal and respond to it by switching on the expression of a set of *floral meristem-identity* genes in the apical meristem.

Mutations affecting the regulation of *Flc* expression alter the time of flowering and thus the ability of a plant to flourish in a given climate. The whole

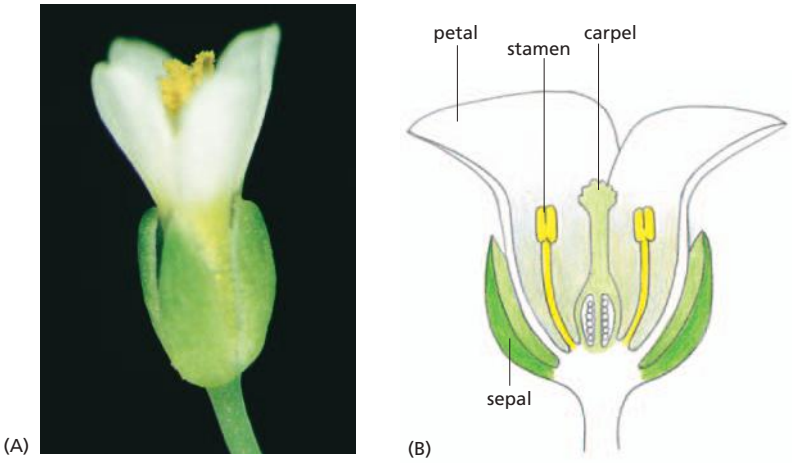


Figure 22–126 The structure of an *Arabidopsis* flower. (A) Photograph. (B) Schematic cross-sectional view. The basic plan, as shown in (B), is common to most flowering dicotyledonous plants. (A, courtesy of Leslie Sieburth.)

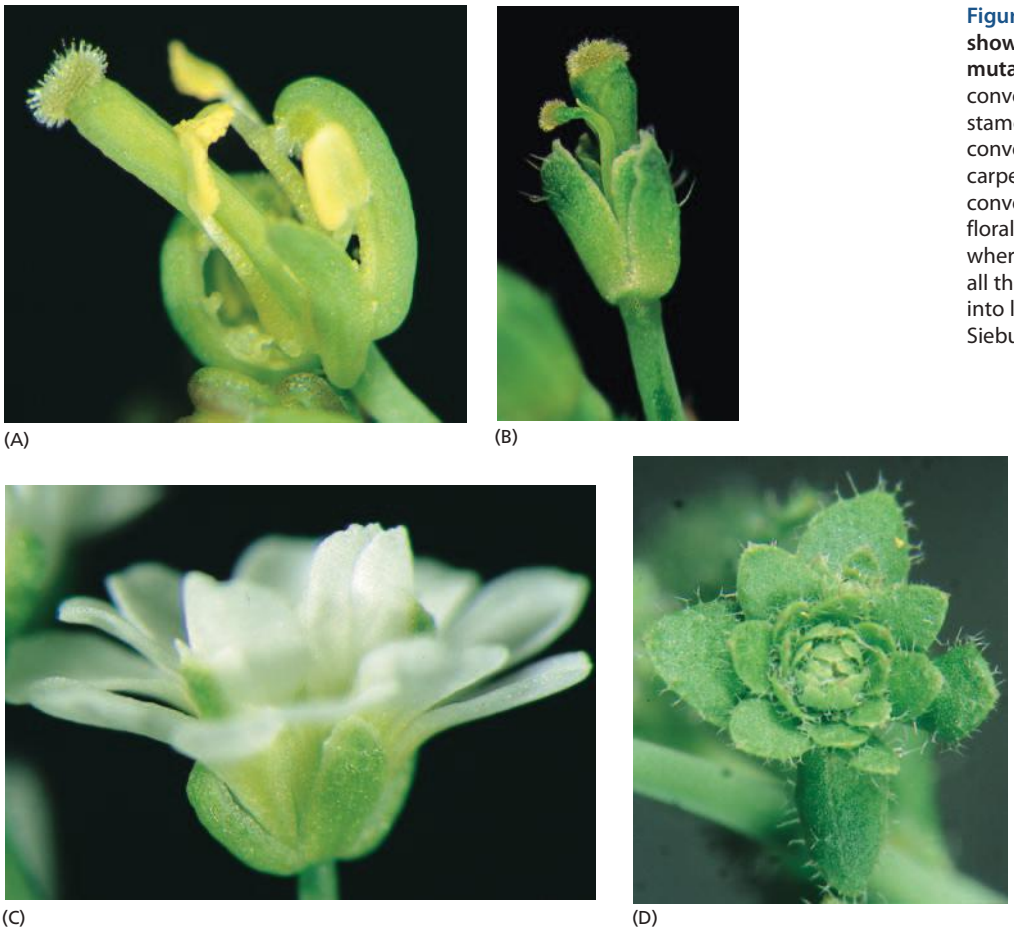


Figure 22–127 *Arabidopsis* flowers showing a selection of homeotic mutations. (A) In *Apetala2,* sepals are converted into carpels and petals into stamens; (B) In *Apetala3,* petals are converted into sepals and stamens into carpels; (C) In *Agamous,* stamens are converted into petals and carpels into floral meristem. (D) In a triple mutant where these three functions are defective, all the organs of the flower are converted into leaves. (A–C, courtesy of Leslie Sieburth; D, courtesy of Mark Running.)

control system governing the switch to flowering thus is of vital importance for agriculture, especially in an era of rapid climate change.

Homeotic Selector Genes Specify the Parts of a Flower

By switching on the floral meristem-identity genes, the apical meristem abandons its chances of continuing vegetative growth and gambles its future on the production of gametes. Its cells embark on a strictly finite program of growth and differentiation: by a modification of the ordinary mechanisms for generating leaves, a series of whorls of specialized appendages are formed in a precise order—typically sepals first, then petals, then stamens carrying anthers containing pollen, and lastly carpels containing eggs (see Panel 22–1). By the end of this process the meristem has disappeared, but among its progeny it has created germ cells.

The series of modified leaves forming a flower can be compared to the series of body segments forming a fly. In plants, as in flies, one can find homeotic mutations that convert one part of the pattern to the character of another. The mutant phenotypes can be grouped into at least four classes, in which different but overlapping sets of organs are altered (**Figure 22–127**). The first or 'A' class, exemplified by the *Apetala2* mutant of *Arabidopsis*, has its two outermost whorls transformed: the sepals are converted into carpels and the petals into stamens. The second or 'B' class, exemplified by *Apetala3,* has its two middle whorls transformed: the petals are converted into sepals and the stamens into carpels. The third or 'C' class, exemplified by *Agamous,* has its two innermost whorls transformed, with a more drastic consequence: the stamens are converted into petals, the carpels are missing, and in their place the central cells of the flower behave as a floral meristem, which begins the developmental performance all over again, generating another abnormal set of sepals and petals nested inside

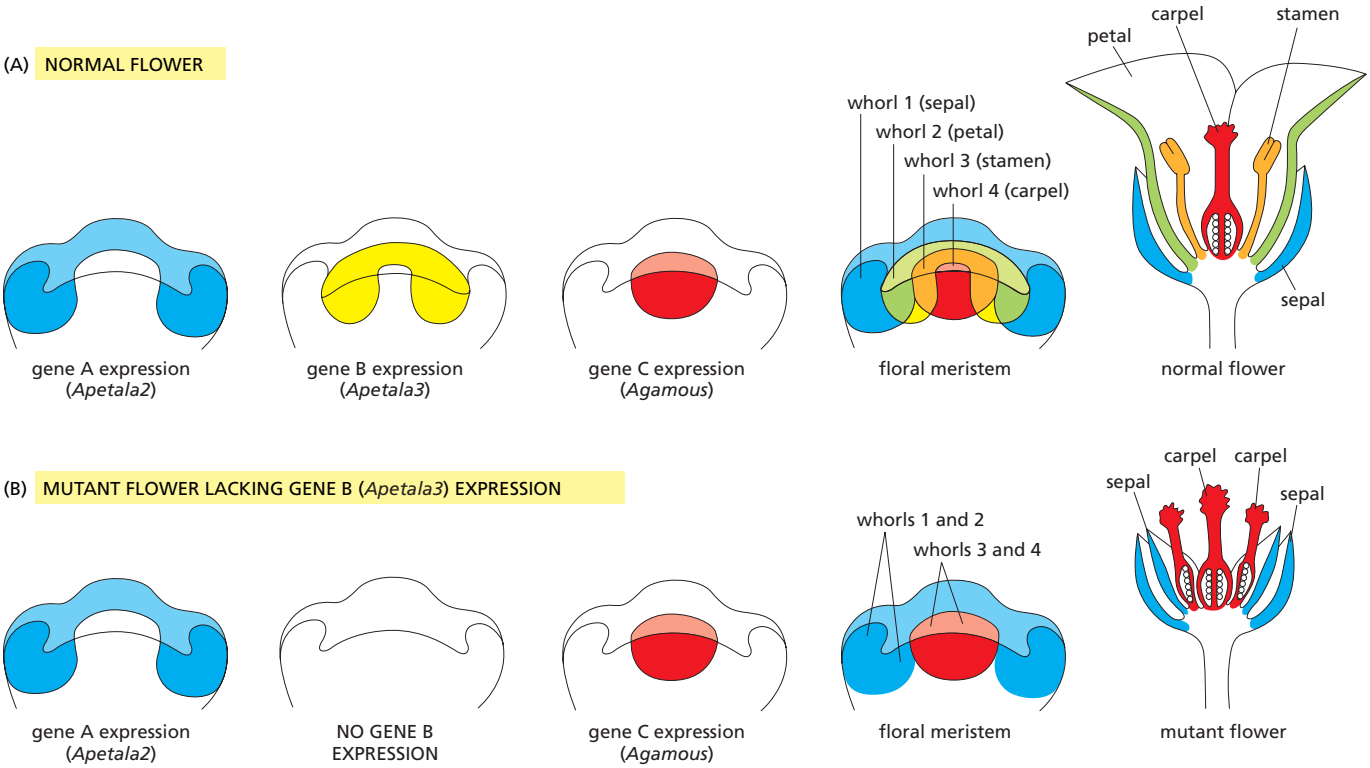


Figure 22–128 Homeotic selector gene expression in an *Arabidopsis* flower. (A) Diagram of the normal expression patterns of the three genes whose mutant phenotypes are illustrated in Figure 22–127A–C. All three genes code for gene regulatory proteins. The colored shading on the flower indicates which organ develops from each whorl of the meristem, but does not imply that the homeotic selector genes are still expressed at this stage. (B) The patterns in a mutant where the *Apetala3* gene is defective. Because the character of the organs in each whorl is defined by the set of homeotic selector genes that they express, the stamens and petals are converted into sepals and carpels. The consequence of a deficiency of a gene of class A, such as *Apetala2*, is slightly more complex: the absence of this class A gene product allows the class C gene to be expressed in the outer two whorls as well as the inner two, causing these outer whorls to develop as carpels and stamens, respectively. Deficiency of a class C gene prevents the central region from undergoing terminal differentiation as a carpel and causes it instead to continue growth as a meristem, generating more and more sepals and petals.

the first and, potentially, another nested inside that, and so on, indefinitely. A fourth class, the *Sepallata* mutants, has its three inner whorls all transformed into sepals.

These phenotypes identify four classes of homeotic selector genes, which, like the homeotic selector genes of *Drosophila*, all code for gene regulatory proteins. These are expressed in different domains and define the differences of cell state that give the different parts of a normal flower their different characters, as shown in **Figure 22–128**. The gene products collaborate to form protein complexes that drive expression of the appropriate downstream genes. In a triple mutant where the A, B, and C genetic functions are all absent, one obtains in place of a flower an indefinite succession of tightly nested leaves (see Figure 22–127D). Conversely, in a transgenic plant where genes of the A, B, and *Sepallata* classes are all expressed together outside their normal domains, leaves are transformed into petals. Leaves therefore represent a "ground state" in which none of these homeotic selector genes are expressed, while the other types of organ result from expressing the genes in different combinations.

Similar studies have been carried out in other plant species, and a similar set of phenotypes and genes have been identified: plants, no less than animals, have conserved their homeotic selector gene systems. Gene duplication has played a large part in the evolution of these genes: several of them, required in different organs of the flower, have clearly homologous sequences. These are not of the homeobox class but are members of another family of gene regulatory proteins (the so-called MADS family), also found in yeast and in vertebrates.

Clearly, plants and animals have independently found very similar solutions to many of the fundamental problems of multicellular development.

Summary

The development of a flowering plant, like that of an animal, begins with division of a fertilized egg to form an embryo with a polarized organization: the apical part of the embryo will form the shoot, the basal part, the root, and the middle part, the stem. At first, cell division occurs throughout the body of the embryo. As the embryo grows, however, addition of new cells becomes restricted to small regions known as meristems. Apical meristems, at shoot tips and root tips, will persist throughout the life of the plant, enabling it to grow by sequentially adding new body parts at its periphery. Typically, the shoot generates a repetitive series of modules, each consisting of a segment of stem, a leaf, and an axillary bud. Polarized transport of auxin controls the positioning of the primordia of these structures as they arise in the neighborhood of the meristem. An axillary bud is a potential new meristem, capable of giving rise to a side branch; the environment—and long-range hormonal signals within the plant can control the development of the plant by regulating bud activation. Mutations that alter the rules for activating axillary buds can have a drastic effect on the shape and structure of the plant; a single such mutation—one of about five key genetic alterations—accounts for a large part of the dramatic difference between modern maize and its wild ancestor, teosinte.

The small weed Arabidopsis thaliana *is widely used as a model organism for genetic studies and is the first plant to have had its genome completely sequenced. As in animals, genes governing plant development can be identified through genetic screens and their functions tested by genetic manipulations. Such studies have begun to reveal the molecular mechanisms by which the internal organization of each plant module is sketched out on a microscopic scale through cell–cell interactions in the neighborhood of the apical meristem. The meristem itself appears to be maintained by a local feedback loop, in which cells expressing the gene regulatory protein Wuschel provide a positive stimulus, and a negative feedback dependent on the Clavata cell–cell signaling pathway keeps the meristem from becoming too big.*

Environmental cues—especially light that is appropriately timed—can cause the expression of genes that switch the apical meristem from a leaf-forming to a flower-forming mode. The parts of a flower—its sepals, petals, stamens, and carpels—are formed by a modification of the mechanism for development of leaves, and the differences between these parts are controlled by homeotic selector genes that are closely analogous (although not homologous) to those of animals.

REFERENCES

General

Carroll SB, Grenier JK & Weatherbee SD (2004) From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design, 2nd ed. Malden, MA: Blackwell Science.

Gilbert SF (2006) Developmental Biology, 8th ed. Sunderland, MA: Sinauer.

Nüsslein-Volhard C (2006) Coming to Life: How Genes Drive Development. London: Yale University Press

Wolpert L, Jessell TM, Lawrence P et al (2006) Principles of Development, 3rd ed. Oxford: Oxford University Press.

Universal Mechanisms of Animal Development

Gilbert SF & Raunio AM (eds) (1997) Embryology: Constructing The Organism. Sunderland, MA: Sinauer.

Giraldez AJ, Cinalli RM, Glasner ME et al (2005) MicroRNAs regulate brain morphogenesis in zebrafish. *Science* 308:833–838.

Lander AD (2007) Morpheus unbound: reimagining the morphogen gradient. *Cell* 128:245–256.

Meinhardt H & Gierer A (2000) Pattern formation by local self-activation and lateral inhibition. *BioEssays* 22:753–760.

Spemann H (1938) Embryonic Development and Induction. New Haven: Yale University Press. Reprinted 1988, New York: Garland Publishing.

Tyson JJ, Chen KC & Novak B (2003) Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr Opin Cell Biol* 15:221–231.

Caenorhabditis elegans: Development From the Perspective of the Individual Cell

Girard LR, Fiedler TJ, Harris TW et al (2007) WormBook: the online review of *Caenorhabditis elegans* biology. *Nucleic Acids Res* 35:D472–475.

Harris WA & Hartenstein V (1991) Neuronal determination without cell division in *Xenopus* embryos. *Neuron* 6:499–515.

Kamath RS, Fraser AG, Dong Y et al (2003) Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature* 421:231–237.

Lyczak R, Gomes JE & Bowerman B (2002) Heads or tails: cell polarity and axis formation in the early *Caenorhabditis elegans* embryo. *Dev Cell* 3:157–166.

Metzstein MM, Stanfield GM & Horvitz HR (1998) Genetics of programmed cell death in *C. elegans*: past, present and future. *Trends Genet* 14:410–416.

Pasquinelli AE, Reinhart BJ, Slack F et al (2000) Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* 408:86–89.

Strome S & Lehmann R (2007) Germ versus soma decisions: lessons from flies and worms. *Science* 316:392–393.

Drosophila and the Molecular Genetics of Pattern Formation: Genesis of the Body Plan

Bate M & Martinez Arias A (eds) (1993) The Development of *Drosophila melanogaster*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Gregor T, Bialek W, de Ruyter van Steveninck RR, Tank DW & Wieschaus EF (2005) Diffusion and scaling during early embryonic pattern formation. *Proc Natl Acad Sci U S A* 102:18403–18407.
Huynh JR & St Johnston D (2004) The origin of asymmetry: early polarisation of the *Drosophila* germline cyst and oocyte. *Curr Biol* 14:R438–449.
Lawrence PA (1992) The Making of a Fly: The Genetics of Animal Design. Oxford: Blackwell Scientific.
Nusslein-Volhard C & Wieschaus E (1980) Mutations affecting segment number and polarity in *Drosophila*. *Nature* 287:795–801.
Peel AD, Chipman AD & Akam M (2005) Arthropod segmentation: beyond the *Drosophila* paradigm. *Nature Rev Genet* 6:905–916.
Stathopoulos A & Levine M (2004) Whole-genome analysis of *Drosophila* gastrulation. *Curr Opin Genet Dev* 14:477–484.

Homeotic Selector Genes and the Patterning of the Anteroposterior Axis

Lemons D & McGinnis W (2006) Genomic evolution of Hox gene clusters. *Science* 313:1918–1922.
Lewis EB (1978) A gene complex controlling segmentation in *Drosophila*. *Nature* 276:565–570.
McGinnis W, Garber RL,Wirz J et al (1984) A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans. *Cell* 37:403–408.
Ringrose L & Paro R (2007) Polycomb/Trithorax response elements and epigenetic memory of cell identity. *Development* 134:223–232.
Wellik DM & Capecchi MR (2003) Hox10 and Hox11 genes are required to globally pattern the mammalian skeleton. *Science* 301:363–367.

Organogenesis and the Patterning of Appendages

Bellaiche Y & Schweisguth F (2001) Lineage diversity in the *Drosophila* nervous system. *Curr Opin Genet Dev* 11:418–423.
Betschinger J & Knoblich JA (2004) Dare to be different: asymmetric cell division in *Drosophila, C. elegans* and vertebrates. *Curr Biol* 14:R674–685.
Betschinger J, Mechtler K & Knoblich JA (2006) Asymmetric segregation of the tumor suppressor brat regulates self-renewal in *Drosophila* neural stem cells. *Cell* 124:1241–1253.
Irvine KD & Rauskolb C (2001) Boundaries in development: formation and function. *Annu Rev Cell Dev Biol* 17:189–214.
Harvey K & Tapon N (2007) The Salvador-Warts-Hippo pathway—an emerging tumour-suppressor network. *Nature Rev Cancer* 7:182–191.
O'Farrell P (2004). How metazoans reach their full size : the natural history of bigness, In Cell Growth: The Control of Cell Size, MN Hall et al, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 1–22.
Panganiban G, Irvine SM, Lowe C et al (1997) The origin and evolution of animal appendages. *Proc. Natl. Acad. Sci. USA* 94:5162–5166.
Rogulja D & Irvine KD (2005) Regulation of cell proliferation by a morphogen gradient. *Cell* 123:449–461.
Tickle C (2003) Patterning systems—from one end of the limb to the other. *Dev Cell* 4:449–458.

Cell Movement and the Shaping of the Vertebrate Body

David NB, Sapede D, Saint-Etienne L et al (2002) Molecular basis of cell migration in the fish lateral line: role of the chemokine receptor CXCR4 and of its ligand, SDF1. *Proc Natl Acad Sci U S A* 99:16297–16302.
Heasman J (2006) Maternal determinants of embryonic cell fate. *Semin Cell Dev Biol* 17:93–98.
Hirokawa N, Tanaka Y, Okada Y & Takeda S (2006) Nodal flow and the generation of left-right asymmetry. *Cell* 125:33–45.
Kalcheim C & Le Douarin NM (1999) The Neural Crest, 2nd ed. Cambridge: Cambridge University Press.
Lewis J (2003) Autoinhibition with transcriptional delay: a simple mechanism for the zebrafish somitogenesis oscillator. *Curr Biol* 13:1398–1408.
Pourquie O (2003) The segmentation clock: converting embryonic time into spatial pattern. *Science* 301:328–330.
Raz E (2004) Guidance of primordial germ cell migration. *Curr Opin Cell Biol* 16:169–173.
Schier AF (2007) The maternal-zygotic transition: death and birth of RNAs. *Science* 316:406–407.
Tabin CJ (2006) The key to left-right asymmetry. *Cell* 127:27–32.
Takeichi M, Nakagawa S, Aono S et al (2000) Patterning of cell assemblies regulated by adhesion receptors of the cadherin superfamily. *Philos Trans R Soc Lond B Biol Sci* 355:885–890.

The Mouse

Ghabrial A, Luschnig S, Metzstein MM & Krasnow MA (2003) Branching morphogenesis of the *Drosophila* tracheal system. *Annu Rev Cell Dev Biol* 19:623–647.
Hogan BL & Kolodziej PA (2002) Organogenesis: molecular mechanisms of tubulogenesis. *Nature Rev Genet* 3:513–523.
Larsen WJ (2001) Human Embryology, 3rd ed: Churchill Livingstone.
Lu CC, Brennan J & Robertson EJ (2001) From fertilization to gastrulation: axis formation in the mouse embryo. *Curr Opin Genet Dev* 11:384–392.
Smith AG (2001) Embryo-derived stem cells: of mice and men. *Annu Rev Cell Dev Biol* 17:435–462.
Weaver M, Dunn NR & Hogan BL (2000) Bmp4 and Fgf10 play opposing roles during lung bud morphogenesis. *Development* 127:2695–2704.
Zernicka-Goetz M (2006) The first cell-fate decisions in the mouse embryo: destiny is a matter of both chance and choice. *Curr Opin Genet Dev* 16:406–412.

Neural Development

Bi G & Poo M (2001) Synaptic modification by correlated activity: Hebb's postulate revisited. *Annu Rev Neurosci* 24:139–166.
Flanagan JG (2006) Neural map specification by gradients. *Curr Opin Neurobiol* 16:59–66.
Hubel DH & Wiesel TN (1965) Binocular interaction in striate cortex of kittens reared with artificial squint. *J Neurophysiol* 28:1041–1059.
Kandel ER, Schwartz JH & Jessell TM (2000) Principles of Neural Science, 4th ed. New York: McGraw-Hill.
Nägerl UV, Eberhorn N, Cambridge SB & Bonhoeffer T (2004) Bidirectional activity-dependent morphological plasticity in hippocampal neurons. *Neuron* 44:759–767.
Pearson BJ & Doe CQ (2004) Specification of temporal identity in the developing nervous system. *Annu Rev Cell Dev Biol* 20:619–647.
Sanes DH, Reh TA & Harris WA (2000) Development of the Nervous System. San Diego: Academic Press.
Tessier-Lavigne M (2002) Wiring the brain: the logic and molecular mechanisms of axon guidance and regeneration. *Harvey Lect* 98:103–143.
Waites CL, Craig AM & Garner CC (2005) Mechanisms of vertebrate synaptogenesis. *Annu Rev Neurosci* 28:251–274.
Wang GX & Poo MM (2005) Requirement of TRPC channels in netrin-1-induced chemotropic turning of nerve growth cones. *Nature* 434:898–904.
Zou Y, Stoeckli E, Chen H & Tessier-Lavigne M (2000) Squeezing axons out of the gray matter: a role for slit and semaphorin proteins from midline and ventral spinal cord. *Cell* 102:363–375.

Plant Development

Bäurle I & Dean C (2006) The timing of developmental transitions in plants. *Cell* 125:655–664.
Doebley J, Stec A & Hubbard L (1997) The evolution of apical dominance in maize. *Nature* 386:485–488.
Heisler MG, Ohno C, Das P et al (2005) Patterns of auxin transport and gene expression during primordium development revealed by live imaging of the *Arabidopsis* inflorescence meristem. *Curr Biol* 15:1899–1911.
Howell SH (1998) Molecular Genetics of Plant Development. Cambridge: Cambridge University Press.
Jack T (2001) Relearning our ABCs: new twists on an old model. *Trends Plant Sci.* 6:310–316.
Schoof H, Lenhard M, Haecker A et al (2000) The stem cell population of Arabidopsis shoot meristems is maintained by a regulatory loop between the CLAVATA and WUSCHEL genes. *Cell* 100:635–644.

Specialized Tissues, Stem Cells, and Tissue Renewal

23

Cells evolved originally as free-living individuals, but the cells that matter most to us, as human beings, are specialized members of a multicellular community. They have lost features needed for independent survival and acquired peculiarities that serve the needs of the body as a whole. Although they share the same genome, they are spectacularly diverse: there are more than 200 different named cell types in the human body (see our web site for a list). These collaborate with one another to form many different tissues, arranged into organs performing widely varied functions. To understand them, it is not enough to analyze them in a culture dish: we need also to know how they live, work, and die in their natural habitat, the intact body.

In Chapters 7 and 21, we saw how the various cell types become different in the embryo and how cell memory and signals from their neighbors enable them to remain different thereafter. In Chapter 19, we discussed the building technology of multicellular tissues—the devices that bind cells together and the extracellular materials that give them support. In this chapter, we consider the functions and lifestyles of the specialized cells in the adult body of a vertebrate. We describe how cells work together to perform their tasks, how new specialized cells are born, how they live and die, and how the architecture of tissues is preserved despite the constant replacement of old cells by new. We examine in particular the role played in many tissues by *stem cells*—cells that are specialized to provide an indefinite supply of fresh differentiated cells where these are lost, discarded, or needed in greater numbers.

We discuss these topics through a series of examples—some chosen because they illustrate important general principles, others because they highlight favorite objects of study, still others because they pose intriguing problems that cell biology has yet to solve. Finally, we shall confront the practical question that underlies the current storm of interest in stem cells: How can we use our understanding of the processes of cell differentiation and tissue renewal to improve upon nature, and make good those injuries and failings of the human body that have hitherto seemed to be beyond repair?

In This Chapter

EPIDERMIS AND ITS RENEWAL BY STEM CELLS 1417

SENSORY EPITHELIA 1429

THE AIRWAYS AND THE GUT 1434

BLOOD VESSELS, LYMPHATICS, AND ENDOTHELIAL CELLS 1445

RENEWAL BY MULTIPOTENT STEM CELLS: BLOOD CELL FORMATION 1450

GENESIS, MODULATION, AND REGENERATION OF SKELETAL MUSCLE 1463

FIBROBLASTS AND THEIR TRANSFORMATIONS: THE CONNECTIVE-TISSUE CELL FAMILY 1467

STEM-CELL ENGINEERING 1476

EPIDERMIS AND ITS RENEWAL BY STEM CELLS

We begin with a very familiar tissue: the skin. Like almost all tissues, skin is a complex of several different cell types. To perform its basic function as a barrier, the outer covering of the skin depends on a variety of supporting cells and structures, many of which are required in most other tissues also. It needs mechanical support, largely provided by a framework of extracellular matrix, mainly secreted by *fibroblasts*. It needs a blood supply to bring nutrients and oxygen and to remove waste products and carbon dioxide, and this requires a network of blood vessels, lined with *endothelial cells*. These vessels also provide access routes for cells of the immune system to defend against infection: *macrophages* and *dendritic* cells, to phagocytose invading pathogens and help activate *lymphocytes,* and lymphocytes themselves, to mediate more sophisticated adaptive immune system responses (discussed in Chapter 24). *Nerve fibers* are needed too, to convey sensory information from the tissue to the central nervous system, and to

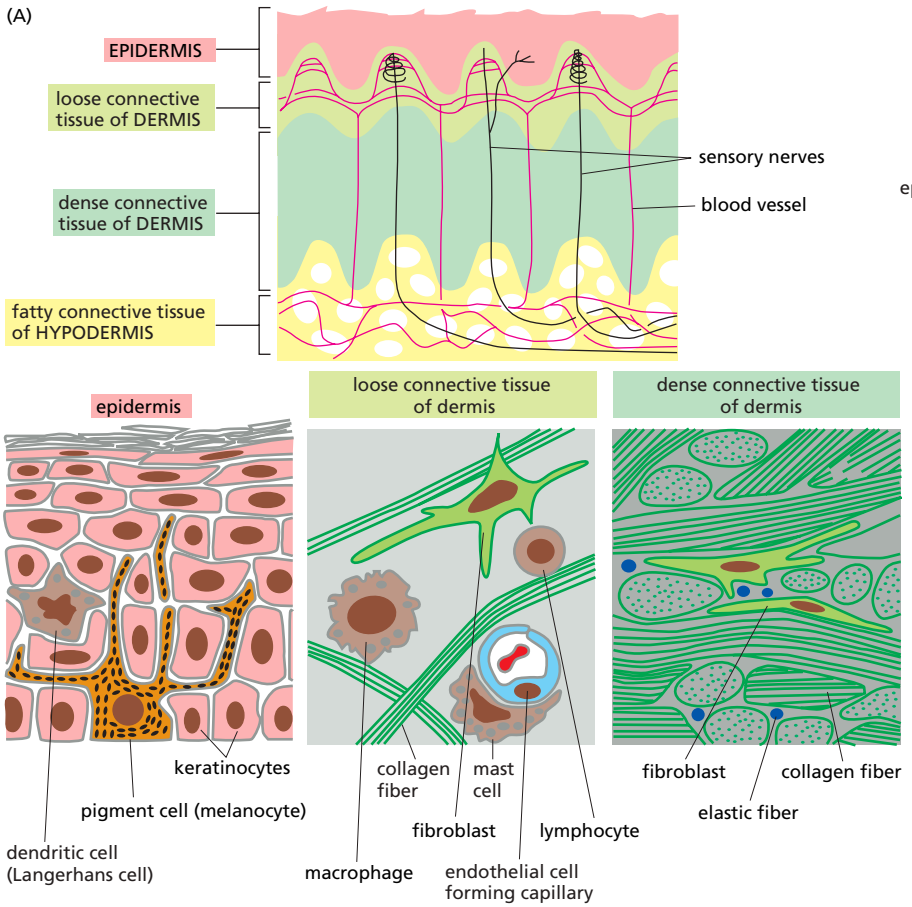




Figure 23–1 Mammalian skin. (A) These diagrams show the cellular architecture of thick skin. (B) Micrograph of a cross section through the sole of a human foot, stained with hematoxylin and eosin. The skin can be viewed as a large organ composed of two main tissues: the epidermis, and the underlying connective tissue, which consists of the dermis and the hypodermis. Each tissue is composed of several different cell types. The dermis and hypodermis are richly supplied with blood vessels and nerves. Some nerve fibers extend into the epidermis.

deliver signals in the opposite direction for glandular secretion and smooth muscle contraction.

Figure 23–1 illustrates the architecture of the skin and shows how it satisfies all these requirements. An epithelium, the **epidermis**, forms the outer covering, creating a waterproof barrier that is self-repairing and continually renewed. Beneath this lies a relatively thick layer of connective tissue, which includes the tough collagen-rich **dermis** (from which leather is made) and the underlying fatty *subcutaneous layer* or *hypodermis*. In the skin, as elsewhere, the connective tissue, with vessels and nerves running through it, provides most of the general supportive functions listed above. The epidermis, however, is the fundamental, quintessential component of the skin—the tissue that is peculiar to this organ, even though not the major part of its bulk. Appendages such as hairs, fingernails, sebaceous glands, and sweat glands develop as specializations of the epidermis (**Figure 23–2**). Complex mechanisms regulate the distribution of these structures and their distinctive patterns of growth and renewal. The regions of less specialized, more or less flat epithelium covering the body surface between the hair follicles and other appendages are called *interfollicular epidermis*. This has a simple organization, and it provides a good introduction to the way in which tissues of the adult body are continually renewed.

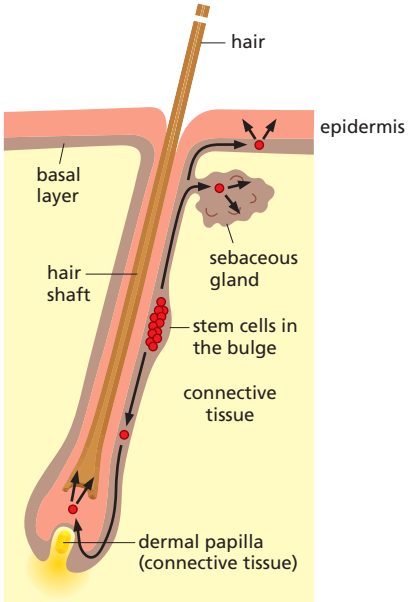


Figure 23–2 A hair follicle and its associated sebaceous gland. These structures form as specializations of the epidermis. The hair grows upward from the papilla at its base. The sebaceous gland contains cells loaded with lipid, which is secreted to keep the hair properly oiled. The whole structure undergoes cycles of growth, regression (when the hair falls out), and reconstruction. Like the rest of the epidermis, it depends on stem cells for its growth and reconstruction in each cycle. An important group of stem cells *(red),* able to give rise to both hair follicle and interfollicular epidermis, lie in a region called the bulge, just below the sebaceous gland.

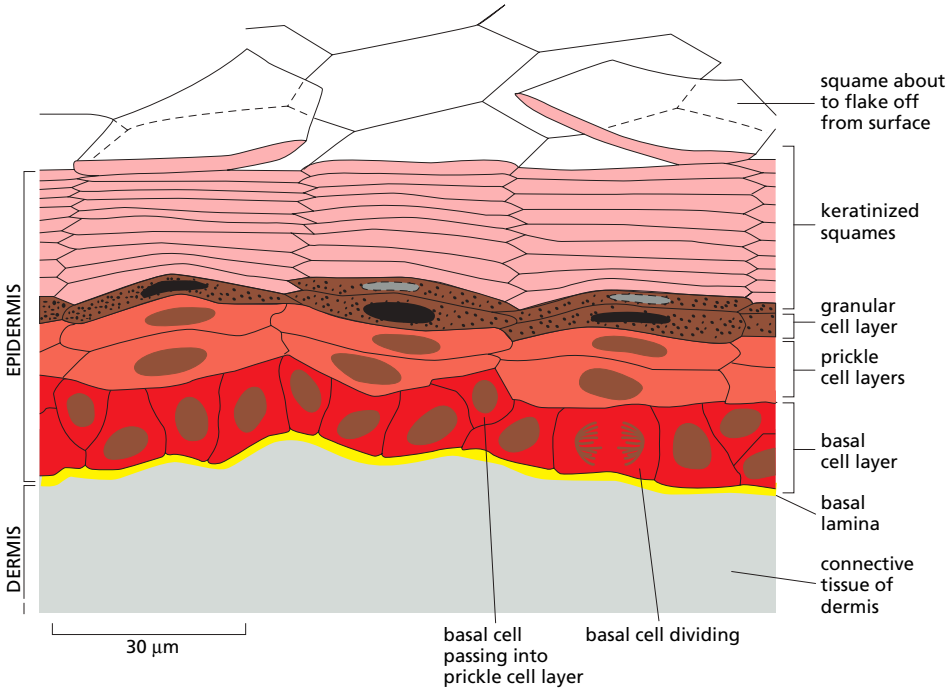


Figure 23–3 The multilayered structure of the epidermis, as seen in a mouse. The outlines of the keratinized squames are revealed by swelling them in a solution containing sodium hydroxide. The highly ordered hexagonal arrangement of interlocking columns of cells shown here occurs only in some sites where the epidermis is thin. In human skin, the stacks of squames are usually many times higher and less regular, and where the skin is very thick mitotic cells are seen not only in the basal layer but also in the first few cell layers above it. In addition to the cells destined for keratinization, the deep layers of the epidermis include small numbers of different types of cells, as indicated in Figure 23–1, including dendritic cells, called Langerhans cells, derived from bone marrow; melanocytes (pigment cells) derived from the neural crest; and Merkel cells, which are associated with nerve endings in the epidermis.

Epidermal Cells Form a Multilayered Waterproof Barrier

The interfollicular epidermis is a multilayered *(stratified)* epithelium composed largely of *keratinocytes* (so named because their characteristic differentiated activity is the synthesis of keratin intermediate filament proteins, which give the epidermis its toughness) (**Figure 23–3**). These cells change their appearance from one layer to the next. Those in the innermost layer, attached to an underlying basal lamina, are termed *basal cells*, and it is usually only these that divide. Above the basal cells are several layers of larger *prickle cells* (**Figure 23–4**), whose numerous desmosomes—each a site of anchorage for thick tufts of keratin filaments—are just visible in the light microscope as tiny prickles around the cell surface (hence the name). Beyond the prickle cells lies the thin, darkly staining granular cell layer (see Figure 23–3). It is at this level that the cells are sealed together to form a waterproof barrier. Mice that fail to form this barrier because of a genetic defect die from rapid fluid loss soon after birth, even though their skin appears normal in other respects.

The granular layer, with its barrier to the movement of water and solutes, marks the boundary between the inner, metabolically active strata and the outermost layer of the epidermis, consisting of dead cells whose intracellular organelles have disappeared. These outermost cells are reduced to flattened scales, or *squames*, filled with densely packed keratin. The plasma membranes of both the squames and the outer granular cells are reinforced on their cytoplasmic surface by a thin (12 nm), tough, cross-linked layer of proteins, including a cytoplasmic protein called involucrin. The squames themselves are normally so compressed and thin that their boundaries are hard to make out in the light microscope, but soaking in sodium hydroxide solution (or a warm bath tub) makes them swell slightly, and their outlines can then be seen (see Figure 23–3).

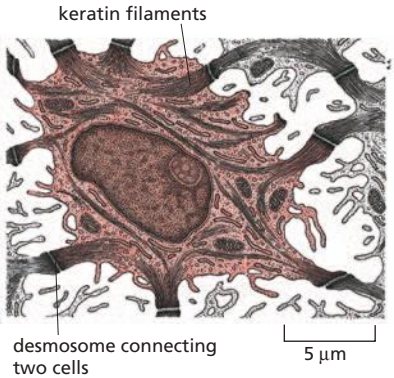


Figure 23–4 A prickle cell. This drawing, from an electron micrograph of a section of the epidermis, shows the bundles of keratin filaments that traverse the cytoplasm and are inserted at the desmosome junctions that bind the prickle cell *(red)* to its neighbors. Nutrients and water diffuse freely through the intercellular spaces in the metabolically active layers of the epidermis occupied by the prickle cells. Farther out, at the level of the granular cells, there is a waterproof barrier that is thought to be created by a sealant material that the granular cells secrete. (From R.V. Krstić, Ultrastructure of the Mammalian Cell: an Atlas. Berlin: Springer-Verlag, 1979.)

Differentiating Epidermal Cells Express a Sequence of Different Genes as They Mature

Let us now set this static picture in motion to see how the epidermis is continually renewed. While some basal cells are dividing, adding to the population in the basal layer, others (their sisters or cousins) are slipping out of the basal cell layer into the prickle cell layer, taking the first step on their outward journey. When they reach the granular layer, the cells start to lose their nucleus and cytoplasmic organelles, through a degradative mechanism that involves partial activation of the machinery of apoptosis; in this way, the cells are transformed into the keratinized squames of the keratinized layer. These finally flake off from the surface of the skin (and become a main constituent of household dust). The time from birth of a cell in the basal layer of the human skin to its loss by shedding from the surface is about a month, depending on body region.

As the new keratinocyte in the basal layer is transformed into the squame in the outermost layers (see Figure 23–4), it steps through a succession of different states of gene expression, synthesizing a succession of different members of the keratin protein family. Meanwhile other characteristic proteins, such as involucrin, also begin to be synthesized as part of a coordinated program of **terminal cell differentiation**—the process in which a precursor cell acquires its final specialized characteristics and usually permanently stops dividing. The whole program is initiated in the basal layer. It is here that the fates of the cells are decided.

Stem Cells in the Basal Layer Provide for Renewal of the Epidermis

Humans renew the outer layers of their epidermis a thousand times over in the course of a lifetime. In the basal layer, there have to be cells that can remain undifferentiated and carry on dividing for this whole period, continually throwing off descendants that commit to differentiation, leave the basal layer, and are eventually discarded. The process can be maintained only if the basal cell population is self-renewing. It must therefore contain some cells that generate a mixture of progeny, including daughters that remain undifferentiated like their parent, as well as daughters that differentiate. Cells with this property are called **stem cells**. They have so important a role in such a variety of tissues that it is useful to have a formal definition.

The defining properties of a stem cell are as follows:

1. It is not itself terminally differentiated (that is, it is not at the end of a pathway of differentiation).
2. It can divide without limit (or at least for the lifetime of the animal).
3. When it divides, each daughter has a choice: it can either remain a stem cell, or it can embark on a course that commits it to terminal differentiation (**Figure 23–5**).

Stem cells are required wherever there is a recurring need to replace differentiated cells that cannot themselves divide. The stem cell itself has to be able to divide—that is part of the definition—but it should be noted that it does not necessarily have to divide rapidly; in fact, stem cells usually divide at a relatively slow rate.

The need for stem cells arises in many different tissues. Thus, stem cells are of many types, specialized for the genesis of different classes of terminally differentiated cells—epidermal stem cells for epidermis, intestinal stem cells for intestinal epithelium, hemopoietic stem cells for blood, and so on. Each stem-cell system nevertheless raises similar fundamental questions. What are the distinguishing features of the stem cell in molecular terms? What factors determine

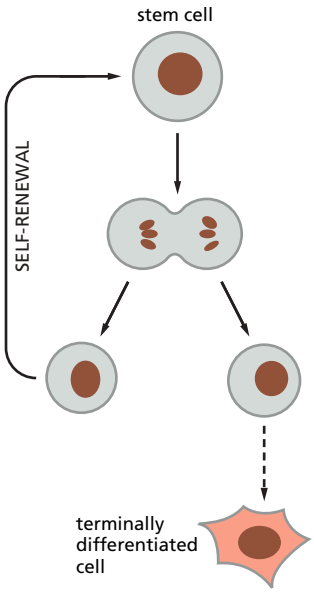


Figure 23–5 The definition of a stem cell. Each daughter produced when a stem cell divides can either remain a stem cell or go on to become terminally differentiated. In many cases, the daughter that opts for terminal differentiation undergoes additional cell divisions before terminal differentiation is completed.

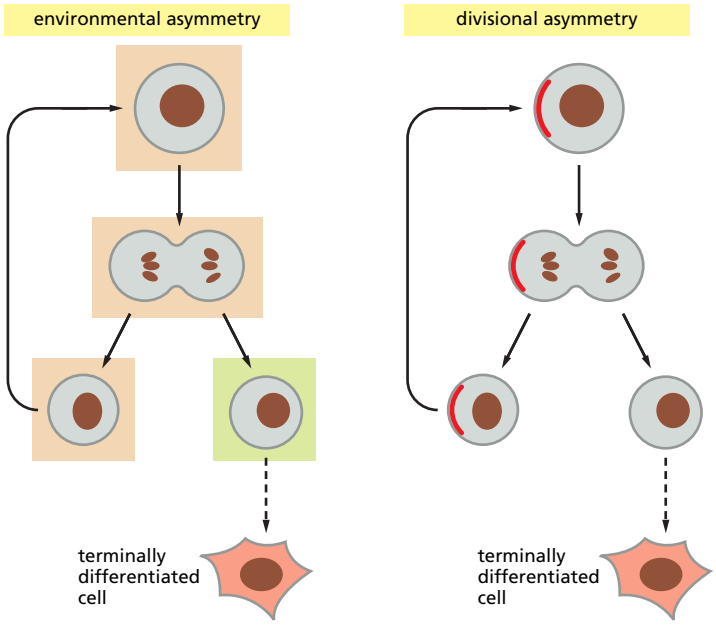


Figure 23–6 Two ways for a stem cell to produce daughters with different fates. In the strategy based on environmental asymmetry, the daughters of the stem cell are initially similar and are directed into different pathways according to the environmental influences that act on them after they are born. The environment is shown as *colored shading* around the cell. With this strategy, the number of stem cells can be increased or reduced to fit the niche available for them. In the strategy based on divisional asymmetry, the stem cell has an internal asymmetry and divides in such a way that its two daughters are already endowed with different determinants at the time of their birth. In some cases, the choice between the alternative fates may be made at random for each daughter, but with a defined probability, like a coin-toss, reflecting the intrinsic randomness or "noise" in all genetic control systems (discussed in Chapter 7).

whether it divides or stays quiescent? What decides whether a given daughter cell commits to differentiation or remains a stem cell? And where the stem cell can give rise to more than one kind of differentiated cell—as is very often the case—what determines which differentiation pathway is followed?

The Two Daughters of a Stem Cell Do Not Always Have to Become Different

At steady state, to maintain a stable stem-cell population, precisely 50% of the daughters of stem cells in each cell generation must remain as stem cells. In principle, this could be achieved in two ways—through *environmental asymmetry* or through *divisional asymmetry* (**Figure 23–6**). In the first strategy, the division of a stem cell could generate two initially similar daughters whose fates would be governed by their subsequent environment or by some random process with an appropriate environmentally controlled probability; 50% of the *population* of daughters would remain as stem cells, but the two daughters of an individual stem cell in the population might often have the same fate. At the opposite extreme, the stem cell division could be always strictly asymmetric, producing one daughter that inherits the stem-cell character and another that inherits factors that force it to embark on differentiation. The neuroblasts of the *Drosophila* central nervous system, discussed in Chapter 22, are an example of cells that show this type of divisional asymmetry. This strategy in its strict form has a drawback, however: it means that the existing stem cells can never increase their numbers, and any loss of stem cells is irreparable, unless by recruitment of some other type of cell to become a stem cell. The strategy of control by environmental asymmetry is more flexible.

In fact, if a patch of epidermis is destroyed, the surrounding epidermal cells repair the damage by migrating in and proliferating to cover the denuded area. In this process, a new self-renewing patch of epidermis is established, implying that additional stem cells have been generated to make up for the loss. These must have been produced by symmetric divisions in which one stem cell gives rise to two. In this way, the stem cell population adjusts its numbers to fit the available niche.

Observations such as these suggest that the maintenance of stem cell character in the epidermis might be controlled by contact with the basal lamina, with a loss of contact triggering the start of terminal differentiation, and maintenance of contact serving to preserve stem cell potential. This idea contains a grain of truth, but it is not the whole truth. As we now explain, not all the cells in the basal layer have the capacity to serve as stem cells.

The Basal Layer Contains Both Stem Cells and Transit Amplifying Cells

Basal keratinocytes can be dissociated from intact epidermis and can proliferate in a culture dish, giving rise to new basal cells and to terminally differentiated cells. Even within a population of cultured basal keratinocytes that all seem undifferentiated, there is great variation in the ability to proliferate. When human keratinocytes are taken singly and tested for their ability to found new colonies, some seem unable to divide at all, others go through only a few division cycles and then halt, and still others divide enough times to form large colonies. This proliferative potential directly correlates with the expression of the β1 subunit of integrin, which helps mediate adhesion to the basal lamina. Clusters of cells with high levels of this molecule are found in the basal layer of the intact human epidermis also, and they are thought to contain the stem cells (**Figure 23–7**). We still do not have definitive markers for the stem cells themselves, and we still do not understand in molecular terms what it is that fundamentally defines the stem-cell state. This is one of the key problems of stem-cell biology, and we shall say more about it in later sections of the chapter.

Paradoxically, many if not all of the epidermal cells that generate large colonies in culture seem to be cells that themselves as a rule divide rarely. One line of evidence comes from experiments in which a pulse of the thymidine analog bromodeoxyuridine (BrdU) is given to a young animal, in which the epidermis is growing rapidly, or to a mature animal following an injury that provokes rapid repair. One then waits for many days or weeks before fixing the tissue and staining with an antibody that recognizes DNA in which BrdU has been incorporated. The BrdU is taken up by any cell that is in S phase of the division cycle at the time of the initial pulse. Because the BrdU would be expected then to be diluted by half at each subsequent cell division, any cells that remain strongly labeled at the time of fixation are assumed to have undergone few or no divisions since replicating their DNA at the time of the pulse. Such *label-retaining cells* can be seen scattered among unlabeled or lightly labeled cells in the basal layer of the epidermis even after a period of several months, and large numbers are visible in hair follicles, in a region called the bulge (see Figure 23–2). Ingenious labeling procedures indicate that the label-retaining cells, in the hair follicle at least, are in fact stem cells: when a new cycle of hair growth begins after an old hair has been shed, the label-retaining cells in the bulge at last divide and contribute the

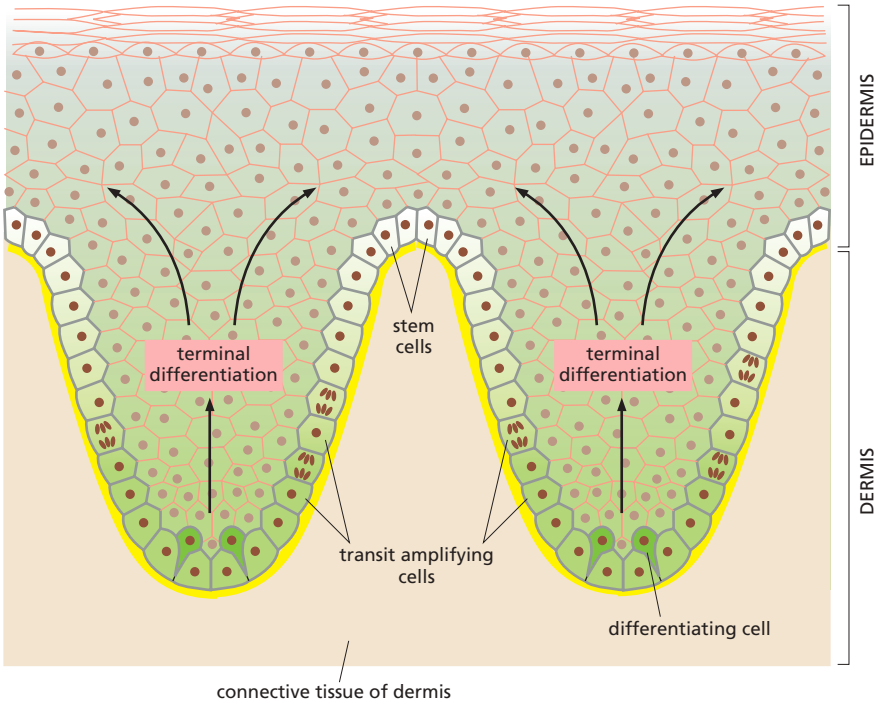


Figure 23–7 The distribution of stem cells in human epidermis, and the pattern of epidermal cell production. The diagram is based on specimens in which the location of the stem cells was identified by staining for β1 integrin, and that of the differentiating cells by staining for keratin-10, a marker of keratinocyte differentiation; dividing cells were identified by labeling with BrdU, a thymidine analog that is incorporated into cells in S phase of the cell division cycle. The stem cells seem to be clustered near the tips of the dermal papillae. They divide infrequently, giving rise (through a sideways movement) to transit amplifying cells, which occupy the intervening regions. The transit amplifying cells divide frequently, but for a limited number of division cycles, at the end of which they begin to differentiate and slip out of the basal layer. The precise distribution of stem cells and transit amplifying cells varies from one region of epidermis to another. (Adapted from S. Lowell et al., *Curr. Biol.* 10:491–500, 2000. With permission from Elsevier.)

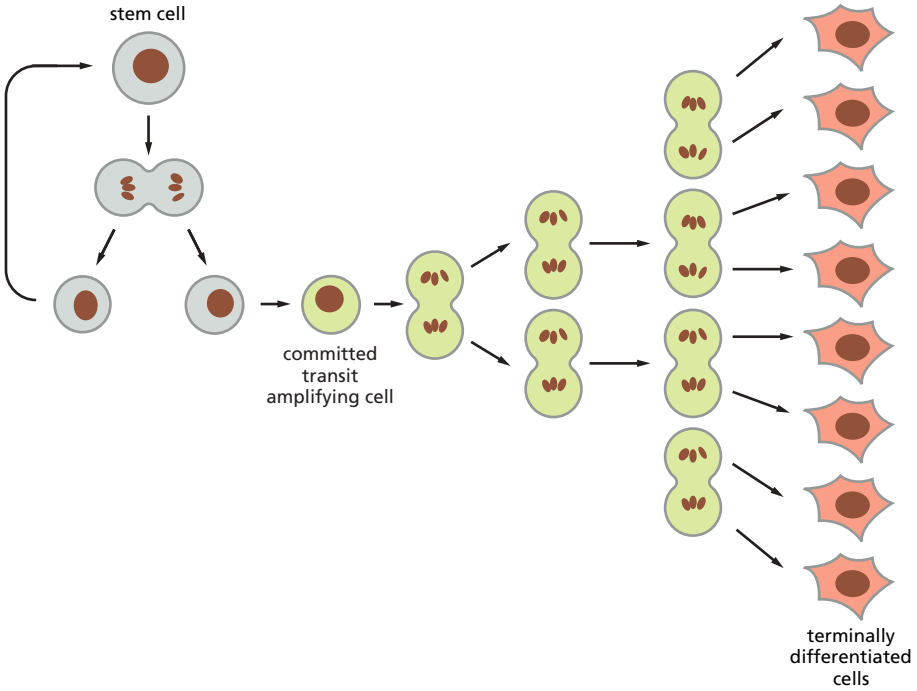


Figure 23–8 Transit amplifying cells. Stem cells in many tissues divide only rarely but give rise to transit amplifying cells—daughters committed to differentiation that go through a limited series of more rapid divisions before completing the process. In the example shown here, each stem cell division gives rise in this way to eight terminally differentiated progeny.

cells that go to form the regenerated hair follicle. Although it is not certain that all the stem cells of the hair follicle have this label-retaining character, some clearly do, and the same seems to be true of the stem cells in the interfollicular epidermis. Moreover, basal cells expressing β1 integrin at a high level—the cells that can give rise to large colonies in culture —are rarely seen dividing.

Mixed with these cells there are others that divide more frequently—but only for a limited number of division cycles, after which they leave the basal layer and differentiate. These latter cells are called **transit amplifying cells**—"transit", because they are in transit from a stem-cell character to a differentiated character; and "amplifying", because the division cycles they go through have the effect of amplifying the number of differentiated progeny that result from a single stem-cell division (**Figure 23–8**). In this way, a small population of stem cells that divide only rarely can generate a plentiful supply of new differentiated cells.

Transit Amplifying Divisions Are Part of the Strategy of Growth Control

Transit amplifying cells are a common feature of stem cell systems. This means that in most such systems there are few true stem cells and they are mixed with a much larger number of progeny cells that have only a limited capacity to divide. As discussed in Chapter 20, the same seems to be true not only of normal self-renewing tissues but also for many cancers, where only a small minority of cells in the tumor cell population are capable of serving as cancer stem cells. Why should this be? There are several possible answers, but a part of the explanation probably lies in the strategy by which large multicellular animals (such as mammals) control the sizes of their cell populations.

The proportions of the parts of the body are mostly determined early, during development, by means of signals that operate over distances of a few hundred cell diameters at most: for each organ or tissue, a small rudiment or founder cell population is delimited in this way. The founder cell populations must then grow, but—in mammals at least—only up to a certain definite limit, at which point they must stop.

One way to halt growth at a certain size is by feedback signals that operate over much larger distances in the mature organism; we shall see that such signals indeed play an important part in controlling the growth of at least some tissues. Another strategy, however, is to endow each founder cell with an internal

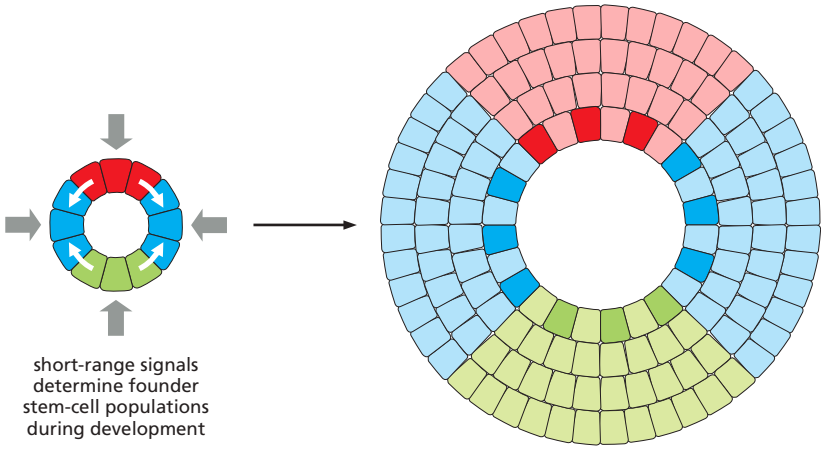


Figure 23–9 One way to define the size of a large organ. In the embryo, short-range signals determine small groups of cells as founders of the different cell populations. Each founder can be programmed then to divide a certain number of times, giving rise to a large set of cells in the adult. If the adult organ is to be renewed while maintaining its proper size, the founders can be programmed to divide as stem cells, giving rise at each division, on average, to one daughter that remains as a stem cell and another that is programmed to go through a set number of transit amplifying divisions.

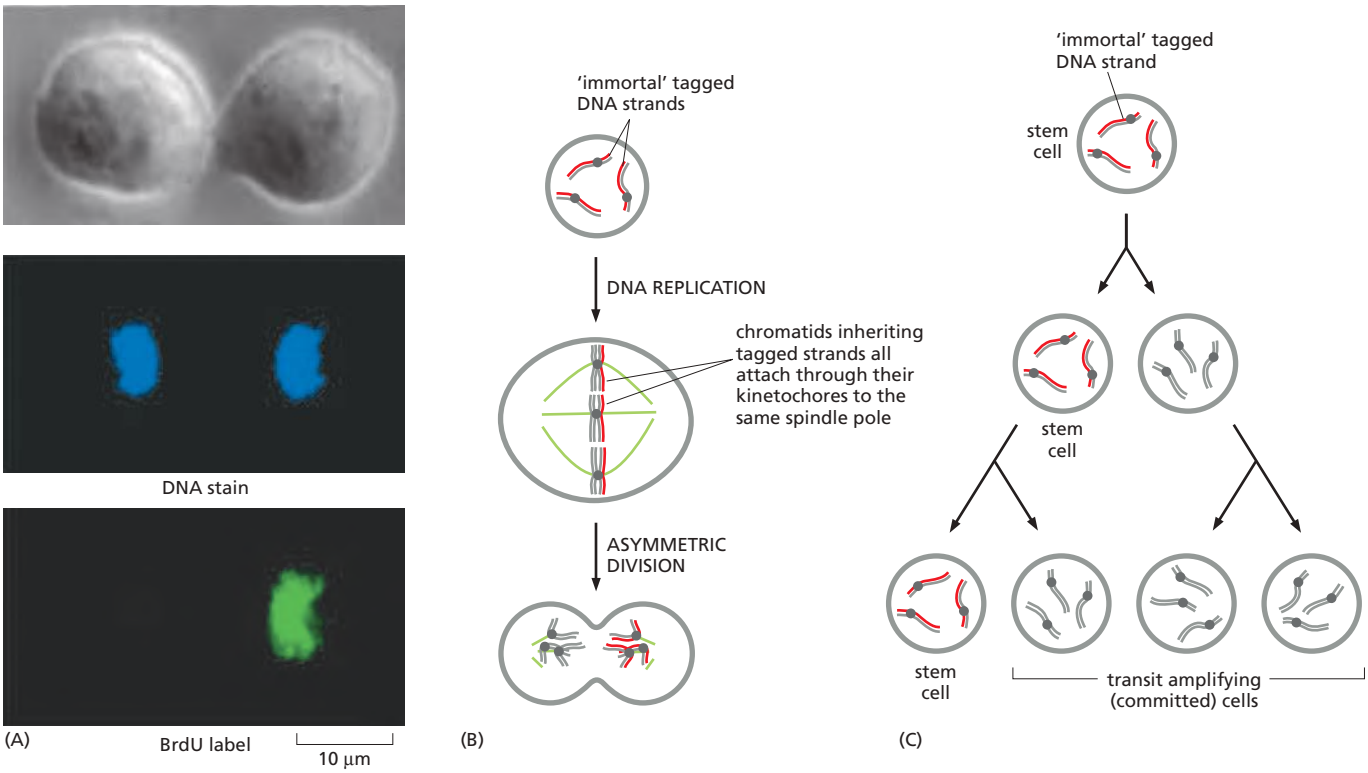
program dictating that it shall divide a limited number of times and then stop. In this way, short-range signals during development can define the size of large final structures (**Figure 23–9**). But if that is the strategy, how can the adult tissue be continually renewed? A solution is to specify the founder cells as stem cells, able to continue dividing indefinitely, but producing at each division one daughter that remains as a stem cell and one that is programmed to go through a limited number of transit amplifying divisions and then stop.

This is certainly an oversimplified and incomplete account of the control of tissue growth and renewal, but it helps to explain why cells that are programmed to undergo long sequences of cell divisions and then halt are such a common feature of animal development and why tissue renewal by stem cells so often involves transit amplifying divisions.

Stem Cells of Some Tissues Selectively Retain Original DNA Strands

Stem cells in many tissues appear to be label-retaining cells. As we have just explained, this has generally been assumed to be because, having incorporated a tracer such as BrdU into their DNA during a period of BrdU exposure, the stem cells then divide rarely, so that the label is only slowly diluted by newly synthesized DNA. There is, however, another possible interpretation: regardless of whether they divide fast or slowly, the stem cells might segregate their DNA strands asymmetrically, in such a way that in every division, and for every chromosome, the specific DNA strand that was originally labeled is retained in the daughter cell that remains a stem cell. This original strand would presumably have to have acquired some sort of special tag, designating it as a stem-cell strand and ensuring that it segregates asymmetrically, into the daughter that remains a stem cell (see Figure 23–6), along with all the similarly tagged DNA strands of the other chromosomes; in this way, the old labeled strands would be retained in the stem cells from cell generation to cell generation. The tag might, for example, take the form of some special kinetochore protein that remains associated with the old DNA strand at the centromere of each chromosome during DNA replication and then engages with some asymmetry in the mitotic spindle so as to ensure that the stem-cell daughter receives all the daughter chromosomes carrying the tag. In each stem-cell generation, the same original tagged DNA strands would then serve as templates for production of the new sets of DNA strands to be despatched into the transit amplifying cells in the following generation (**Figure 23–10**).

This "immortal strand" hypothesis may seem a lot to swallow, given that no mechanism for such tagging and segregation of individual DNA strands has yet been identified. Yet there is increasing evidence suggesting that the immortal strand hypothesis is correct. Muscle (described later in this chapter) provides an



example. When BrdU is supplied during a period of production of muscle stem cells and the subsequent fate of the cells is followed as they divide and proliferate, it is possible to detect small clones of cells, or pairs of sister cells, within which all the BrdU label is concentrated in a single cell, even though all the cells share a common origin from a single ancestor cell that took up the label initially. Similar observations have been reported in studies of other types of stem cells, and, importantly, this behavior has not been seen in cell populations that do not contain stem cells. The immortal strand hypothesis would not only explain why stem cells retain labeled DNA indefinitely, but would also imply that asymmetric division is a fundamental stem-cell property, with the corollary that any increase in the number of stem cells must require special conditions to confer the immortality tag on additional, newly synthesized DNA strands. The immortal strand hypothesis was originally proposed in the 1970s as a mechanism for stem cells to avoid accumulating cancer-promoting mutations during DNA replication. Reduction of the risk of cancer could be one of its benefits.

The Rate of Stem-Cell Division Can Increase Dramatically When New Cells Are Needed Urgently

Whatever the mechanism of stem-cell maintenance may be, the use of transit amplifying divisions brings several benefits. First, it means that the number of stem cells can be small and their division rate can be low, even when terminally differentiated cells have to be produced rapidly in large numbers. This reduces the cumulative burden of genetic damage, since most mutations occur in the course of DNA replication and mitosis, and mutations occurring in cells that are not stem cells are discarded in the course of tissue renewal. The likelihood of cancer is thus reduced. If the immortal strand hypothesis is correct, so that stem cells always retain the original "immortal" template DNA strands, the risk is still further reduced, since most sequence errors introduced during DNA replication will be in the newly synthesized strands, which the stem cells ultimately discard.

Second, and perhaps more important, a low stem-cell division rate in normal circumstances allows for dramatic increase when there is an urgent need - for example, in wound repair. The stem cells can then be roused to divide rapidly, and the additional division cycles can both amplify the stock of stem

Figure 23–10 The immortal strand hypothesis. (A) Experimental evidence. Here, stem cells of skeletal muscle (members of the muscle satellite cell population, discussed later in this chapter) have been placed in culture and allowed to divide for 4 days in the presence of BrdU to label newly synthesized DNA strands. The cells have then been allowed to divide for 1 day in the absence of BrdU. The photographs show a pair of sister cells at the end of this procedure: one has inherited BrdU, the other has not. This implies that daughter chromosomes carrying DNA strands synthesized during the cell divisions that occurred in the presence of BrdU have all been inherited by the one cell, while those carrying only either pre-existing or subsequently synthesized DNA strands have been inherited by the other. This phenomenon, in which old and new DNA strands are asymmetrically allocated to different daughter cells, is seen only in cell populations that include stem cells. (B) The pattern of DNA strand inheritance in stem cells according to the immortal strand hypothesis. One strand in each chromosome in the stem cell is somehow tagged as the immortal strand and is retained by the stem-cell daughter. (C) This original DNA strand remains available through all subsequent stem-cell generations as a template for production of chromosomes of transit amplifying cells. (A, from V. Shinin, B. Gayraud-Morel, D. Gomès and S. Tajbakhsh, *Nat. Cell Biol.* 8:677–687, 2006. With permission from Macmillan Publishers Ltd.)

cells and increase steeply the production of cells committed to terminal differentiation. Thus, for example, when a patch of hairy skin is cut away, the slowly dividing stem cells in the bulge region of surviving hair follicles near the wound are switched into rapid proliferation, and some of their progeny move out as new stem cells to form fresh interfollicular epidermis to cover the wounded patch of body surface.

Many Interacting Signals Govern Epidermal Renewal

Cell turnover in the epidermis seems at first glance a simple matter, but the simplicity is deceptive. There are many points in the process that have to be controlled according to circumstances: the rate of stem-cell division; the probability that a stem-cell daughter will remain a stem cell; the number of cell divisions of the transit amplifying cells; the timing of exit from the basal layer, and the time that the cell then takes to complete its differentiation program and be sloughed from the surface. Regulation of these steps must enable the epidermis to respond to rough usage by becoming thick and callused, and to repair itself when wounded. In specialized regions of epidermis, such as those that form hair follicles, with their own specialized subtypes of stem cells, yet more controls are needed to organize the local pattern.

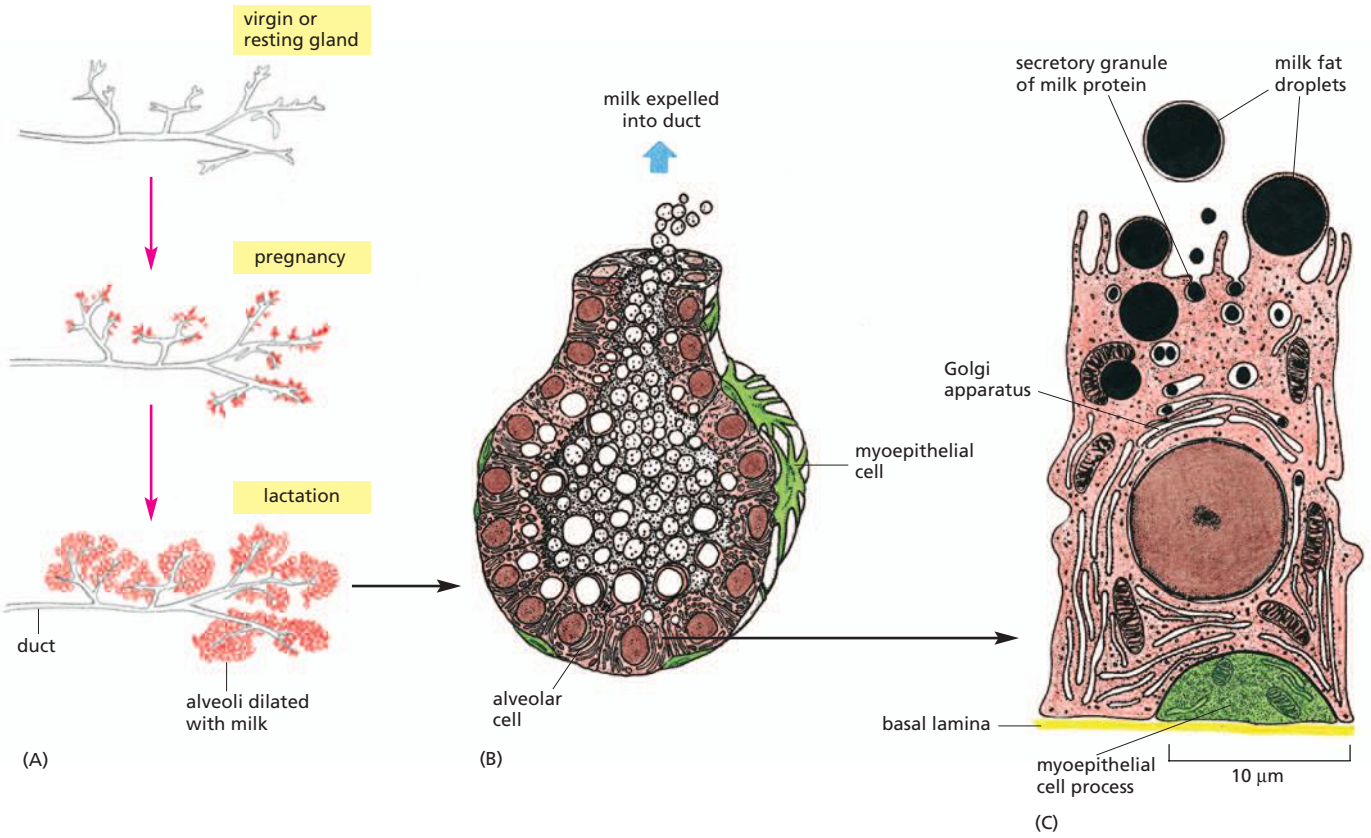
Each of the control points has its own importance, and a multitude of molecular signals is needed to regulate them all, so as to keep the body surface always properly covered. As we suggested earlier, one important influence is contact with the basal lamina, signaled via integrins in the plasma membrane of the cells. If cultured basal keratinocytes are held in suspension, instead of being allowed to settle and attach to the bottom of the culture dish, they all stop dividing and differentiate. To remain as an epidermal stem cell, it is apparently necessary (although not sufficient) to be attached to the basal lamina or other extracellular matrix. This requirement helps ensure that the size of the stem cell population does not increase without limit. If crowded out of their regular niche on the basal lamina, the cells lose their stem cell character. When this rule is broken, as in some cancers, the result can be an ever-growing tumor.

Most of the other cell communication mechanisms described in Chapter 15 are also implicated in the control of epidermal renewal, either in signaling between cells within the epidermis or in signaling between epidermis and dermis. The EGF, FGF, Wnt, Hedgehog, Notch, BMP/TGFβ, and integrin signaling pathways are all involved (and we shall see that the same is true of most other tissues). Overactivation of the Hedgehog pathway, for example, can cause epidermal cells to carry on dividing after they have left the basal layer, and mutations in components of this pathway are responsible for many epidermal cancers. At the same time, Hedgehog signaling helps to guide the choice of differentiation pathway: a deficit of Hedgehog signaling leads to loss of sebaceous glands, while an excess can cause sebaceous glands to develop in regions where they would never normally form. Similarly, loss of Wnt signaling leads to failure of hair follicle development, while excessive activation of this pathway causes extra hair follicles to form and to grow excessively so that they give rise to tumors.

Notch signaling, in contrast, seems to restrict the size of the stem cell population, inhibiting neighbors of stem cells from remaining as stem cells and causing them to become transit amplifying cells instead. And TGFβ has a key role in signaling to the dermis during the repair of skin wounds, promoting the formation of collagen-rich scar tissue. The precise individual functions of all the various signaling mechanisms in the epidermis are only beginning to be disentangled.

The Mammary Gland Undergoes Cycles of Development and Regression

In specialized regions of the body surface, various other types of cells develop from the embryonic epidermis. In particular, secretions such as sweat, tears,



saliva, and milk are produced by cells segregated in deep-lying glands that originate as ingrowths of the epidermis. These epithelial structures have functions and patterns of renewal quite different from those of keratinizing regions.

The mammary glands are the largest and most remarkable of these secretory organs. They are the defining feature of mammals and an important concern in many ways: not only for nourishment of babies and attraction of the opposite sex, but also as the basis for a large industry—the dairy industry—and as the site of some of the commonest forms of cancer. Mammary tissue illustrates most dramatically that developmental processes continue in the adult body; and it shows how cell death by apoptosis permits cycles of growth and regression.

Milk production must be switched on when a baby is born and switched off when the baby is weaned. During pregnancy, the producer cells of the milk factory are generated; at weaning, they are destroyed. A "resting" adult mammary gland consists of branching systems of ducts embedded in fatty connective tissue; this is the future plumbing network that will deliver milk to the nipple. The ducts are lined with an epithelium that includes mammary stem cells. These stem cells can be identified by a functional test, in which the cells of the mammary tissue are dissociated, sorted according to the cell surface markers that they express, and transplanted back into appropriate host tissue (a mammary fat pad). This assay reveals that a small subset of the total epithelial cells have stem-cell potential. A single one of these cells, estimated to be about one in 5000 of the total mammary epithelial population but more concentrated within a subpopulation expressing certain markers, can proliferate indefinitely and give rise to an entire new mammary gland with all its epithelial cell types. This reconstituted gland is able to go through the full program of differentiation required for milk production. As a first step toward milk production, the steroid hormones that circulate during pregnancy (estrogen and progesterone) cause the duct cells to proliferate, increasing their numbers several hundred-fold. In a process that depends on local activation of the Wnt signaling pathway, the terminal regions of the ducts grow and branch, forming little dilated outpocketings, or

Figure 23–11 The mammary gland. (A) The growth of alveoli from the ducts of the mammary gland during pregnancy and lactation. Only a small part of the gland is shown. The "resting" gland contains a small amount of inactive glandular tissue embedded in a large amount of fatty connective tissue. During pregnancy an enormous proliferation of the glandular tissue takes place at the expense of the fatty connective tissue, with the secretory portions of the gland developing preferentially to create alveoli. (B) One of the milk-secreting alveoli with a basket of contractile myoepithelial cells *(green)* embracing it (see also Figure 23–47E). (C) A single type of secretory alveolar cell produces both the milk proteins and the milk fat. The proteins are secreted in the normal way by exocytosis, while the fat is released as droplets surrounded by plasma membrane detached from the cell. (B, after R. Krstić, Die Gewebe des Menschen und der Säugetiere. Berlin: Springer-Verlag, 1978; C, from D.W. Fawcett, A Textbook of Histology, 12th ed. New York: Chapman and Hall, 1994.)

alveoli, containing secretory cells (**Figure 23–11**). Milk secretion begins only when these cells are stimulated by the different combination of hormones circulating in the mother after the birth of the baby, especially prolactin from the pituitary gland. Prolactin binds to receptors on the surface of the mammary epithelial cells and thereby activates a pathway that switches on expression of milk protein genes. As in the epidermis, signals from the extracellular matrix, mediated by integrins, are also essential: the milk-producing cells can only respond to prolactin if they are also in contact with the basal lamina. A further tier of hormonal control governs the actual ejection of milk from the breast: the stimulus of suckling causes cells of the hypothalamus (in the brain) to release the hormone *oxytocin*, which travels via the bloodstream to act on *myoepithelial cells*. These musclelike cells originate from the same epithelial precursor population as the secretory cells of the breast, and they have long spidery processes that embrace the alveoli. In response to oxytocin they contract, thereby squirting milk out of the alveoli into the ducts.

Eventually, when the baby is weaned and suckling stops, the secretory cells die by apoptosis, and most of the alveoli disappear. Macrophages rapidly clear away the dead cells, matrix metalloproteinases degrade the surplus extracellular matrix, and the gland reverts to its resting state. This ending of lactation seems to be induced by the accumulation of milk, rather than by a hormonal mechanism. If one subset of mammary ducts is obstructed so that no milk can be discharged, the secretory cells that supply it commit mass suicide by apoptosis, while other regions of the gland survive and continue to function. The apoptosis is triggered by a complex array of factors that accumulate where milk secretion is blocked.

Cell division in the growing mammary gland is regulated not only by hormones but also by local signals passing between cells within the epithelium and between the epithelial cells and the connective tissue, or *stroma,* in which the epithelial cells are embedded. All the signals listed earlier as important in controlling cell turnover in the epidermis are also implicated in controlling events in the mammary gland. Again, signals delivered via integrins play a crucial part: deprived of the basal lamina adhesions that activate integrin signaling, the epithelial cells fail to respond normally to hormonal signals. Faults in these interacting control systems underlie some of the commonest forms of cancer, and we need to understand them better.

Summary

Skin consists of a tough connective tissue, the dermis, overlaid by a multilayered waterproof epithelium, the epidermis. The epidermis is continually renewed from stem cells, with a turnover time, in humans, on the order of a month. Stem cells, by definition, are not terminally differentiated and have the ability to divide throughout the organism's lifetime, yielding some progeny that differentiate and others that remain stem cells. The epidermal stem cells lie in the basal layer, attached to the basal lamina; under normal conditions, their division rate is low. Progeny that become committed to differentiation go through several rapid transit amplifying divisions in the basal layer, and then stop dividing and move out toward the surface of the skin. They progressively differentiate, switching from expression of one set of keratins to expression of another until, eventually, their nuclei degenerate, producing an outer layer of dead keratinized cells that are continually shed from the surface.

The fate of the daughters of a stem cell is controlled by interactions with the basal lamina, mediated by integrins and by signals from neighboring cells. Some types of stem cells may also be internally programmed to divide asymmetrically so as to create one stem-cell daughter and one daughter committed to eventual differentiation; this may involve selective segregation of original "immortal" template DNA strands into the stem-cell daughter. Environmental controls, however, allow two stem cells to be generated from one during repair processes and can trigger steep increases in the rate of stem-cell division. Factors such as Wnt and Hedgehog signal proteins not only regulate the rate of cell proliferation according to need, but can also drive specialization of epidermal cells to form structures such as hair follicles and sebaceous glands.

These and other organs connected to the epidermis, such as the mammary glands, have their own stem cells and their own distinct patterns of cell turnover. In the breast, for example, circulating hormones stimulate the cells to proliferate, differentiate, and make milk; the cessation of suckling triggers the milk-secreting cells to die by apoptosis, in response to a combination of factors that build up where milk fails to be drained away.

SENSORY EPITHELIA

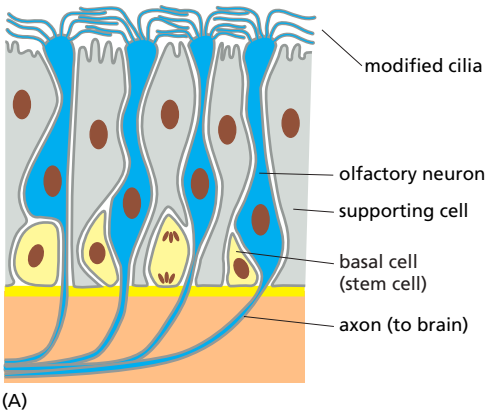
We sense the smells, sounds, and sights of the external world through another class of specializations of the epithelium that cover our body surface. The sensory tissues of the nose, the ears, and the eyes—and, indeed, if we look back to origins in the early embryo, the whole of the central nervous system—all arise from the same sheet of cells, the *ectoderm*, that gives rise to the epidermis. These structures have several features in common, and related systems of genes govern their development (discussed in Chapter 22). They all retain an epithelial organization, but it is very different from that of the ordinary epidermis or of the glands that derive from it.

The nose, the ear, and the eye are complex organs, with elaborate devices to collect signals from the external world and to deliver them, filtered and concentrated, to the sensory epithelia, where they can act on the nervous system. The sensory epithelium in each organ is the key component, although it is small relative to all the ancillary apparatus. It is the part that has been most highly conserved in evolution, not only from one vertebrate to another, but also between vertebrates and invertebrates.

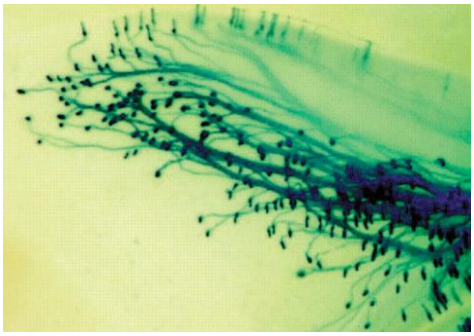
Within each sensory epithelium lie sensory cells that act as *transducers*, converting signals from the outside world into an electrical form that the nervous system can interpret. In the nose, the sensory transducers are *olfactory sensory neurons*; in the ear, *auditory hair cells*; and in the eye, *photoreceptors*. All of these cell types are either neurons or neuron-like. Each carries at its apical end a specialized structure that detects the external stimulus and converts it to a change in the membrane potential. At its basal end, each makes synapses with neurons that relay the sensory information to specific sites in the brain.

Olfactory Sensory Neurons Are Continually Replaced

In the olfactory epithelium of the nose (**Figure 23–12**A), a subset of the epithelial cells differentiate as **olfactory sensory neurons**. These cells have modified, immotile cilia on their free surfaces (see Figure 15–46), containing odorant receptor proteins, and a single axon extending from their basal end toward the brain (Figure 23–12B). *Supporting cells* that span the thickened epithelium and have properties similar to those of glial cells in the central nervous system hold



(A)


(B)
200 μm


(C)
500 μm

Figure 23–12 Olfactory epithelium and olfactory neurons. (A) Olfactory epithelium consists of supporting cells, basal cells, and olfactory sensory neurons. The basal cells are the stem cells for production of the olfactory neurons. Six to eight modified cilia project from the apex of the olfactory neuron and contain the odorant receptors. (B) This micrograph shows olfactory neurons in the nose of a genetically modified mouse in which the *LacZ* gene has been inserted into an odorant receptor locus, so that all the cells that would normally express that particular receptor now also make the enzyme β-galactosidase. The β-galactosidase is detected through the blue product of the enzymatic reaction that it catalyzes. The cell bodies *(dark blue)* of the marked olfactory neurons, lying scattered in the olfactory epithelium, send their axons *(light blue)* toward the brain (out of the picture to the right). (C) A cross section of the *left* and *right* olfactory bulbs, stained for β-galactosidase. Axons of all the olfactory neurons expressing the same odorant receptor converge on the same glomeruli *(red arrows)* symmetrically placed within the bulbs on the *right* and *left* sides of the brain. Other glomeruli (unstained) receive their inputs from olfactory neurons expressing other odorant receptors. (B and C, from P. Mombaerts et al., *Cell* 87:675–686, 1996. With permission from Elsevier.)

the neurons in place and separate them from one another. The sensory surfaces are kept moist and protected by a layer of fluid secreted by cells sequestered in glands that communicate with the exposed surface. Even with this protection, however, each olfactory neuron survives for only a month or two, and so a third class of cells—the *basal cells*—is present in the epithelium to generate replacements for the olfactory neurons that are lost. The population of basal cells, lying in contact with the basal lamina, includes stem cells for the production of the neurons.

As discussed in Chapter 15, the genome contains a remarkably large number of odorant receptor genes—about 1000 in a mouse or a dog, and about 350 (plus many more that are degenerate and non-functional) in a human. Each olfactory neuron most probably expresses only one of these genes, enabling the cell to respond to one particular class of odorants (generally small organic molecules) sharing some structural feature that the odorant receptor protein recognizes. But regardless of the odor, every olfactory neuron responds in the same way—it sends a train of action potentials back along its axon to the brain. The discriminating sensibility of an individual olfactory neuron is therefore useful only if its axon delivers its messages to the specific relay station in the brain that is dedicated to the particular range of odors that the neuron senses. These relay stations are called *glomeruli*. They are located in structures called the olfactory bulbs (one on each side of the brain), with about 1800 glomeruli in each bulb (in the mouse). Olfactory neurons expressing the same odorant receptor are widely scattered in the olfactory epithelium, but their axons all converge on the same glomerulus (Figure 23–12C). As new olfactory neurons are generated, replacing those that die, they must in turn send their axons to the correct glomerulus. The odorant receptor proteins thus have a second function: guiding the growing tips of the new axons along specific paths to the appropriate target glomeruli in the olfactory bulbs. If it were not for the continual operation of this guidance system, a rose might smell in one month like a lemon, in the next like rotting fish.

Auditory Hair Cells Have to Last a Lifetime

The sensory epithelium responsible for hearing is the most precisely and minutely engineered of all the tissues in the body (**Figure 23–13**). Its sensory cells, the **auditory hair cells**, are held in a rigid framework of supporting cells and overlaid by a mass of extracellular matrix (the tectorial membrane), in a structure called the *organ of Corti*. The hair cells convert mechanical stimuli into electrical signals. Each has a characteristic organ-pipe array of giant microvilli (called *stereocilia)* protruding from its surface as rigid rods, filled with cross-linked actin filaments, and arranged in ranks of graded height. The dimensions of each such array are specified with extraordinary accuracy according to the location of the hair cell in the ear and the frequency of sound that it has to

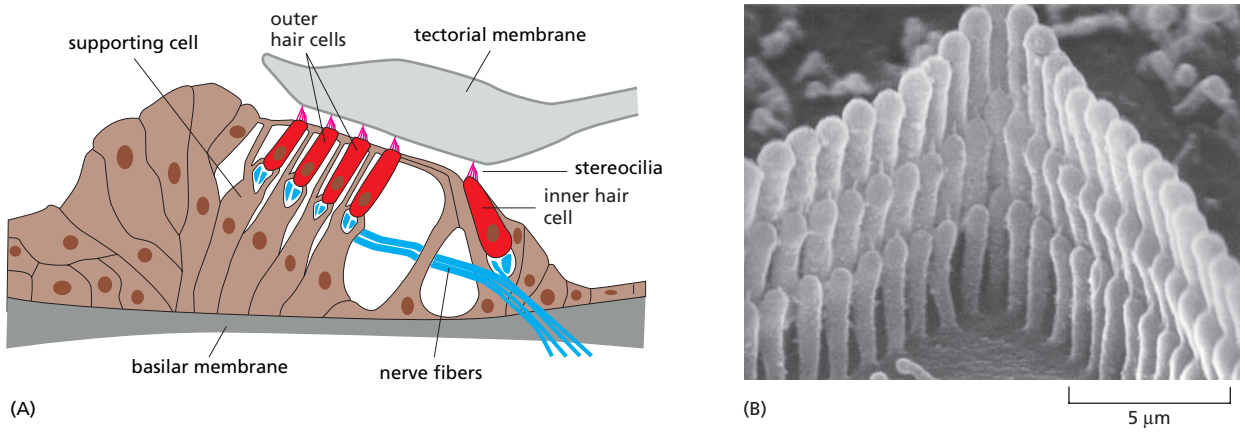


Figure 23–13 Auditory hair cells. (A) A diagrammatic cross section of the auditory apparatus (the organ of Corti) in the inner ear of a mammal shows the auditory hair cells held in an elaborate epithelial structure of supporting cells and overlaid by a mass of extracellular matrix (the tectorial membrane). The epithelium containing the hair cells sits on the basilar membrane—a thin, resilient sheet of tissue that forms a long, narrow partition separating two fluid-filled channels. Sound creates pressure waves in these channels and makes the basilar membrane vibrate up and down. (B) This scanning electron micrograph shows the apical surface of an outer auditory hair cell, with the characteristic organ-pipe array of giant microvilli (stereocilia). The inner hair cells, of which there are just 3500 in each human ear, are the principal auditory receptors. The outer hair cells, roughly four times more numerous in humans, are thought to form part of a feedback mechanism that regulates the mechanical stimulus delivered to the inner hair cells. (B, from J.D. Pickles, *Prog. Neurobiol.* 24:1–42, 1985. With permission from Elsevier.)

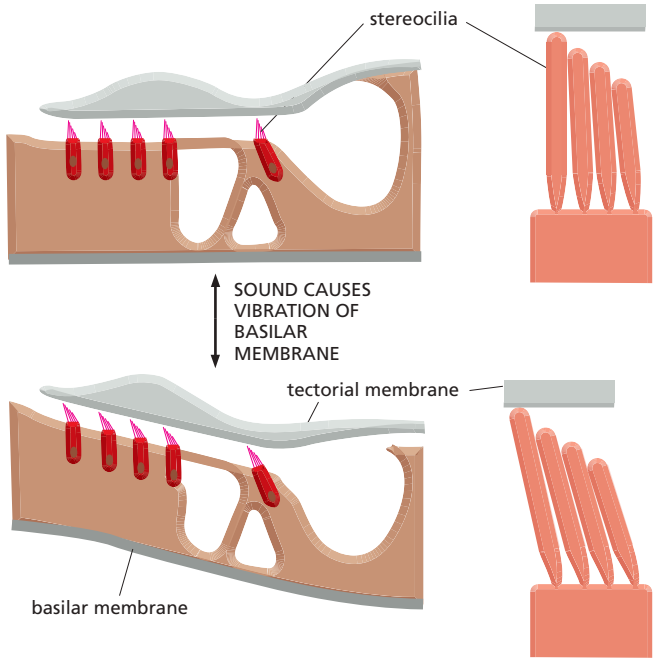


Figure 23–14 How a relative movement of the overlying extracellular matrix (the tectorial membrane) tilts the stereocilia of auditory hair cells in the organ of Corti in the inner ear of a mammal. The stereocilia behave as rigid rods hinged at the base and bundled together at their tips.

respond to. Sound vibrations rock the organ of Corti, causing the bundles of stereocilia to tilt (**Figure 23–14**) and mechanically gated ion channels in the membranes of the stereocilia to open or close (**Figure 23–15**). The flow of electric charge carried into the cell by the ions alters the membrane potential and thereby controls the release of neurotransmitter at the cell's basal end, where the cell synapses with a nerve ending. <TCCA> <CATA>

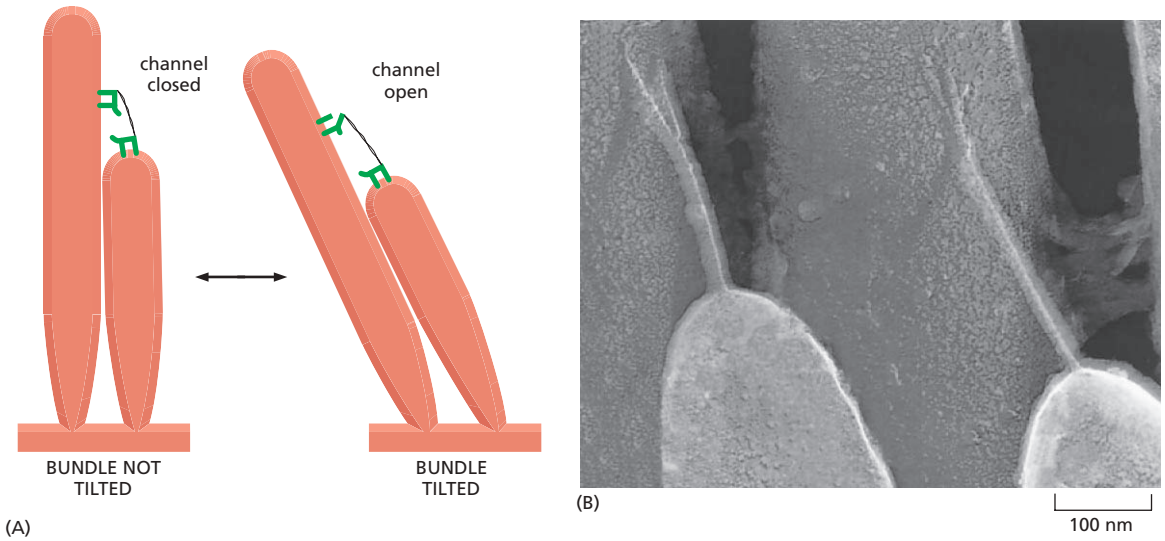


Figure 23–15 How a sensory hair cell works. (A) The cell functions as a transducer, generating an electrical signal in response to sound vibrations that rock the organ of Corti and so cause the stereocilia to tilt. A fine filament runs more or less vertically upward from the tip of each shorter stereocilium to attach at a higher point on its adjacent taller neighbor. Tilting the bundle puts tension on the filaments, which pull on mechanically gated ion channels in the membrane of the stereocilia. Opening of these channels allows an influx of positive charge, depolarizing the hair cell. (B) An electron micrograph of the filaments extending from the tops of two stereocilia. Each filament consists, in part at least, of members of the cadherin superfamily of cell–cell adhesion molecules. Mutant individuals lacking these specific cadherins lack the filaments and are deaf.

By extraordinarily delicate mechanical measurements, correlated with electrical recordings from a single hair cell as the bundle of stereocilia is deflected by pushing with a flexible glass probe, it is possible to detect an extra "give" of the bundle as the mechanically gated channels yield to the applied force and are pulled open. In this way it can be shown that the force required to open a single one of the hypothesized channels is about 2×10^{-13} newtons and that its gate swings through a distance of about 4 nm as it opens. The mechanism is astonishingly sensitive: the faintest sounds that we can hear have been estimated to stretch the filaments by an average of 0.04 nm, which is just under half the diameter of a hydrogen atom. (B, from B. Kachar et al., *Proc. Natl Acad. Sci. U.S.A.* 97:13336–13341, 2000. With permission from National Academy of Sciences.)

In humans and other mammals, the auditory hair cells, unlike olfactory neurons, have to last a lifetime. If they are destroyed by disease, toxins, or excessively loud noise, they do not regenerate and the resultant hearing loss is permanent. But in other vertebrates, destruction of auditory hair cells triggers the supporting cells to divide and behave as stem cells, generating progeny that can differentiate as replacements for the hair cells that are lost. With better understanding of how this regeneration process is regulated, we may one day be able to induce the auditory epithelium to repair itself in humans also.

So far, one treatment is known that can bring about the partial regeneration of auditory hair cells in an adult mammal. The technique uses a virus (an adenovirus) engineered to contain a copy of the *Atoh1* gene, coding for a gene regulatory protein that is known to drive the differentiation of hair cells during development. Guinea pigs that have been deafened by exposure to a toxin that destroys hair cells can be treated by injection of this viral construct into a damaged ear. Many of the surviving supporting cells then become infected with the viral construct and express *Atoh1*. This converts them into functioning hair cells, and the animal partially recovers its hearing in the treated ear.

Most Permanent Cells Renew Their Parts: the Photoreceptor Cells of the Retina

The neural retina is the most complex of the sensory epithelia. It consists of several cell layers organized in a way that seems perverse. The neurons that transmit signals from the eye to the brain (called *retinal ganglion cells*) lie closest to the external world, so that the light, focused by the lens, must pass through them to reach the photoreceptor cells. The **photoreceptors**, which are classified as *rods* or *cones*, according to their shape, lie with their photoreceptive ends, or outer segments, partly buried in the *pigment epithelium* (**Figure 23–16**). Rods and cones contain different *visual pigments*—photosensitive complexes of *opsin* protein with the light-absorbing small molecule *retinal*. Rods, whose visual pigment is called *rhodopsin*, are especially sensitive at low light levels,

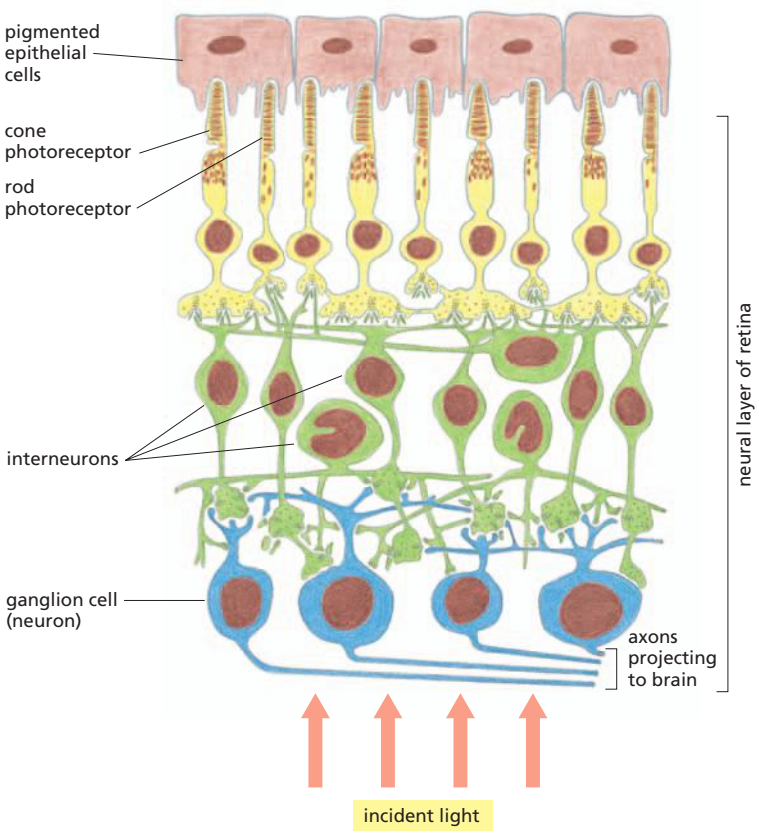


Figure 23–16 The structure of the retina. When light stimulates the photoreceptors, the resulting electrical signal is relayed via interneurons to the ganglion cells, which then convey the signal to the brain. A population of specialized supporting cells (not shown here) occupies the spaces between the neurons and photoreceptors in the neural retina. (Modified from J.E. Dowling and B.B. Boycott, *Proc. R. Soc. Lond. B Biol. Sci.* 166:80–111, 1966. With permission from Royal Society.)

while cones (of which there are three types in humans, each with a different opsin, giving a different spectral response) detect color and fine detail.

The outer segment of a photoreceptor appears to be a modified cilium with a characteristic ciliumlike arrangement of microtubules in the region connecting the outer segment to the rest of the cell (**Figure 23–17**). The remainder of the outer segment is almost entirely filled with a dense stack of membranes in which the photosensitive complexes are embedded; light absorbed here produces an electrical response, as discussed in Chapter 15. At their opposite ends, the photoreceptors form synapses on interneurons, which relay the signal to the retinal ganglion cells (see Figure 23–16).

Photoreceptors in humans, like human auditory hair cells, are permanent cells that do not divide and are not replaced if destroyed by disease or by a misdirected laser beam. The photosensitive molecules of visual pigment, however, are not permanent but are continually degraded and replaced. In rods (although not, curiously, in cones), this turnover is organized in an orderly production line, which can be analyzed by following the passage of a cohort of radiolabeled protein molecules through the cell after a short pulse of radioactive amino acid (**Figure 23–18**). The radiolabeled proteins can be traced from the Golgi apparatus in the inner segment of the cell to the base of the stack of membranes in the outer segment. From here they are gradually displaced toward the tip as new material is fed into the base of the stack. Finally (after about 10 days in the rat), on reaching the tip of the outer segment, the labeled proteins and the layers of membrane in which they are embedded are phagocytosed (chewed off and digested) by the cells of the pigment epithelium.

This example illustrates a general point: even though individual cells of certain types persist, very little of the adult body consists of the same molecules that were laid down in the embryo.

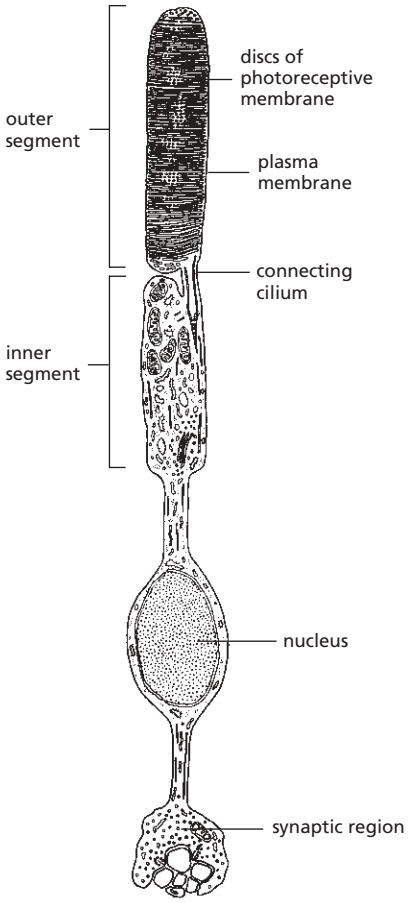


Figure 23–17 A rod photoreceptor.

Summary

Most sensory receptor cells, like epidermal cells and nerve cells, derive from the epithelium forming the outer surface of the embryo. They transduce external stimuli into electrical signals, which they relay to neurons via chemical synapses. Olfactory receptor cells in the nose are themselves full-fledged neurons, sending their axons to the brain. They have a lifetime of only a month or two, and are continually replaced by new cells derived from stem cells in the olfactory epithelium. Each olfactory neuron expresses just

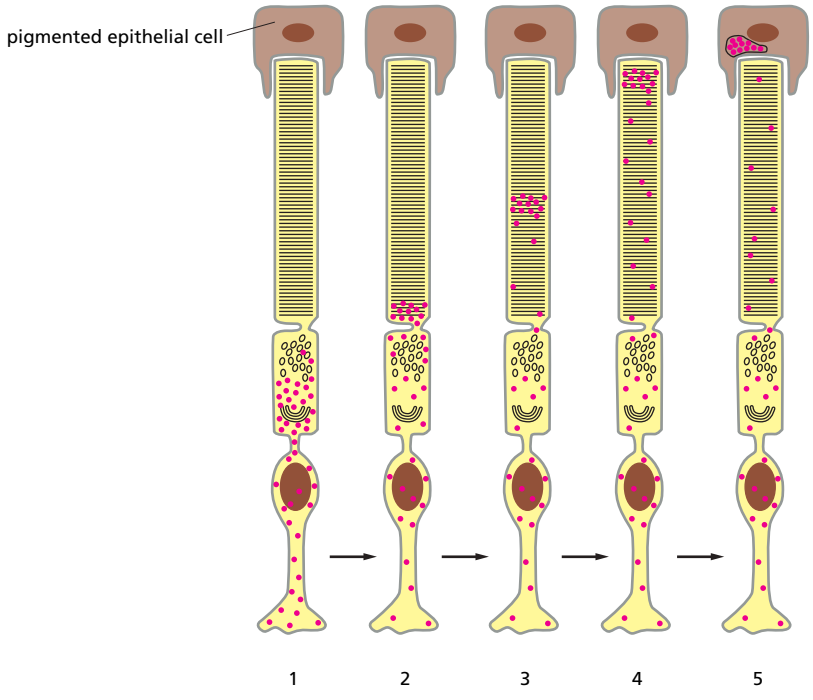


Figure 23–18 Turnover of membrane protein in a rod cell. Following a pulse of ^{3}H-leucine, the passage of radiolabeled proteins through the cell is followed by autoradiography. *Red dots* indicate sites of radioactivity. The method reveals only the ^{3}H-leucine that has been incorporated into proteins; the rest is washed out during tissue preparation. (1) The incorporated leucine is first seen concentrated in the neighborhood of the Golgi apparatus. (2) From there it passes to the base of the outer segment into a newly synthesized disc of photoreceptive membrane. (3–5) New discs are formed at a rate of three or four per hour (in a mammal), displacing the older discs toward the pigment epithelium.

one of the hundreds of different olfactory receptor proteins for which genes exist in the genome, and the axons from all olfactory neurons expressing the same receptor protein navigate to the same glomeruli in the olfactory bulbs of the brain.

Auditory hair cells—the receptor cells for sound—unlike olfactory receptor cells, have to last a lifetime, in mammals at least, although artificial expression of a hair-cell differentiation gene, Atoh1, can convert surviving supporting cells into hair cells where hair cells have been destroyed. Hair cells have no axon but make synaptic contact with nerve terminals in the auditory epithelium. They take their name from the hair-like bundle of stereocilia (giant microvilli) on their outer surface. Sound vibrations tilt the bundle, pulling mechanically gated ion channels on the stereocilia into an open configuration to excite the cell electrically.

Photoreceptor cells in the retina of the eye absorb photons in molecules of visual pigment (opsin protein plus retinal) held in stacks of membrane in the photoreceptor outer segments, triggering an electrical excitation by a more indirect intracellular signaling pathway. Although the photoreceptor cells themselves are permanent and irreplaceable, the stacks of opsin-rich membrane that they contain undergo continual renewal.

THE AIRWAYS AND THE GUT

The examples we have discussed so far represent a small selection of the tissues and cell types that derive from the outer layer of the embryo—the ectoderm. They are enough, however, to illustrate how diverse these cells can be, in form, function, lifestyle, and pattern of replacement. The innermost layer of the embryo—the *endoderm*, forming the primitive gut tube—gives rise to another whole zoo of cell types lining the digestive tract and its appendages. We begin with the lungs.

Adjacent Cell Types Collaborate in the Alveoli of the Lungs

The airways of the lungs are formed by repeated branching of a system of tubes that originated in the embryo from an outpocketing of the gut lining, as discussed in Chapter 22 (see Figure 22–92). Repeated tiers of branching terminate in several hundred million air-filled sacs—the **alveoli**. Alveoli have thin walls, closely apposed to the walls of blood capillaries so as to allow exchange of O_2 and CO_2 with the bloodstream (**Figure 23–19**).

To survive, the cells lining the alveoli must remain moist. At the same time, they must serve as a gas container that can expand and contract with each breath in and out. This creates a problem. When two wet surfaces touch, they become stuck together by surface tension in the layer of water between them—an effect that operates more powerfully the smaller the scale of the structure. There is a risk, therefore, that the alveoli may collapse and be impossible to reinflate. To solve the problem, two types of cells are present in the lining of the alveoli. *Type I alveolar cells* cover most of the wall: they are thin and flat (*squamous*) to allow gas exchange. *Type II alveolar cells* are interspersed among them. These are plump and secrete *surfactant,* a phospholipid-rich material that forms a film on the free water surfaces and reduces surface tension, making the alveoli easy to reinflate even if they collapse. The production of adequate amounts of surfactant in the fetus, starting at about 5 months of pregnancy in humans, marks the beginning of the possibility of independent life. Premature babies born before this stage are unable to inflate their lungs and breathe; those born after it can do so and, with intensive care, can survive.

Goblet Cells, Ciliated Cells, and Macrophages Collaborate to Keep the Airways Clean

Higher up in the airways we find different combinations of cell types, serving different purposes. The air we breathe is full of dust, dirt, and air-borne

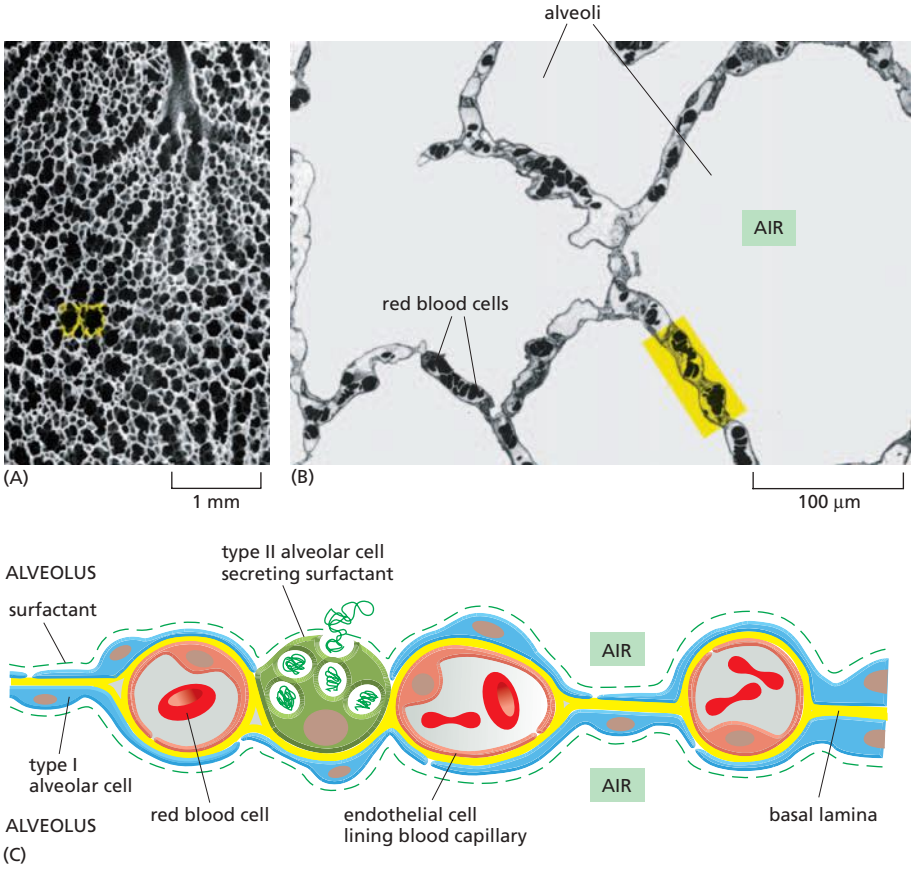


Figure 23–19 Alveoli in the lung. (A) Scanning electron micrograph at low magnification, showing the sponge-like texture created by the many air-filled alveoli. A bronchiole (small tubular airway) is seen at the top, communicating with the alveoli. (B) Transmission electron micrograph of a section through a region corresponding to the *yellow* box in (A) showing the alveolar walls, where gas exchange occurs. (C) Diagram of the cellular architecture of a piece of alveolar wall, corresponding to the *yellow* box in (B). (A, from P. Gehr et al., *Respir. Physiol.* 44:61–86, 1981. With permission from Elsevier; B, courtesy of Peter Gehr, from D.W. Fawcett, A Textbook of Histology, 12th ed. New York: Chapman and Hall, 1994.)

microorganisms. To keep the lungs clear and healthy, this debris must be constantly swept out. To perform this task, a relatively thick *respiratory epithelium* lines the larger airways (**Figure 23–20**). This epithelium consists of three differentiated cell types: *goblet cells* (so named because of their shape), which secrete mucus; *ciliated cells*, with cilia that beat; and a small number of *endocrine cells*, secreting serotonin and peptides that act as local mediators. These signal molecules affect nerve endings and other neighboring cells in the respiratory tract, so as to help regulate the rate of mucus secretion and ciliary beating, the contraction of surrounding smooth muscle cells that can constrict the airways, and other functions. Basal cells are also present, and serve as stem cells for renewal of the epithelium.

The mucus secreted by the goblet cells forms a viscoelastic blanket about 5 μm thick over the tops of the cilia. The cilia, all beating in the same direction, at a rate of about 12 beats per second, sweep the mucus out of the lungs, carrying with it the debris that has become stuck to it. This conveyor belt for the removal of rubbish from the lungs is called the *mucociliary escalator*. Of course, some inhaled particles may reach the alveoli themselves, where there is no escalator. Here, the unwanted matter is removed by yet another class of specialized cells, *macrophages*, which roam the lungs, engulfing foreign matter and killing and digesting bacteria. Many millions of macrophages, loaded with debris, are swept out of the lungs every hour on the mucociliary escalator.

At the upper end of the respiratory tract, the wet mucus-covered respiratory epithelium gives way abruptly to stratified squamous epithelium. This cell sheet is structured for mechanical strength and protection, and, like epidermis, it consists of many layers of flattened cells densely packed with keratin. It differs from epidermis in that it is kept moist and its cells retain their nucleus even in the outermost layers. Abrupt boundaries of epithelial cell specialization, such as that between the mucous and the stratified squamous epithelium of the respiratory tract, are also found in other parts of the body, but very little is known about how they are created and maintained.

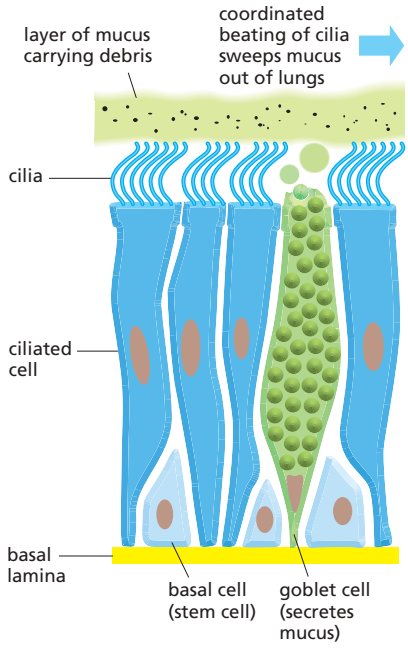


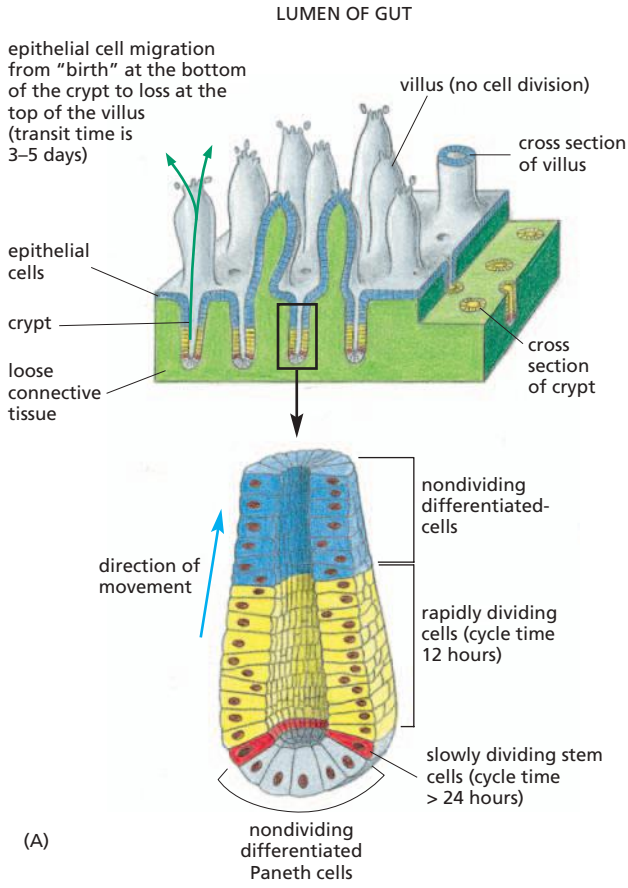
Figure 23–20 Respiratory epithelium. The goblet cells secrete mucus, which forms a blanket over the tops of the ciliated cells. The regular, coordinated beating of the cilia sweeps the mucus up and out of the airways, carrying any debris that is stuck to it. The mechanism that coordinates the ciliary beating is a mystery, but it seems to reflect an intrinsic polarity in the epithelium. If a segment of rabbit trachea is surgically reversed, it carries on sweeping mucus, but in the wrong direction, back down toward the lung, in opposition to adjacent unreversed portions of trachea.

The Lining of the Small Intestine Renews Itself Faster Than Any Other Tissue

Only air-breathing vertebrates have lungs, but all vertebrates, and almost all invertebrate animals, have a gut—that is, a digestive tract lined with cells specialized for the digestion of food and absorption of the nutrient molecules released by the digestion. These two activities are hard to carry on at the same time, as the processes that digest food in the lumen of the gut are liable also to digest the lining of the gut itself, including the cells that absorb the nutrients. The gut uses several strategies to solve the problem.

The fiercest digestive processes, involving acid hydrolysis as well as enzyme action, are conducted in a separate reaction vessel, the stomach. The products are then passed on to the small intestine, where the nutrients are absorbed and enzymatic digestion continues, but at a neutral pH. The different regions of the gut lining consist of correspondingly different mixtures of cell types. The stomach epithelium includes cells that secrete acid, and other cells that secrete digestive enzymes that work at acid pH. Conversely, glands (in particular the pancreas) that discharge into the initial segment of the small intestine contain cells that secrete bicarbonate to neutralize the acidity, along with other cells that secrete digestive enzymes that work at neutral pH. The lining of the intestine, downstream from the stomach, contains both absorptive cells and cells specialized for the secretion of mucus, which covers the epithelium with a protective coat. In the stomach, too, mucus cells line most exposed surfaces. And, in case these measures are not enough, the whole lining of the stomach and intestine is continually renewed and replaced by freshly generated cells, with a turnover time of a week or less.

The renewal process has been studied best in the small intestine (**Figure 23–21**). The lining of the small intestine (and of most other regions of the gut) is a single-layered epithelium. This epithelium covers the surfaces of the *villi* that project into the lumen, and it lines the *crypts* that descend into the underlying



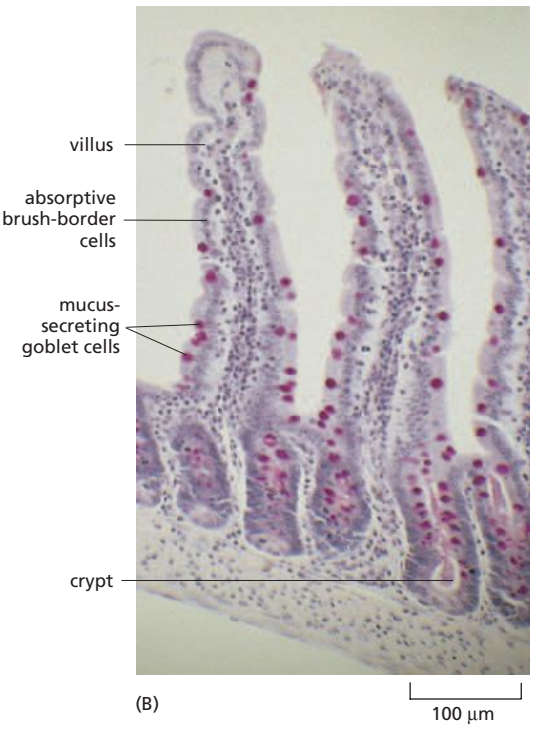


Figure 23–21 Renewal of the gut lining. (A) The pattern of cell turnover and the proliferation of stem cells in the epithelium that forms the lining of the small intestine. The *colored arrow* shows the general upward direction of cell movement onto the villi, but some cells, including a proportion of the goblet and enteroendocrine cells, stay behind and differentiate while still in the crypts. The nondividing differentiated cells (Paneth cells) at the very bottom of the crypts also have a finite lifetime and are continually replaced by progeny of the stem cells. (B) Photograph of a section of part of the lining of the small intestine, showing the villi and crypts. Note how mucus-secreting goblet cells (stained *red*) are interspersed among other cell types. Enteroendocrine cells are less numerous and less easy to identify without special stains. See Figure 23–22 for the structure of these cells.

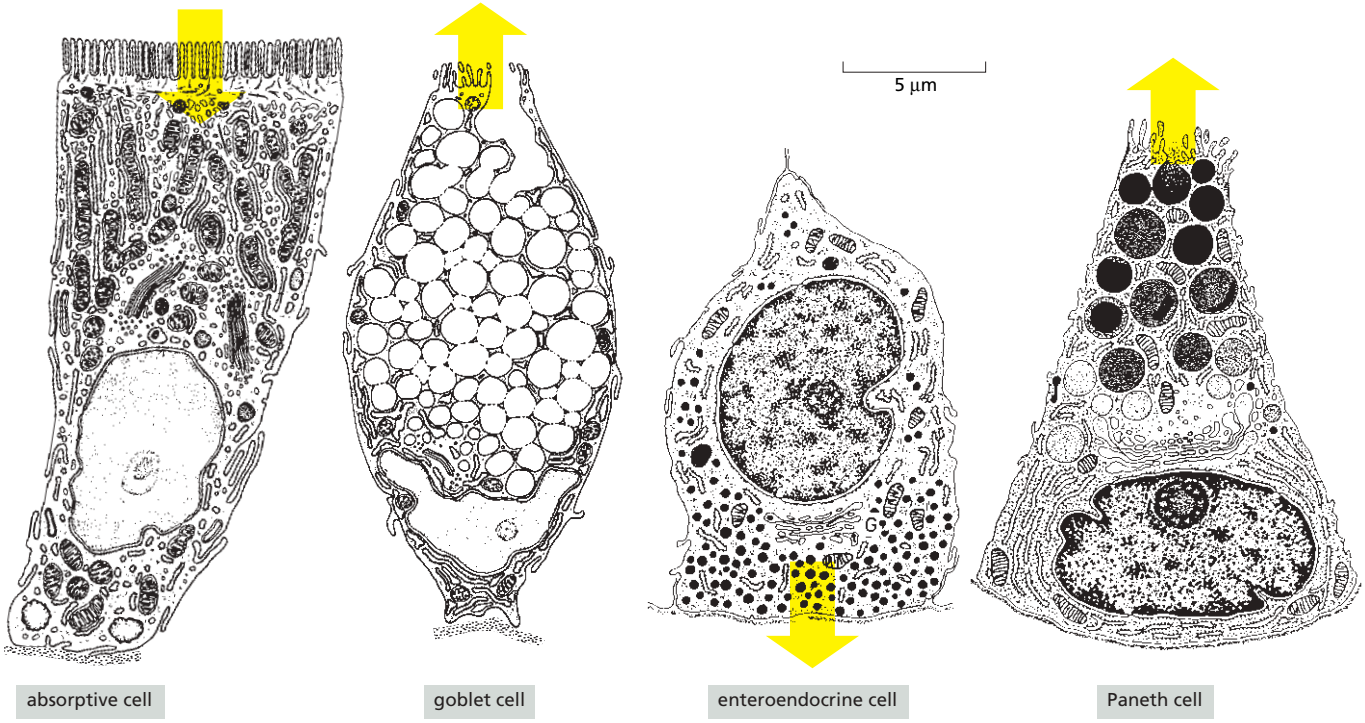


Figure 23–22 The four main differentiated cell types found in the epithelial lining of the small intestine. All of these are generated from undifferentiated multipotent stem cells living near the bottoms of the crypts (see Figure 23–21). The microvilli on the apical surface of the absorptive (brush-border) cell provide a 30-fold increase of surface area, not only for the import of nutrients but also for the anchorage of enzymes that perform the final stages of extracellular digestion, breaking down small peptides and disaccharides into monomers that can be transported across the cell membrane. Broad *yellow* arrows indicate direction of secretion or uptake of materials for each type of cell. (After T.L. Lentz, Cell Fine Structure. Philadelphia: Saunders, 1971; R. Krstić, Illustrated Encyclopedia of Human Histology. Berlin: Springer-Verlag, 1984.)

connective tissue. Dividing stem cells lie in a protected position in the depths of the crypts. These generate four types of differentiated progeny (**Figure 23–22**):

1. *Absorptive cells* (also called *brush-border cells* or *enterocytes*) have densely packed microvilli on their exposed surfaces to increase their active surface area for the uptake of nutrients. They both absorb nutrients and secrete (or carry on their exterior surfaces) hydrolytic enzymes that perform some of the final steps of extracellular digestion, breaking down food molecules in preparation for transport across the plasma membrane.
2. *Goblet cells* (as in respiratory epithelium) secrete mucus.
3. *Paneth cells* form part of the innate immune defense system (discussed in Chapter 24) and secrete (along with some growth factors) *cryptdins*—proteins of the defensin family that kill bacteria (see Figure 24–46).
4. *Enteroendocrine cells*, of more than 15 different subtypes, secrete serotonin and peptide hormones, such as cholecystokinin (CCK), that act on neurons and other cell types in the gut wall and regulate the growth, proliferation, and digestive activities of cells of the gut and other tissues. Cholecystokinin, for example, is released from enteroendocrine cells in response to the presence of nutrients in the gut and binds to receptors on nearby sensory nerve endings, which relay a signal to the brain to stop you feeling hungry after you have eaten enough.

The absorptive, goblet, and enteroendocrine cells travel mainly upward from the stem-cell region, by a sliding movement in the plane of the epithelial sheet, to cover the surfaces of the villi. In analogy with the epidermis, it is thought that the most rapidly proliferating precursor cells in the crypt are in a transit amplifying stage, already committed to differentiation but undergoing several divisions on their way out of the crypt, before they stop dividing and differentiate terminally. Within 2–5 days (in the mouse) after emerging from the crypts, the cells reach the tips of the villi, where they undergo the initial stages of

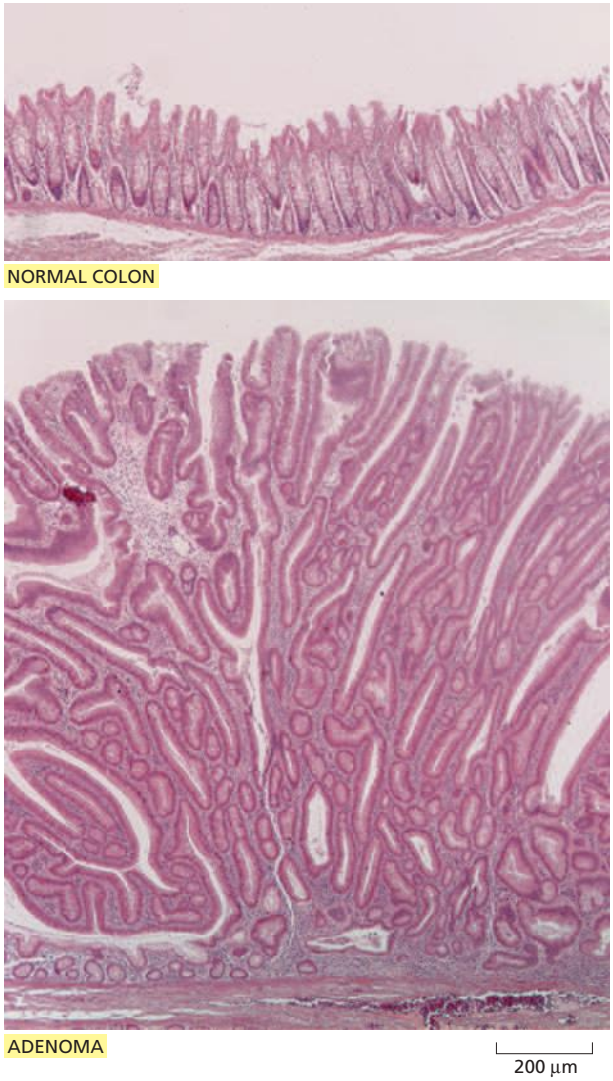


Figure 23–23 An adenoma in the human colon, compared with normal tissue from an adjacent region of the same person's colon. The specimen is from a patient with an inherited mutation in one of his two copies of the *Apc* gene. A mutation in the other *Apc* gene copy, occurring in a colon epithelial cell during adult life, has given rise to a clone of cells that behave as though the Wnt signaling pathway is permanently activated. As a result, the cells of this clone form an adenoma—an enormous, steadily expanding mass of giant cryptlike structures.

apoptosis and are finally discarded into the gut lumen. The Paneth cells are produced in much smaller numbers and have a different migration pattern. They live at the bottom of the crypts, where they too are continually replaced, although not so rapidly, persisting for about 20 days (in the mouse) before undergoing apoptosis and being phagocytosed by their neighbors. The stem cells, too, remain at or near the bottoms of the crypts. What keeps them there, and what restricts cell division to the crypts? How are the migrations controlled so that some cells move up while others stay down? What are the molecular signals that organize the whole stem-cell system, and how do they work?

Wnt Signaling Maintains the Gut Stem-Cell Compartment

The beginnings of an answer to these questions came from the study of cancer of the colon and rectum (the lower end of the gut). As discussed in Chapter 20, some people have a hereditary predisposition to this disease and, in advance of the cancer, develop large numbers of small precancerous tumors (adenomas) in the lining of their large intestine (**Figure 23–23**). The appearance of these tumors suggests that they have arisen from intestinal crypt cells that have failed to halt their proliferation in the normal way, and so have given rise to excessively large cryptlike structures. The cause can be traced to mutations in the *Apc* (*Adenomatous Polyposis Coli*) gene: the tumors arise from cells that have lost both gene copies. *Apc* codes for a protein that prevents inappropriate activation of the Wnt signaling pathway, so that loss of APC is presumed to mimic the effect of continual exposure to a Wnt signal. The suggestion, therefore, is that Wnt signaling normally keeps

crypt cells in a proliferative state, and cessation of exposure to Wnt signaling normally makes them stop dividing as they leave the crypt. Indeed, mice that are homozygous for a knockout mutation in the *Tcf4* gene, coding for a gene regulatory protein that is required as an effector of Wnt signaling in the gut, make no crypts, fail to renew their gut epithelium, and die soon after birth.

Experiments with transgenic mice confirm the importance of Wnt signaling and reveal other regulators that act together with Wnt to organize the gut-cell production line and keep it running correctly. Using the Cre/lox technique with an inducible promoter for Cre (as described in Chapter 8, p. 567), it is possible, for example, to knock out the *Apc* gene in gut epithelial cells abruptly, at any chosen time in the life of the mouse. Within a few days, the gut structure is transformed: the crypt-like regions of proliferative cells are greatly enlarged, villi are reduced, and the numbers of terminally differentiated cells are drastically diminished. Conversely, one can make a transgenic mouse in which the gut epithelial cells all secrete a diffusible inhibitor of Wnt signaling. These animals, in which Wnt signaling is blocked, form scarcely any crypts and have hardly any proliferating cells in their gut epithelium. Instead, almost all the gut lining cells are fully differentiated non-dividing absorptive cells; but goblet cells, enteroendocrine cells, and Paneth cells are missing. Thus Wnt signaling not only keeps cells in a proliferative state but is also needed to make them competent to give rise to the full range of ultimate differentiated cell types.

Notch Signaling Controls Gut Cell Diversification

What then causes the cells to diversify as they differentiate? Notch signaling has this function in many other systems, where it mediates lateral inhibition—a competitive interaction that drives neighboring cells toward different fates (see Chapters 15 and 22, Figures 15–75 and 22–60). All the essential components of the Notch pathway are expressed in the crypts; it seems that Wnt signaling switches on their expression. When Notch signaling is abruptly blocked by knocking out one of these essential components, within a few days all the cells in the crypts differentiate as goblet cells and cease dividing; conversely, when Notch signaling is artificially activated in all the cells, no goblet cells are produced and the crypt-like regions of cell proliferation are enlarged.

From the effects of all these manipulations of Wnt and Notch signaling, we arrive at a simple picture of how the two pathways combine to govern the production of differentiated cells from the intestinal stem cells (**Figure 23–24**). Wnt

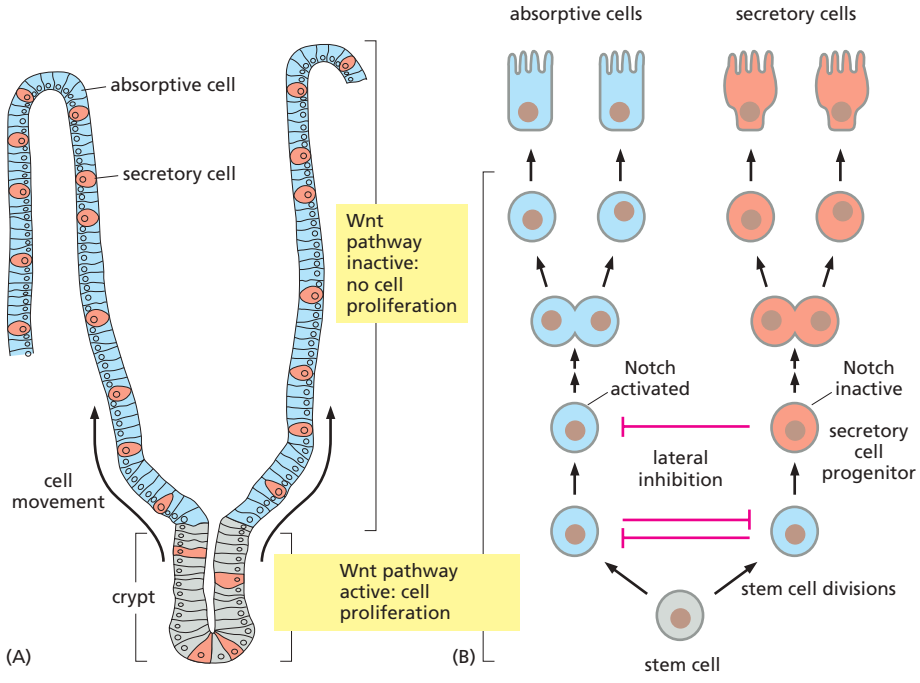


Figure 23–24 How Wnt and Notch signaling pathways combine to control the production of differentiated cells from stem cells in the intestine. (A) Wnt signaling maintains proliferation in the crypt, where the stem cells reside and their progeny become committed to diverse fates. (B) Wnt signaling in the crypt drives expression of the components of the Notch signaling pathway in that region; Notch signaling is thus active in the crypt and, through lateral inhibition, forces cells there to diversify. Both pathways must be activated in the same cell to keep it as a stem cell. The progeny of the stem cell continue dividing under the influence of Wnt even after they become committed to a specific differentiated fate, but the timing of these transit amplifying divisions in relation to commitment is not understood in detail.

signaling promotes cell proliferation and confers competence for the full range of modes of differentiation, while preventing differentiation from occurring immediately; in this way, it defines the crypt and maintains the stem cells. But Wnt signaling also, at the same time, activates the expression of Notch pathway components, and Notch signaling within the crypt population mediates lateral inhibition, which forces the cells to diversify, in such a way that some become singled out to deliver lateral inhibition, while others receive it. Cells of the former class express Notch ligands and activate Notch in their neighbors, but escape from Notch activation themselves; as a result, they become committed to differentiate as secretory cells. Cells of the latter class—the majority—are kept in an opposite state, with Notch activated and ligand expression inhibited; as a result, they retain competence to differentiate in any of a variety of ways and to engage in lateral-inhibition competition with their neighbors. Both classes of cells (with the exception of some secretory subtypes) continue dividing so long as they are in the crypt, under the influence of Wnt. But when cells leave the crypt and lose exposure to Wnt signaling, the competition halts, division stops, and the cells differentiate according to their individual states of Notch activation at that time—as absorptive cells if Notch is still activated, as secretory cells if it is not.

This is certainly not the whole story of events in the crypt. It does not explain, for example, how the various subclasses of secretory cells (goblet, enteroendocrine, and Paneth) become different from one another. Nor does it say anything about the distinction that many experts believe to exist between true stem cells and the more rapidly dividing transit amplifying cells within the crypt. Several different members of each of the families of Wnt and Notch pathway components are expressed in the crypt epithelium and in the connective tissue around the base of the crypts, and probably have differing effects. Moreover, other signaling pathways also have crucial functions in organizing the system.

Ephrin–Eph Signaling Controls the Migrations of Gut Epithelial Cells

One of the most remarkable features of the gut stem-cell system is the steady, ordered, selective migration of cells from crypt to villus. Differentiating absorptive, goblet, and enteroendocrine cells stream out of the crypts and up the villi (**Figure 23–25**); stem cells remain deep in the crypts; and Paneth cells migrate right down to the crypt bottoms. This pattern of movements, which segregates the different groups of cells, depends on yet another cell–cell signaling pathway. Wnt signaling stimulates the expression of cell-surface receptors of the EphB family (discussed in Chapter 15) in the cells in the crypt; however, as cells differentiate, they switch off expression of these receptors, and switch on instead expression of the ligands, cell-surface proteins of the ephrinB family (**Figure 23–26**A). There is one exception: the Paneth cells retain expression of the EphB

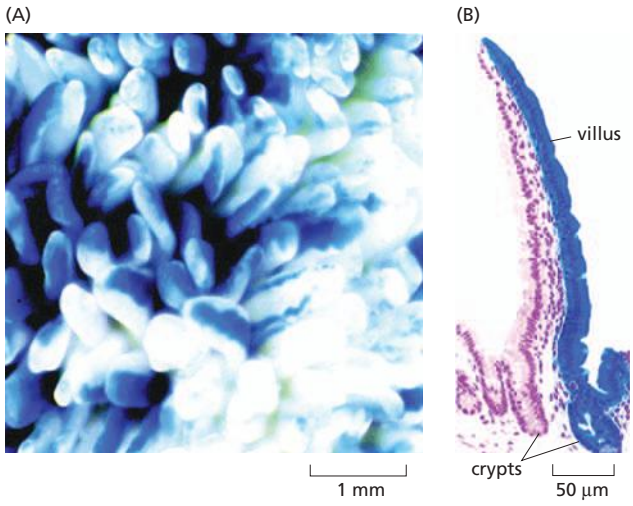


Figure 23–25 Migration of cells from crypts onto villi. In this mouse intestine, a random subset of epithelial cells was induced to undergo a mutation during late fetal life, causing the mutant cells to express a *LacZ* transgene, coding for an enzyme that can be detected by the blue product of the reaction that it catalyses. By 6 weeks after birth, each crypt has become populated by the progeny of a single stem cell and thus appears either totally blue or totally white, according to whether that stem cell was or was not genetically marked in this way. Several crypts contribute to a single villus, each sending a stream of differentiated cells outward along it. (A) Low-magnification surface view of part of the lining of the intestine, showing many villi, each receiving streams of cells from several crypts. (B) Detail of a single villus and adjacent crypts in cross-section. In the example shown, the streams from different crypts have remained unmixed, so that the villus appears blue on one side and white on the other; more commonly, there is some mixing, giving a less orderly result. (From M.H. Wong, J.R. Saam, T.S. Stappenbeck, C.H. Rexer and J.I. Gordon, *Proc. Natl Acad. Sci. U.S.A.* 97:12601–12606, 2000. With permission from National Academy of Sciences.)

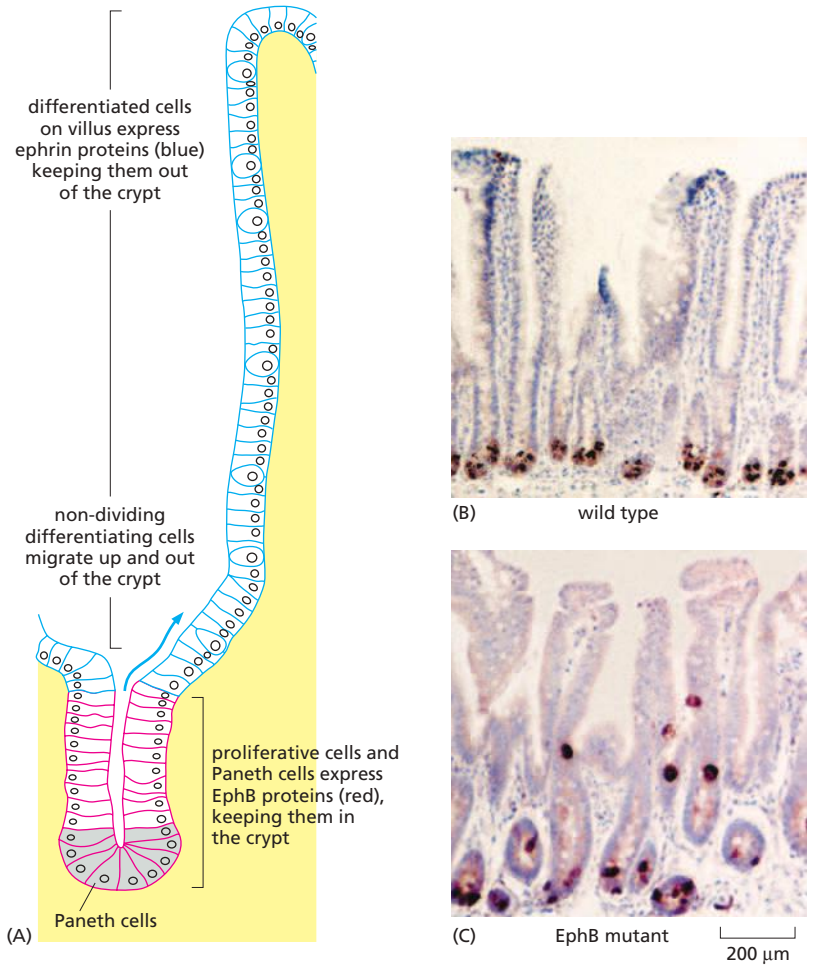
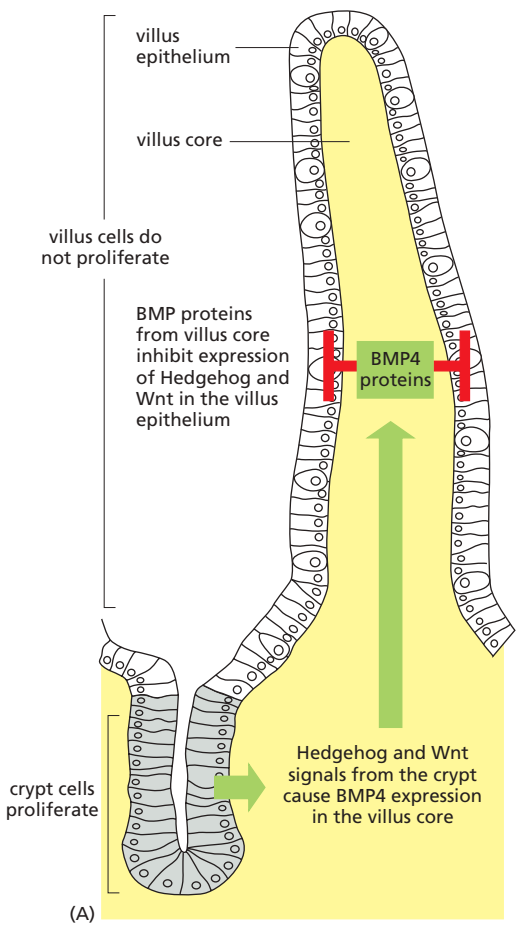


Figure 23–26 Ephrin–Eph signaling controls cell segregation between crypts and villi. (A) Proliferative cells (including the stem cells) and Paneth cells express EphB proteins, while the differentiated, nondividing cells that cover the villi express ephrinB proteins. The repulsive cell–cell interaction mediated by encounters between these two types of cell-surface molecules keeps the two classes of cells segregated. (B) In a normal gut, as a result, Paneth cells (*brown* stain) and dividing cells remain confined to the bottoms of the crypts. (C) In a mutant where EphB proteins are defective, cells that should stay in the crypts wander out onto the villi. (Adapted from E. Batlle et al., *Cell* 111:251–263, 2002. With permission from Elsevier.)

proteins. Thus EphB expression is characteristic of cells that stay in the crypts, while ephrinB expression is characteristic of cells moving out onto the villi. In various other tissues, cells expressing Eph proteins are repelled by contacts with cells expressing ephrins (see Chapter 22, Figure 22–106). It seems that the same is true in the gut lining, and that this mechanism serves to keep the cells in their proper places. In EphB knockout mutants, the populations become mixed, so that, for example, Paneth cells wander out onto the villi (Figure 23–26C). Loss of EphB genes in intestinal cancers correlates with the onset of invasive behavior by the tumor cells.

Wnt, Hedgehog, PDGF, and BMP Signaling Pathways Combine to Delimit the Stem-Cell Niche

Clearly, the gut stem cells cannot exist without the special environment that the crypt provides for them. This **stem cell niche** is as essential as the stem cells themselves. How is it created and maintained? The mechanism seems to depend on a complex interplay of signals between the epithelium and the underlying connective tissue. Exchange of Wnt, Hedgehog, and PDGF signals between the two tissues, and between different regions of the crypt–villus axis, leads to a restriction of Wnt signaling to the neighborhood of the crypts. The epithelial cells in the crypts produce both Wnt proteins and the receptors that respond to them, creating a positive feedback loop that presumably helps to make Wnt pathway activation in this region self-sustaining. At the same time, signals exchanged with the connective tissue lead to expression of BMP proteins in the connective-tissue cells forming the core of the villi (**Figure 23–27**). These cells signal to the adjacent villus epithelium to inhibit the development of misplaced crypts: blocking BMP signaling disrupts the whole organization



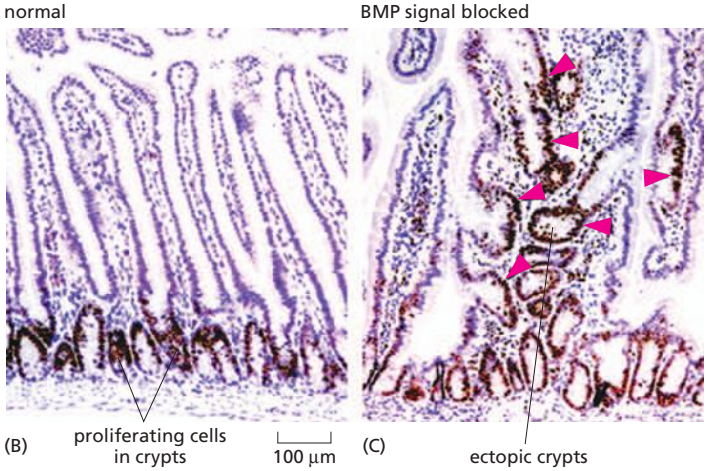


Figure 23–27 Signals defining the intestinal stem-cell niche. (A) Diagram of the signaling system. Signal proteins of the Hedgehog and Wnt families are expressed by the epithelial cells in the base of each crypt, which also express Wnt receptors and experience high levels of Wnt pathway activation. The connective-tissue cells underlying the epithelium express both Hedgehog receptors and Wnt receptors. The combined effect of the signals from the crypt base, perhaps in conjunction with other signals, is to provoke the connective-tissue cells that lie in the core of each villus to express BMP proteins. The BMP proteins act on the epithelium of the villus, preventing its cells from forming crypts. (B) Cross section of a region of normal intestinal epithelium. The brown stain marks proliferative cells, which are confined to the crypts. (C) Similarly stained section of intestine of a transgenic mouse expressing an inhibitor of BMP signaling. Crypts containing dividing cells have developed ectopically, along the sides of the misshapen villi. (B and C, courtesy of A. Haramis et al., *Science* 303:1684–1686, 2004. With permission from AAAS.)

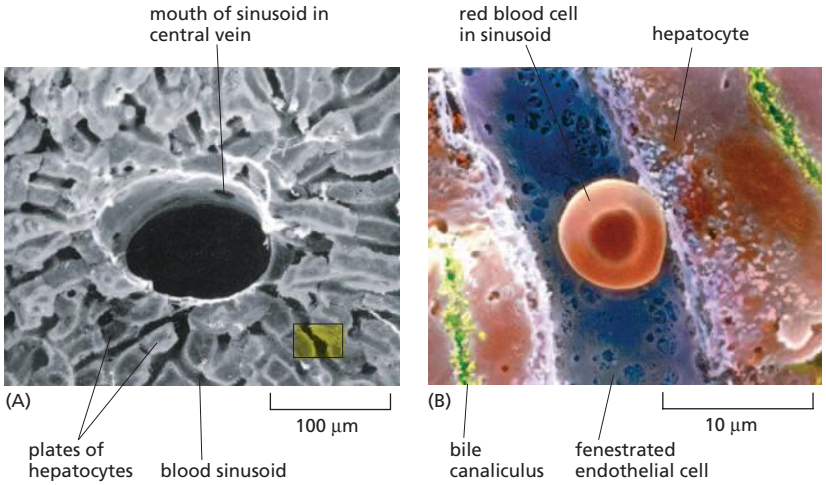
and causes misplaced crypts to form as invaginations of proliferating epithelium along the sides of the villi.

The Liver Functions as an Interface Between the Digestive Tract and the Blood

As we have seen, the functions of the gut are divided between a variety of cell types. Some cells are specialized for the secretion of hydrochloric acid, others for the secretion of enzymes, others for the absorption of nutrients, and so on. Some of these cell types are closely intermingled in the wall of the gut, whereas others are segregated in large glands that communicate with the gut and originate in the embryo as outgrowths of the gut epithelium.

The liver is the largest of these glands. It develops at a site where a major vein runs close to the wall of the primitive gut tube, and the adult organ retains a special relationship with the blood. Cells in the liver that derive from the primitive gut epithelium—the **hepatocytes**—are arranged in interconnected sheets and cords, with blood-filled spaces called sinusoids running between them (**Figure 23–28**). The blood is separated from the surface of the hepatocytes by a single layer of flattened endothelial cells that covers the exposed faces of the hepatocytes. This structure facilitates the chief functions of the liver, which depend on the exchange of metabolites between hepatocytes and the blood.

The liver is the main site at which nutrients that have been absorbed from the gut and then transferred to the blood are processed for use by other cells of the body. It receives a major part of its blood supply directly from the intestinal tract (via the portal vein). Hepatocytes synthesize, degrade, and store a vast number of substances. They play a central part in the carbohydrate and lipid metabolism of the body as a whole, and they secrete most of the protein found in blood plasma. At the same time, the hepatocytes remain connected with the lumen of the gut via a system of minute channels (or canaliculi) and larger ducts



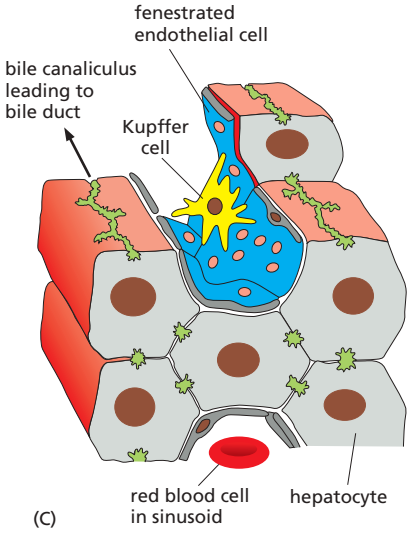


Figure 23–28 The structure of the liver. (A) A scanning electron micrograph of a portion of the liver, showing the irregular sheets and cords of hepatocytes and the many small channels, or sinusoids, for the flow of blood. The larger channels are vessels that distribute and collect the blood that flows through the sinusoids. (B) Detail of a sinusoid (enlargement of region similar to that marked by yellow rectangle at lower right in [A]). (C) Schematized diagram of the fine structure of the liver. A single thin sheet of endothelial cells with interspersed macrophagelike Kupffer cells separates the hepatocytes from the bloodstream. Small holes in the endothelial sheet, called fenestrae (Latin for "windows"), allow the exchange of molecules and small particles between the hepatocytes and the bloodstream. Besides exchanging materials with the blood, the hepatocytes form a system of tiny bile canaliculi into which they secrete bile, which is ultimately discharged into the gut via bile ducts. The real structure is less regular than this diagram suggests. (A and B, courtesy of Pietro M. Motta, University of Rome "La Sapienza.")

(see Figure 23–28B,C) and secrete into the gut by this route both waste products of their metabolism and an emulsifying agent, *bile*, which helps in the absorption of fats. Hepatocytes are big cells, and about 50% of them (in an adult human) are polyploid, with two, four, eight, or even more times the normal diploid quantity of DNA per cell.

In contrast to the rest of the digestive tract, there seems to be remarkably little division of labor within the population of hepatocytes. Each hepatocyte seems able to perform the same broad range of metabolic and secretory tasks. These fully differentiated cells can also divide repeatedly, when the need arises, as we explain next.

Liver Cell Loss Stimulates Liver Cell Proliferation

The liver illustrates in a striking way one of the great unsolved problems of developmental and tissue biology: what determines the size of an organ of the body, or the quantity of one type of tissue relative to another? For different organs, the answers are almost certainly different, but there is scarcely any case in which the mechanism is well understood.

Hepatocytes normally live for a year or more and are renewed at a slow rate. Even in a slowly renewing tissue, however, a small but persistent imbalance between the rate of cell production and the rate of cell death would lead to disaster. If 2% of the hepatocytes in a human divided each week but only 1% died, the liver would grow to exceed the weight of the rest of the body within 8 years. Homeostatic mechanisms must operate to adjust the rate of cell proliferation or the rate of cell death, or both, so as to keep the organ at its normal size. This size, moreover, needs to be matched to the size of the rest of the body. Indeed, when the liver of a small dog is grafted into a large dog, it rapidly grows to almost the size appropriate to the host; conversely, when the liver is grafted from a large dog into a small one, it shrinks.

Direct evidence for the homeostatic control of liver cell proliferation comes from experiments in which large numbers of hepatocytes are removed surgically or are intentionally killed by poisoning with carbon tetrachloride. Within a day or so after either sort of damage, a surge of cell division occurs among the surviving hepatocytes, quickly replacing the lost tissue. (If the hepatocytes themselves are totally eliminated, another class of cells, located in the bile ducts, can serve as stem cells for the genesis of new hepatocytes, but usually there is no need for this.) If two-thirds of a rat's liver is removed, for example, a liver of nearly normal size can regenerate from the remainder by hepatocyte proliferation within about 2 weeks. Although many molecules have been implicated in the triggering of this reaction, one of the most important is a protein called *hepatocyte growth factor*. It stimulates hepatocytes to divide in culture, and its production increases steeply (by poorly understood mechanisms) in response to liver damage.

The balance between cell births and cell deaths in the adult liver (and other organs too) does not depend exclusively on the regulation of cell proliferation: cell survival controls also play a part. If an adult rat is treated with the drug phenobarbital, for example, hepatocytes are stimulated to divide, causing the liver to enlarge. When the phenobarbital treatment is stopped, hepatocyte cell death greatly increases until the liver returns to its original size, usually within a week or so. The mechanism of this type of cell survival control is unknown, but it has been suggested that hepatocytes, like most vertebrate cells, depend on signals from other cells for their survival and that the normal level of these signals can support only a certain standard number of hepatocytes. If the number of hepatocytes rises above this (as a result of phenobarbital treatment, for example), hepatocyte death will automatically increase to bring their number back down. It is not known how the appropriate levels of survival factors are maintained.

Tissue Renewal Does Not Have to Depend on Stem Cells: Insulin-Secreting Cells in the Pancreas

Most of the organs of the respiratory and digestive tract, including the lungs, the stomach, and the pancreas, contain a subpopulation of endocrine cells similar to the enteroendocrine cells in the intestines and, like them, generated in the epithelium under the control of the Notch signaling pathway. The *insulin-secreting cells (β cells)* of the pancreas belong in this category. Their mode of renewal has a special importance, because it is the loss of these cells (through autoimmune attack) that is responsible for Type I (juvenile-onset) diabetes and a significant factor also in the Type II (adult-onset) form of the disease. In a normal pancreas, they are sequestered in cell clusters, called *islets of Langerhans* (**Figure 23–29**), where they are grouped with related enteroendocrine cells, secreting other hormones. The islets contain no obvious subset of cells specialized to act as stem cells, yet fresh β cells are continually generated within them. Where do these new cells come from?

The question has been answered by study of transgenic mice in which an ingenious variant of the Cre-Lox technique (described in Chapter 8) was used to produce a marker mutation just in those cells that were expressing the insulin gene at the time a drug was given to activate Cre. In this way, the only cells that became labeled and transmitted the label to their progeny were those that were already differentiated β cells at the time of the treatment. When the mice were analyzed as much as a year later, all the new β cells carried the label, implying that they were descendants of already-differentiated β cells, and not of some undifferentiated stem cell. As in the liver, it seems that the population of differentiated cells here is renewed and enlarged by simple duplication of existing differentiated cells, and not by means of stem cells.

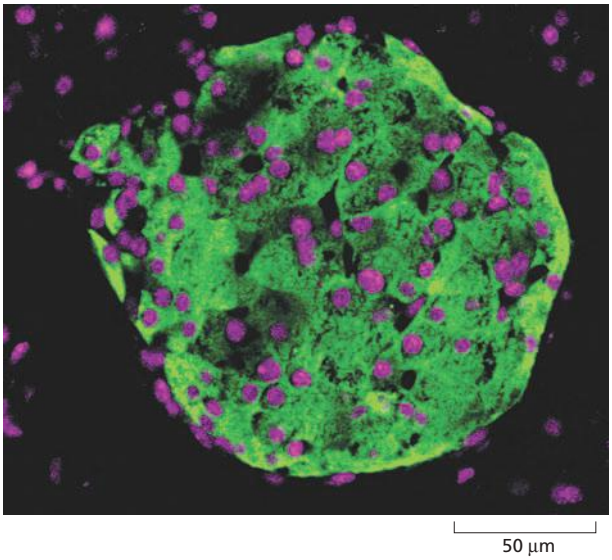


Figure 23–29 An islet of Langerhans in the pancreas. The insulin-secreting cells (β cells) are stained *green* by immunofluorescence. Cell nuclei are stained *purple* with a DNA dye. The surrounding pancreatic exocrine cells (secreting digestive enzymes and bicarbonate via ducts into the gut) are unstained, except for their nuclei. Within the islet, close to its surface, there are also small numbers of cells (unstained) secreting hormones such as glucagon. The insulin-secreting cells replace themselves by simple duplication, without need of specialized stem cells. (Adapted from a photograph courtesy of Yuval Dor. © 2004 Yuval Dor, The Hebrew University, Jerusalem.)

Summary

The lung performs a simple function—gas exchange—but its housekeeping systems are complex. Surfactant-secreting cells help to keep the alveoli from collapsing. Macrophages constantly scour the alveoli for dirt and microorganisms. A mucociliary escalator formed by mucus-secreting goblet cells and beating ciliated cells sweeps debris out of the airways.

In the gut, where more potentially damaging chemical processes occur, constant rapid cell renewal keeps the absorptive epithelium in good repair. In the small intestine, stem cells in the crypts generate new absorptive, goblet, enteroendocrine, and Paneth cells, replacing most of the epithelial lining of the intestine every week. Wnt signaling in the crypts maintains the stem-cell population, while Notch signaling drives diversification of the stem-cell progeny and limits the number that are consigned to a secretory fate. Cell–cell interactions within the epithelium mediated by ephrin–Eph signaling control the selective migration of cells from the crypts upward onto the villi. Interactions between the epithelium and the stroma, involving the Wnt, Hedgehog, PDGF, and BMP pathways organize the pattern of crypts and villi, thereby creating the niches that stem cells inhabit.

The liver is a more protected organ, but it too can rapidly adjust its size up or down by cell proliferation or cell death when the need arises. Differentiated hepatocytes remain able to divide throughout life, showing that a specialized class of stem cells is not always needed for tissue renewal. Similarly, the population of insulin-producing cells in the pancreas is enlarged and renewed by simple duplication of existing insulin-producing cells.

BLOOD VESSELS, LYMPHATICS, AND ENDOTHELIAL CELLS

From the tissues that derive from the embryonic ectoderm and endoderm, we turn now to those derived from *mesoderm*. This middle layer of cells, sandwiched between ectoderm and endoderm, grows and diversifies to provide many sorts of supportive functions. It gives rise to the body's connective tissues, blood cells, and blood and lymphatic vessels, as well as muscle, kidney, and many other structures and cell types. We begin with blood vessels.

Almost all tissues depend on a blood supply, and the blood supply depends on **endothelial cells**, which form the linings of the blood vessels. Endothelial cells have a remarkable capacity to adjust their number and arrangement to suit local requirements. They create an adaptable life-support system, extending by cell migration into almost every region of the body. If it were not for endothelial cells extending and remodeling the network of blood vessels, tissue growth and repair would be impossible. Cancerous tissue is as dependent on a blood supply as is normal tissue, and this has led to a surge of interest in endothelial cell biology. By blocking the formation of new blood vessels through drugs that act on endothelial cells, it may be possible to block the growth of tumors (discussed in Chapter 20).

Endothelial Cells Line All Blood Vessels and Lymphatics

The largest blood vessels are arteries and veins, which have a thick, tough wall of connective tissue and many layers of smooth muscle cells (**Figure 23–30**). The wall is lined by an exceedingly thin single sheet of endothelial cells, the *endothelium,* separated from the surrounding outer layers by a basal lamina. The amounts of connective tissue and smooth muscle in the vessel wall vary according to the vessel's diameter and function, but the endothelial lining is always present. In the finest branches of the vascular tree—the capillaries and sinusoids—the walls consist of nothing but endothelial cells and a basal lamina (**Figure 23–31**), together with a few scattered—but functionally important—*pericytes*. These are cells of the connective-tissue family, related to vascular smooth muscle cells, that wrap themselves around the small vessels (**Figure 23–32**).

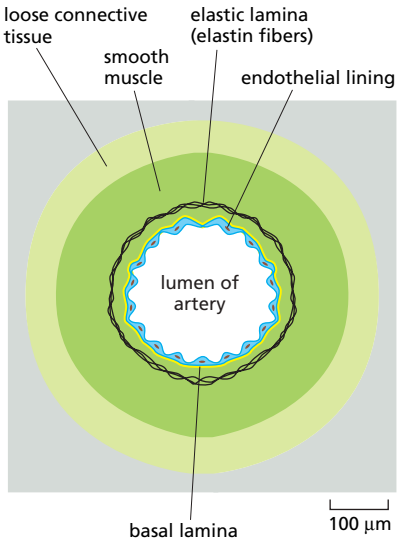


Figure 23–30 Diagram of a small artery in cross section. The endothelial cells, although inconspicuous, are the fundamental component. Compare with the capillary in Figure 23–31.

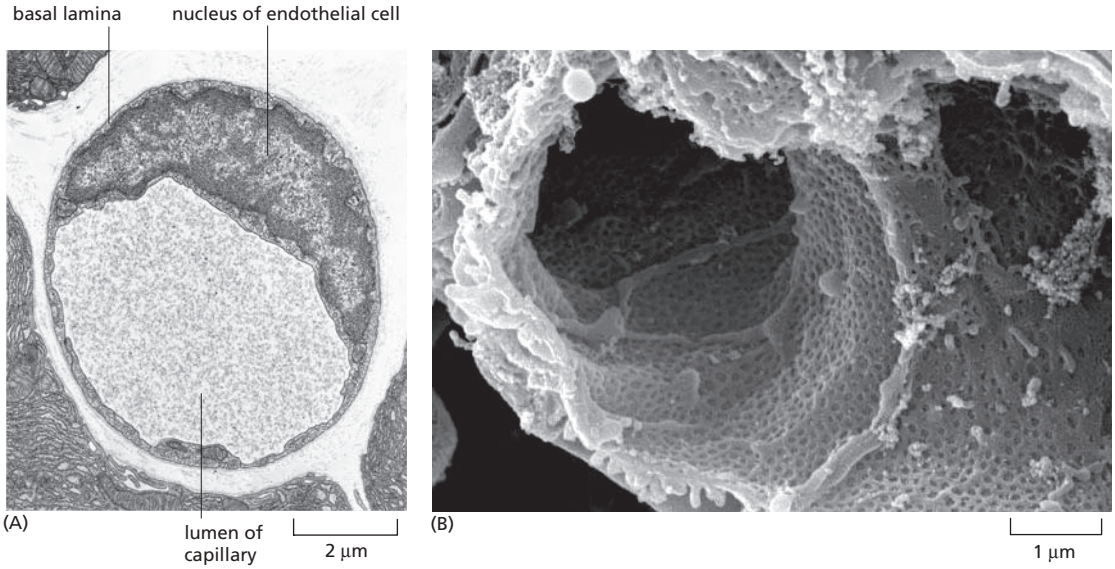


Figure 23–31 Capillaries. (A) Electron micrograph of a cross section of a small capillary in the pancreas. The wall is formed by a single endothelial cell surrounded by a basal lamina. (B) Scanning electron micrograph of the interior of a capillary in a glomerulus of the kidney, where blood filtration occurs to produce urine. Here, as in the liver (see Figure 23–28), the endothelial cells are specialized to form a sieve-like structure, with fenestrae, constructed rather like the pores in the nuclear envelope of eucaryotic cells, allowing water and most molecules to pass freely out of the bloodstream. (A, from R.P. Bolender, *J. Cell Biol.* 61:269–287, 1974. With permission from The Rockefeller University Press; B, courtesy of Steve Gschmeissner and David Shima.)

Less obvious than the blood vessels are the lymphatic vessels. These carry no blood and have much thinner and more permeable walls than the blood vessels. They provide a drainage system for the fluid (lymph) that seeps out of the blood vessels, as well as an exit route for white blood cells that have migrated from blood vessels into the tissues. Less happily, they often also provide the path by which cancer cells escape from a primary tumor to invade other tissues. The lymphatics form a branching system of tributaries all ultimately discharging into a single large lymphatic vessel, the thoracic duct, which opens into a large vein close to the heart. Like blood vessels, lymphatics are lined with endothelial cells.

Thus, endothelial cells line the entire blood and lymphatic vascular system, from the heart to the smallest capillary, and control the passage of materials—and the transit of white blood cells—into and out of the bloodstream. Arteries, veins, and lymphatics all develop from small vessels constructed primarily of endothelial cells and a basal lamina: connective tissue and smooth muscle are added later where required, under the influence of signals from the endothelial cells.

Endothelial Tip Cells Pioneer Angiogenesis

To understand how the vascular system comes into being and how it adapts to the changing needs of tissues, we have to understand endothelial cells. How do they become so widely distributed, and how do they form channels that connect in just the right way for blood to circulate through the tissues and for lymph to drain back to the bloodstream?

Endothelial cells originate at specific sites in the early embryo from precursors that also give rise to blood cells. From these sites the early embryonic endothelial cells migrate, proliferate, and differentiate to form the first rudiments of blood vessels—a process called *vasculogenesis*. Subsequent growth and branching of the vessels throughout the body is mainly by proliferation and movement of the endothelial cells of these first vessels, in a process called **angiogenesis**.

Angiogenesis occurs in a broadly similar way in the young organism as it grows and in the adult during tissue repair and remodeling. We can watch the behavior of the cells in naturally transparent structures, such as the cornea of the eye or the fin of a tadpole, or in tissue culture, or in the embryo. The embryonic



Figure 23–32 Pericytes. The scanning electron micrograph shows pericytes wrapping their processes around a small blood vessel (a post-capillary venule) in the mammary gland of a cat. Pericytes are present also around capillaries, but are much more sparsely distributed there. (From T. Fujiwara and Y. Uehara, *Am. J. Anat.* 170:39–54, 1984. With permission from Wiley-Liss.)

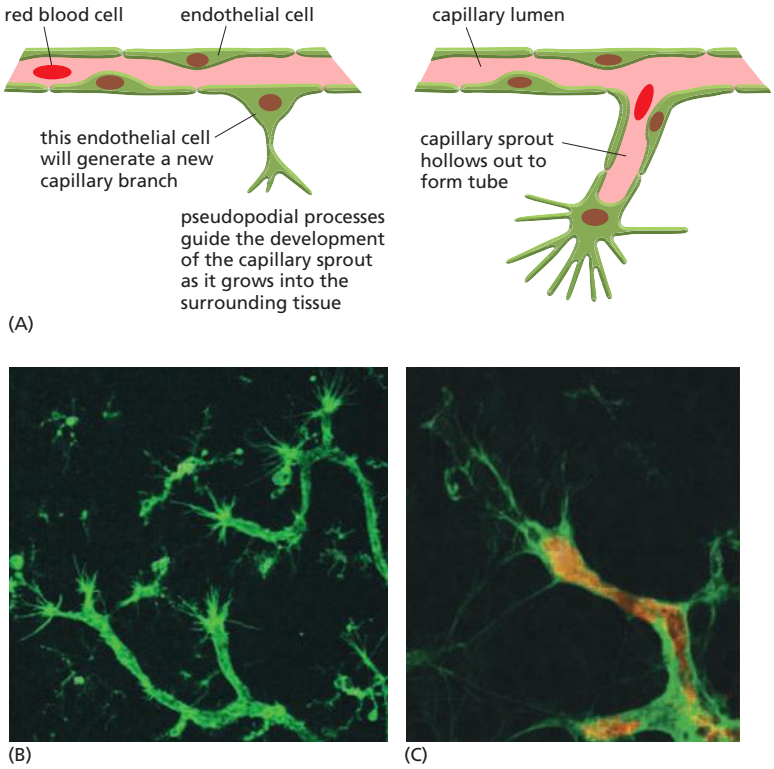


Figure 23–33 Angiogenesis. (A) A new blood capillary forms by the sprouting of an endothelial cell from the wall of an existing small vessel. An endothelial tip cell, with many filopodia, leads the advance of each capillary sprout. The endothelial stalk cells trailing behind the tip cell become hollowed out to form a lumen. (B) Blood capillaries sprouting in the retina of an embryonic mouse. (C) A similar specimen, but with a red dye injected into the bloodstream, revealing the capillary lumen opening up behind the tip cell. (B and C, from H. Gerhardt et al., *J. Cell Biol.* 161:1163–1177, 2003. With permission from The Rockefeller University Press.)

retina, which blood vessels invade according to a predictable timetable, is a convenient example for experimental study. Each new vessel originates as a capillary sprout from the side of an existing capillary or small venule (**Figure 23–33**A). At the tip of the sprout, leading the way, is an endothelial cell with a distinctive character. This *tip cell* has a pattern of gene expression somewhat different from that of the endothelial stalk cells following behind it, and while they divide, it does not; but the tip cell's most striking feature is that it puts out many long filopodia, resembling those of a neuronal growth cone (Figure 23–33B). The stalk cells, meanwhile, become hollowed out to form a lumen (see Figure 23–33A). One can watch this process in the transparent zebrafish embryo: the individual cells develop internal vacuoles that join up with those of their neighbors to create a continuous multicellular tube. <GTTG>

The endothelial tip cells that pioneer the growth of normal capillaries not only look like neuronal growth cones, but also respond similarly to signals in the environment. In fact, many of the same guidance molecules are involved, including semaphorins, netrins, slits, and ephrins, along with the corresponding receptors, which are expressed in the tip cells and guide the vascular sprouts along specific pathways in the embryo, often in parallel with nerves. Perhaps the most important of the guidance molecules for endothelial cells, however, is one that is specifically dedicated to the control of vascular development: *vascular endothelial growth factor*, or *VEGF*. We shall have more to say about it below.

Different Types of Endothelial Cells Form Different Types of Vessel

To create a new circuit for blood flow, a vascular sprout must continue to grow out until it encounters another sprout or vessel with which it can connect. The rules of connection presumably have to be selective, to prevent the formation of undesirable short circuits and to keep the blood and lymphatic systems properly segregated. In fact, endothelial cells of developing arterial, venous, and lymphatic vessels express different genes and have different surface properties. These differences evidently help guide the various types of vessels along different paths, control the selective formation of connections, and govern the development of different types of wall as the vessel enlarges. Arterial endothelial cells, in the embryo at least, express the transmembrane protein ephrinB2,

for example, while the venous arterial cells express the corresponding receptor protein, EphB4 (discussed in Chapter 15). These molecules mediate signaling at sites of cell–cell contact, and they are essential for the development of a properly organized network of vessels.

Expression of the gene regulatory protein Prox1 distinguishes the endothelial cells of lymphatic vessels from arterial and venous endothelial cells. This gene switches on in a subset of endothelial cells in the wall of a large vein (the cardinal vein) in the embryo, converting them into lymphatic progenitors. From these, the whole of the lymphatic vasculature derives by sprouting as described above. Prox1 causes the lymphatic endothelial cells to express receptors for a different member of the VEGF family of guidance molecules, as well as proteins that prevent the lymphatic cells from forming connections with blood vessels.

Tissues Requiring a Blood Supply Release VEGF; Notch Signaling Between Endothelial Cells Regulates the Response

Almost every cell, in almost every tissue of a vertebrate, is located within 50–100 μm of a blood capillary. What mechanism ensures that the system of blood vessels branches into every nook and cranny? How is it adjusted so perfectly to the local needs of the tissues, not only during normal development but also in pathological circumstances? Wounding, for example, induces a burst of capillary growth in the neighborhood of the damage, to satisfy the high metabolic requirements of the repair process (**Figure 23–34**). Local irritants and infections also cause a proliferation of new capillaries, most of which regress and disappear when the inflammation subsides. Less benignly, a small sample of tumor tissue implanted in the cornea, which normally lacks blood vessels, causes blood vessels to grow quickly toward the implant from the vascular margin of the cornea; the growth rate of the tumor increases abruptly as soon as the vessels reach it.

In all these cases, the invading endothelial cells respond to signals produced by the tissue that they invade. The signals are complex, but a key part is played by **vascular endothelial growth factor** (**VEGF**), a distant relative of platelet-derived growth factor (PDGF). The regulation of blood vessel growth to match the needs of the tissue depends on the control of VEGF production, through changes in the stability of its mRNA and in its rate of transcription. The latter control is relatively well understood. A shortage of oxygen, in practically any type of cell, causes an increase in the intracellular concentration of a gene regulatory protein called **hypoxia-inducible factor 1α** (**HIF1α**). HIF1α stimulates transcription of *Vegf* (and of other genes whose products are needed when oxygen is in short supply). The VEGF protein is secreted, diffuses through the tissue (with different isoforms of VEGF diffusing to different extents), and acts on nearby endothelial cells, stimulating them to proliferate, to produce proteases to help them digest their way through the basal lamina of the parent capillary or venule, and to form sprouts. The tip cells of the sprouts detect the VEGF

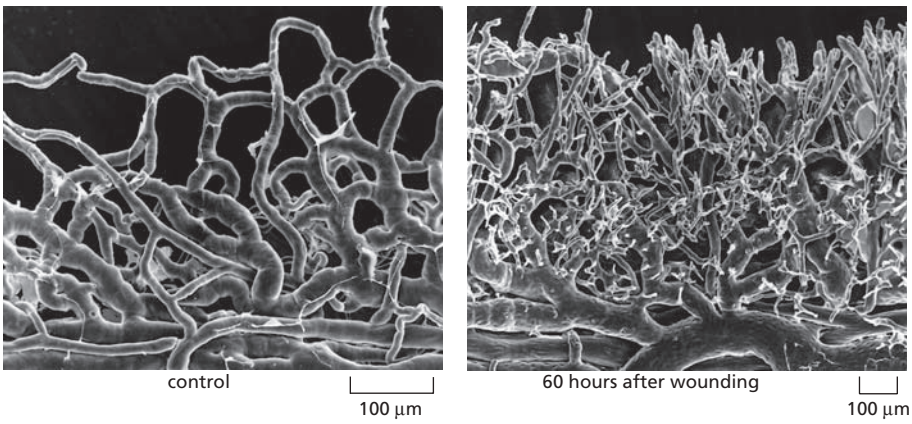


Figure 23–34 New capillary formation in response to wounding. Scanning electron micrographs of casts of the system of blood vessels surrounding the margin of the cornea show the reaction to wounding. The casts are made by injecting a resin into the vessels and letting the resin set; this reveals the shape of the lumen, as opposed to the shape of the cells. Sixty hours after wounding many new capillaries have begun to sprout toward the site of injury, which is just above the top of the picture. Their oriented outgrowth reflects a chemotactic response of the endothelial cells to an angiogenic factor released at the wound. (Courtesy of Peter C. Burger.)

gradient and move toward its source. (Other growth factors, including some members of the fibroblast growth factor family, can also stimulate angiogenesis, mediating reactions to other conditions such as inflammation.)

As the new vessels form, bringing blood to the tissue, the oxygen concentration rises, HIF1α activity declines, VEGF production is shut off, and angiogenesis comes to a halt (**Figure 23–35**). As in all signaling systems, it is as important to switch the signal off correctly as to switch it on. In normal well-oxygenated tissue, continual degradation of the HIF1α protein keeps the concentration of HIF1α low: in the presence of oxygen, an oxygen-requiring enzyme modifies HIF1α so as to target it for degradation. Degradation in turn requires the product of another gene, coding for an E3 ubiquitin ligase subunit, which is defective in a rare disorder called *von Hippel–Lindau (VHL) syndrome*. People with this condition are born with only one functional copy of the *Vhl* gene; mutations occurring at random in the body then give rise to cells with two defective gene copies. These cells contain large quantities of HIF1 regardless of oxygen availability, triggering the continual overproduction of VEGF. The result is development of *hemangioblastomas*, tumors that contain dense masses of blood vessels. The mutant cells that produce the VEGF are apparently themselves encouraged to proliferate by the over-rich nourishment provided by the excess blood vessels, creating a vicious cycle that promotes tumor growth. Loss of the VHL gene product also gives rise to other tumors as well as hemangioblastomas, by mechanisms that may be independent of effects on angiogenesis.

This is not the whole story of how angiogenesis is controlled, however. VEGF and related factors from the target tissue are essential to stimulate and guide angiogenesis, but interactions between one endothelial cell and another, mediated by the Notch signaling pathway, also have a critical role. These interactions control which cells will be singled out to behave as tip cells, extending filopodia and crawling forward to create new vascular sprouts, and they are required to bring this motile behavior to a halt when it is time to stop. Thus, when endothelial sprouts meet and join up to form a vascular circuit, they normally switch off to reduce their sprouting activities. The effect depends on a specific Notch ligand, called Delta4, which is expressed in tip cells and activates Notch in their neighbors; Notch activation leads to reduced expression of VEGF receptors, making the neighbors of the tip cell unresponsive to VEGF. In mutants where Notch signaling is defective, sprouting behavior continues inappropriately and fails to be confined to tip cells. The result is an excessively dense network of ill-organized, dysfunctional vessels that carry little or no blood.

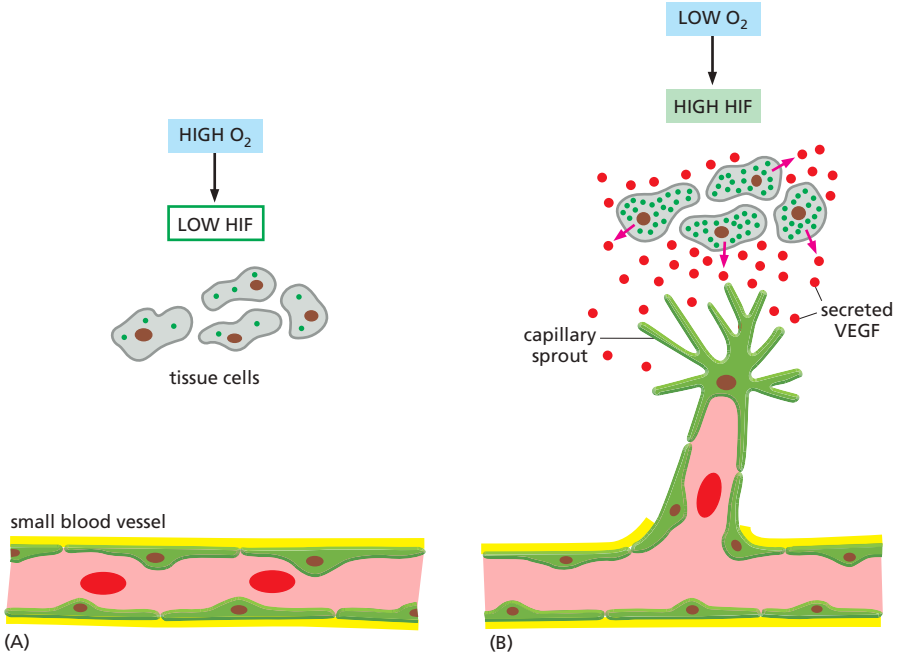


Figure 23–35 The regulatory mechanism controlling blood vessel growth according to a tissue's need for oxygen. Lack of oxygen triggers the secretion of VEGF, which stimulates angiogenesis.

Signals from Endothelial Cells Control Recruitment of Pericytes and Smooth Muscle Cells to Form the Vessel Wall

The vascular network is continually remodeled as it grows and adapts. A newly formed vessel may enlarge; or it may sprout side branches; or it may regress. Smooth muscle and other connective-tissue cells that pack themselves around the endothelium (see Figure 23–32) help to stabilize vessels as they enlarge. This process of vessel wall formation begins with recruitment of pericytes. Small numbers of these cells travel outward in company with the stalk cells of each endothelial sprout. The recruitment and proliferation of pericytes and smooth muscle cells to form a vessel wall depend on PDGF-B secreted by the endothelial cells and on PDGF receptors in the pericytes and smooth muscle cells. In mutants lacking this signal protein or its receptor, these vessel wall cells in many regions are missing. As a result, the embryonic blood vessels develop microaneurysms—microscopic pathological dilatations—that eventually rupture, as well as other abnormalities, reflecting the importance of signals exchanged in both directions between the exterior cells of the wall and the endothelial cells.

Once a vessel has matured, signals from the endothelial cells to the surrounding connective tissue and smooth muscle continue to regulate the vessel's function and structure. For example, the endothelial cells have mechanoreceptors that allow them to sense the shear stress due to flow of blood over their surface. The cells react by generating and releasing the gas NO, thereby signaling to the surrounding cells and inducing changes in the vessel's diameter and wall thickness to accommodate the blood flow. Endothelial cells also mediate rapid responses to neural signals for blood vessel dilation, by releasing NO to make smooth muscle relax in the vessel wall, as discussed in Chapter 15.

Summary

Endothelial cells are the fundamental elements of the vascular system. They form a single cell layer that lines all blood vessels and lymphatics and regulates exchanges between the bloodstream and the surrounding tissues. New vessels originate as endothelial sprouts from the walls of existing small vessels. A specialized motile endothelial tip cell at the leading edge of each sprout puts out filopodia that respond to gradients of guidance molecules in the environment, leading the growth of the sprout like the growth cone of a neuron. The endothelial stalk cells following behind become hollowed out to form a capillary tube. Endothelial cells of developing arteries, veins, and lymphatics express different cell-surface proteins, which may control the way in which they link up to create the vascular networks. Signals from endothelial cells organize the growth and development of the connective-tissue cells that form the surrounding layers of the vessel wall.

A homeostatic mechanism ensures that blood vessels permeate every region of the body. Cells that are short of oxygen increase their concentration of hypoxia-inducible factor (HIF1α), which stimulates the production of vascular endothelial growth factor (VEGF). VEGF acts on endothelial cells, causing them to proliferate and invade the hypoxic tissue to supply it with new blood vessels. The endothelial cells also interact with one another via the Notch pathway. This exchange of Notch signals is necessary to limit the number of cells that behave as tip cells and to halt angiogenic behavior when tip cells meet.

RENEWAL BY MULTIPOTENT STEM CELLS: BLOOD CELL FORMATION

Blood contains many types of cells, with functions that range from the transport of oxygen to the production of antibodies. Some of these cells stay within the vascular system, while others use the vascular system only as a means of transport and perform their function elsewhere. All blood cells, however, have certain similarities in their life history. They all have limited life spans and are produced

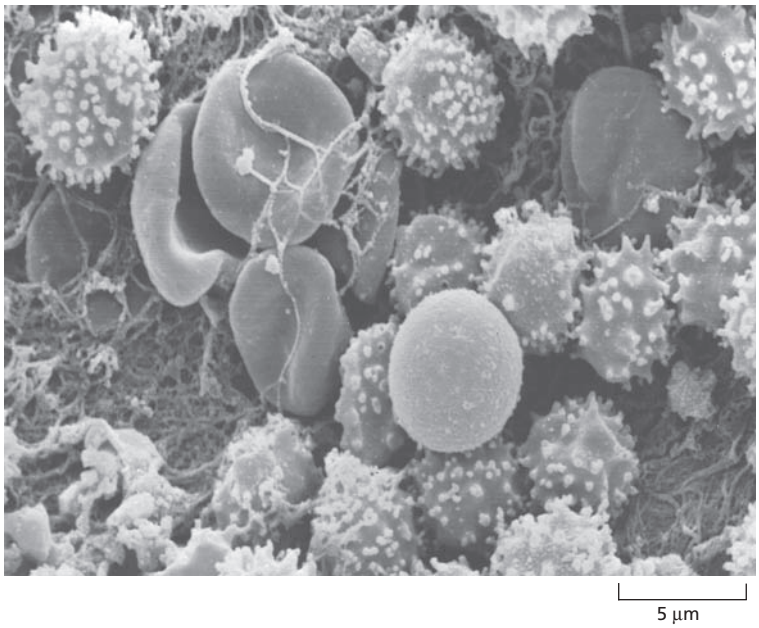


Figure 23–36 Scanning electron micrograph of mammalian blood cells caught in a blood clot. The larger, more spherical cells with a rough surface are white blood cells; the smoother, flattened cells are red blood cells. (Courtesy of Ray Moss.)

throughout the life of the animal. Most remarkably, they are all generated ultimately from a common stem cell in the bone marrow. This *hemopoietic* (blood-forming, also called *hematopoietic*) *stem cell* is thus multipotent, giving rise to all the types of terminally differentiated blood cells as well as some other types of cells, such as osteoclasts in bone, which we discuss later.

Blood cells can be classified as red or white (**Figure 23–36**). The **red blood cells**, or **erythrocytes**, remain within the blood vessels and transport O_2 and CO_2 bound to hemoglobin. The **white blood cells**, or **leucocytes**, combat infection and in some cases phagocytose and digest debris. Leucocytes, unlike erythrocytes, must make their way across the walls of small blood vessels and migrate into tissues to perform their tasks. In addition, the blood contains large numbers of **platelets**, which are not entire cells but small, detached cell fragments or "minicells" derived from the cortical cytoplasm of large cells called *megakaryocytes*. Platelets adhere specifically to the endothelial cell lining of damaged blood vessels, where they help to repair breaches and aid in blood clotting.

The Three Main Categories of White Blood Cells Are Granulocytes, Monocytes, and Lymphocytes

All red blood cells belong in a single class, following the same developmental trajectory as they mature, and the same is true of platelets; but there are many distinct types of white blood cells. White blood cells are traditionally grouped into three major categories—granulocytes, monocytes, and lymphocytes—based on their appearance in the light microscope.

Granulocytes contain numerous lysosomes and secretory vesicles (or granules) and are subdivided into three classes according to the morphology and staining properties of these organelles (**Figure 23–37**). The differences in staining reflect major differences of chemistry and function. *Neutrophils* (also called *polymorphonuclear leucocytes* because of their multilobed nucleus) are the most common type of granulocyte; they phagocytose and destroy microorganisms, especially bacteria, and thus have a key role in innate immunity to bacterial infection, as discussed in Chapter 25. *Basophils* secrete histamine (and, in some species, serotonin) to help mediate inflammatory reactions; they are closely related to *mast cells*, which reside in connective tissues but are also generated from the hemopoietic stem cells. *Eosinophils* help to destroy parasites and modulate allergic inflammatory responses.

Once they leave the bloodstream, **monocytes** (see Figure 23–37D) mature into **macrophages**, which, together with neutrophils, are the main "professional phagocytes" in the body. As discussed in Chapter 13, both types of phagocytic

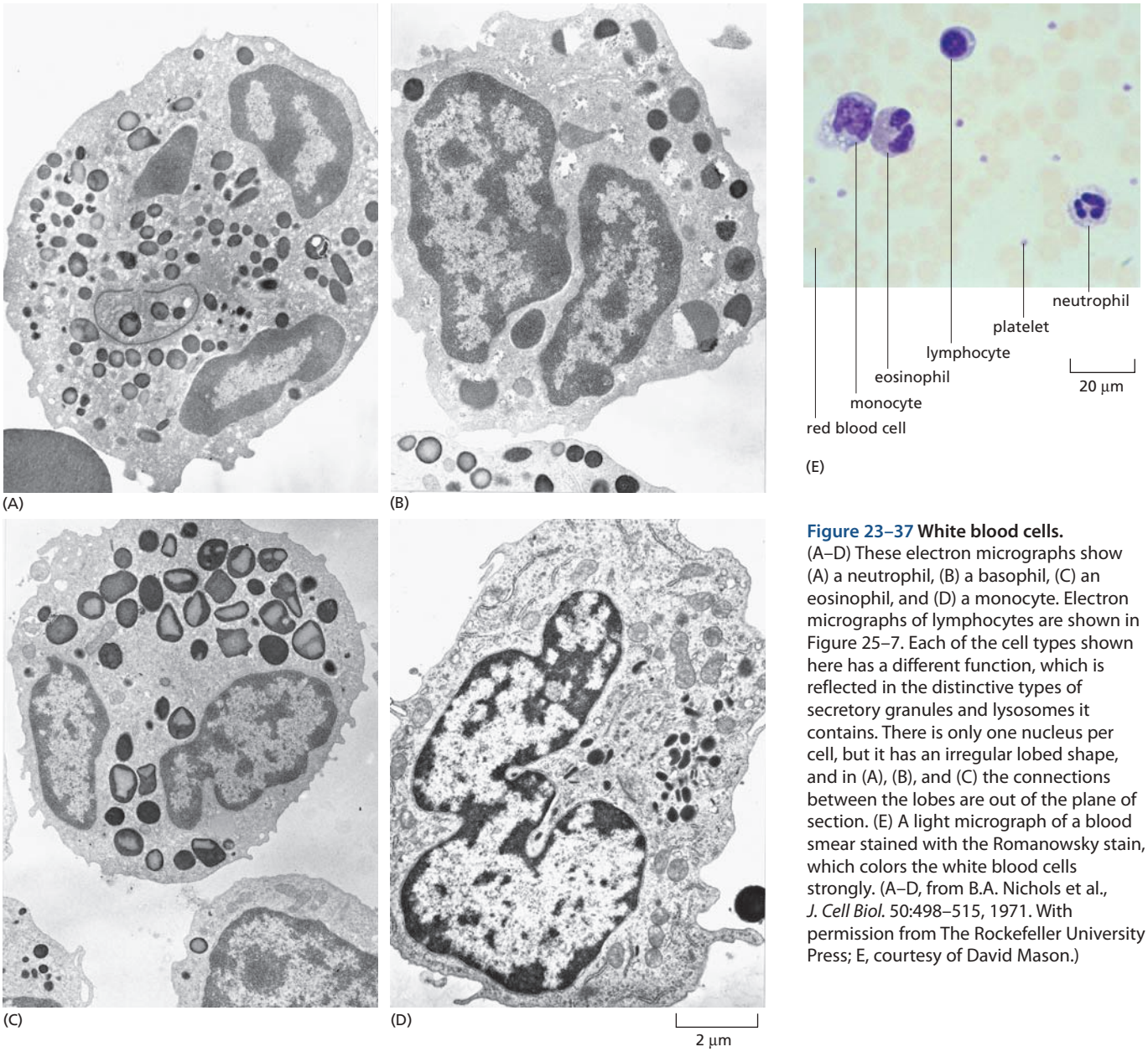


Figure 23–37 White blood cells. (A–D) These electron micrographs show (A) a neutrophil, (B) a basophil, (C) an eosinophil, and (D) a monocyte. Electron micrographs of lymphocytes are shown in Figure 25–7. Each of the cell types shown here has a different function, which is reflected in the distinctive types of secretory granules and lysosomes it contains. There is only one nucleus per cell, but it has an irregular lobed shape, and in (A), (B), and (C) the connections between the lobes are out of the plane of section. (E) A light micrograph of a blood smear stained with the Romanowsky stain, which colors the white blood cells strongly. (A–D, from B.A. Nichols et al., *J. Cell Biol.* 50:498–515, 1971. With permission from The Rockefeller University Press; E, courtesy of David Mason.)

cells contain specialized lysosomes that fuse with newly formed phagocytic vesicles (phagosomes), exposing phagocytosed microorganisms to a barrage of enzymatically produced, highly reactive molecules of superoxide (O_2^-) and hypochlorite (HOCl, the active ingredient in bleach), as well as to attack by a concentrated mixture of lysosomal hydrolases that become activated in the phagosome. Macrophages, however, are much larger and longer-lived than neutrophils. They recognize and remove senescent, dead, and damaged cells in many tissues, and they are unique in being able to ingest large microorganisms such as protozoa.

Monocytes also give rise to *dendritic cells*, such as the *Langerhans cells* scattered in the epidermis. Like macrophages, dendritic cells are migratory cells that can ingest foreign substances and organisms; but they do not have as active an appetite for phagocytosis and are instead specialized as presenters of foreign antigens to lymphocytes to trigger an immune response. Langerhans cells, for example, ingest foreign antigens in the epidermis and carry these trophies back to present to lymphocytes in lymph nodes.

There are two main classes of **lymphocytes**, both involved in immune responses: *B lymphocytes* make antibodies, while *T lymphocytes* kill virus-infected cells and regulate the activities of other white blood cells. In addition,

there are lymphocytelike cells called *natural killer (NK) cells*, which kill some types of tumor cells and virus-infected cells. The production of lymphocytes is a specialized topic discussed in detail in Chapter 25. Here we concentrate mainly on the development of the other blood cells, often referred to collectively as **myeloid cells**.

Table 23–1 summarizes the various types of blood cells and their functions.

The Production of Each Type of Blood Cell in the Bone Marrow Is Individually Controlled

Most white blood cells function in tissues other than the blood; blood simply transports them to where they are needed. A local infection or injury in any tissue rapidly attracts white blood cells into the affected region as part of the inflammatory response, which helps fight the infection or heal the wound.

The inflammatory response is complex and is governed by many different signal molecules produced locally by mast cells, nerve endings, platelets, and white blood cells, as well as by the activation of complement (discussed in Chapters 24 and 25). Some of these signal molecules act on nearby capillaries, causing the endothelial cells to adhere less tightly to one another but making their surfaces adhesive to passing white blood cells. The white blood cells are thus caught like flies on flypaper and then can escape from the vessel by squeezing between the endothelial cells and using digestive enzymes to crawl across the basal lamina. <ACCG> As discussed in Chapter 19, homing receptors called *selectins* mediate the initial binding to endothelial cells, while *integrins* mediate the stronger binding required for the white blood cells to crawl out of

Table 23–1 Blood Cells

TYPE OF CELL	MAIN FUNCTIONS	TYPICAL CONCENTRATION IN HUMAN BLOOD (CELLS/LITER)
Red blood cells (erythrocytes)	transport O_2 and CO_2	5×10^{12}
White blood cells (leucocytes)		
Granulocytes		
Neutrophils (polymorphonuclear leucocytes)	phagocytose and destroy invading bacteria	5×10^9
Eosinophils	destroy larger parasites and modulate allergic inflammatory responses	2×10^8
Basophils	release histamine (and in some species serotonin) in certain immune reactions	4×10^7
Monocytes	become tissue macrophages, which phagocytose and digest invading microorganisms and foreign bodies as well as damaged senescent cells	4×10^8
Lymphocytes		
B cells	make antibodies	2×10^9
T cells	kill virus-infected cells and regulate activities of other leucocytes	1×10^9
Natural killer (NK) cells	kill virus-infected cells and some tumor cells	1×10^8
Platelets (cell fragments arising from *megakaryocytes* in bone marrow)	initiate blood clotting	3×10^{11}

Humans contain about 5 liters of blood, accounting for 7% of body weight. Red blood cells constitute about 45% of this volume and white blood cells about 1%, the rest being the liquid blood plasma.

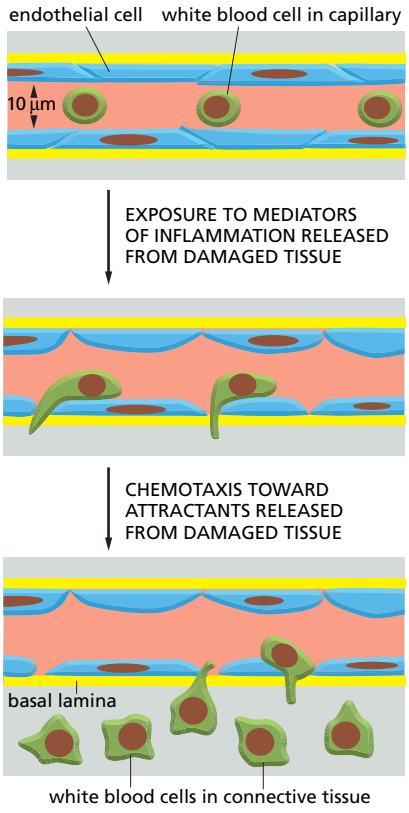


Figure 23–38 The migration of white blood cells out of the bloodstream during an inflammatory response. The response is initiated by signal molecules produced by cells in the neighborhood (mainly in the connective tissue) or by complement activation. Some of these mediators act on capillary endothelial cells, causing them to loosen their attachments to their neighbors so that the capillaries become more permeable. Endothelial cells are also stimulated to express selectins, cell-surface molecules that recognize specific carbohydrates that are present on the surface of leucocytes in the blood and cause them to stick to the endothelium. The inflamed tissues and local endothelial cells secrete other mediators called chemokines, and the chemokines act as chemoattractants, causing the bound leucocytes to crawl between the capillary endothelial cells into the tissue.

the blood vessel (see Figure 19–19). Damaged or inflamed tissues and local endothelial cells secrete other molecules called *chemokines*, which act as chemoattractants for specific types of white blood cells, causing them to become polarized and crawl toward the source of the attractant. As a result, large numbers of white blood cells enter the affected tissue (**Figure 23–38**).

Other signal molecules produced during an inflammatory response escape into the blood and stimulate the bone marrow to produce more leucocytes and release them into the bloodstream. The bone marrow is the key target for such regulation because, with the exception of lymphocytes and some macrophages, most types of blood cells in adult mammals are generated only in the bone marrow. The regulation tends to be cell-type-specific: some bacterial infections, for example, cause a selective increase in neutrophils, while infections with some protozoa and other parasites cause a selective increase in eosinophils. (For this reason, physicians routinely use differential white blood cell counts to aid in the diagnosis of infectious and other inflammatory diseases.)

In other circumstances erythrocyte production is selectively increased—for example, in the process of acclimatization when one goes to live at high altitude, where oxygen is scarce. Thus, blood cell formation, or *hemopoiesis* (also called *hematopoiesis*), necessarily involves complex controls, which regulate the production of each type of blood cell individually to meet changing needs. It is a problem of great medical importance to understand how these controls operate.

In intact animals, hemopoiesis is more difficult to analyze than is cell turnover in a tissue such as the epidermis or the lining of the gut, where a simple, regular spatial organization makes it easy to follow the process of renewal and to locate the stem cells. The hemopoietic tissues do not appear so orderly. However, hemopoietic cells have a nomadic lifestyle that makes them more accessible to experimental study in other ways. It is easy to obtain dispersed hemopoietic cells and to transfer them, without damage, from one animal to another. Moreover, the proliferation and differentiation of individual cells and their progeny can be observed and analyzed in culture, and numerous molecular markers distinguish the various stages of differentiation. Because of this, more is known about the molecules that control blood cell production than about those that control cell production in other mammalian tissues. Studies of hemopoiesis have strongly influenced current ideas about stem-cell systems in general.

Bone Marrow Contains Hemopoietic Stem Cells

Routine staining methods allow us to recognize the different types of blood cells and their immediate precursors in the bone marrow (**Figure 23–39**). Here, these cells are intermingled with one another, as well as with fat cells and other stromal cells (connective-tissue cells), which produce a delicate supporting meshwork of collagen fibers and other extracellular matrix components. In addition, the whole tissue is richly supplied with thin-walled blood vessels, called *blood sinuses*, into which the new blood cells are discharged. **Megakaryocytes** are also present; these, unlike other blood cells, remain in the bone marrow when

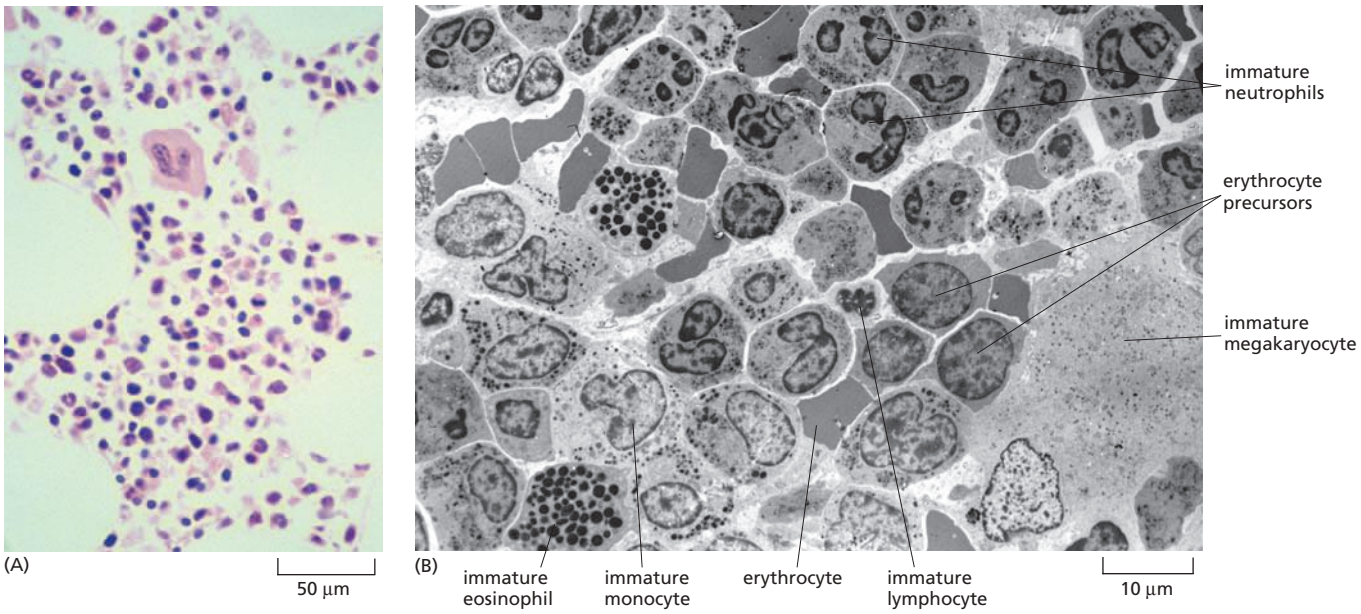


Figure 23–39 Bone marrow. (A) A light micrograph of a stained section. The large empty spaces correspond to fat cells, whose fatty contents have been dissolved away during specimen preparation. The giant cell with a lobed nucleus is a megakaryocyte. (B) A low-magnification electron micrograph. Bone marrow is the main source of new blood cells (except for T lymphocytes, which are produced in the thymus). Note that the immature blood cells of a particular type tend to cluster in "family groups." (A, courtesy of David Mason; B, from J.A.G. Rhodin, Histology: A Text and Atlas. New York: Oxford University Press, 1974.)

mature and are one of its most striking features, being extraordinarily large (diameter up to 60 μm), with a highly polyploid nucleus. They normally lie close beside blood sinuses, and they extend processes through holes in the endothelial lining of these vessels; platelets pinch off from the processes and are swept away into the blood (**Figure 23–40**). <GCAT>

Because of the complex arrangement of the cells in bone marrow, it is difficult to identify in ordinary tissue sections any but the immediate precursors of the mature blood cells. The corresponding cells at still earlier stages of development, before any overt differentiation has begun, look confusingly similar, and although the spatial distribution of cell types has some orderly features, there is no obvious visible characteristic by which we can recognize the ultimate stem cells. To identify and characterize the stem cells, we need a functional assay, which involves tracing the progeny of single cells. As we shall see, this can be done *in vitro* simply by examining the colonies that isolated cells produce in culture. The hemopoietic system, however, can also be manipulated so that such clones of cells can be recognized *in vivo* in the intact animal.

When an animal is exposed to a large dose of x-rays, most of the hemopoietic cells are destroyed and the animal dies within a few days as a result of its

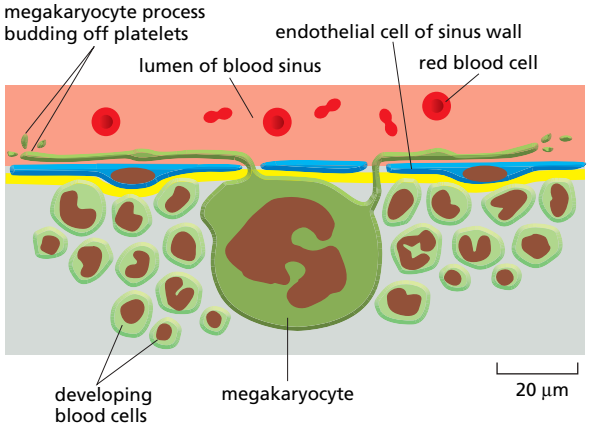


Figure 23–40 A megakaryocyte among other cells in the bone marrow. Its enormous size results from its having a highly polyploid nucleus. One megakaryocyte produces about 10,000 platelets, which split off from long processes that extend through holes in the walls of an adjacent blood sinus.

inability to manufacture new blood cells. The animal can be saved, however, by a transfusion of cells taken from the bone marrow of a healthy, immunologically compatible donor. Among these cells there are some that can colonize the irradiated host and permanently reequip it with hemopoietic tissue (**Figure 23–41**). Such experiments prove that the marrow contains hemopoietic stem cells. They also show how we can assay for the presence of hemopoietic stem cells and hence discover the molecular features that distinguish them from other cells.

For this purpose, cells taken from bone marrow are sorted (using a fluorescence-activated cell sorter) according to the surface antigens that they display, and the different fractions are transfused back into irradiated mice. If a fraction rescues an irradiated host mouse, it must contain hemopoietic stem cells. In this way, it has been possible to show that the hemopoietic stem cells are characterized by a specific combination of cell-surface proteins, and by appropriate sorting we can obtain virtually pure stem cell preparations. The stem cells turn out to be a tiny fraction of the bone marrow population—about 1 cell in 10,000; but this is enough. As few as five such cells injected into a host mouse with defective hemopoiesis are sufficient to reconstitute its entire hemopoietic system, generating a complete set of blood cell types, as well as fresh stem cells.

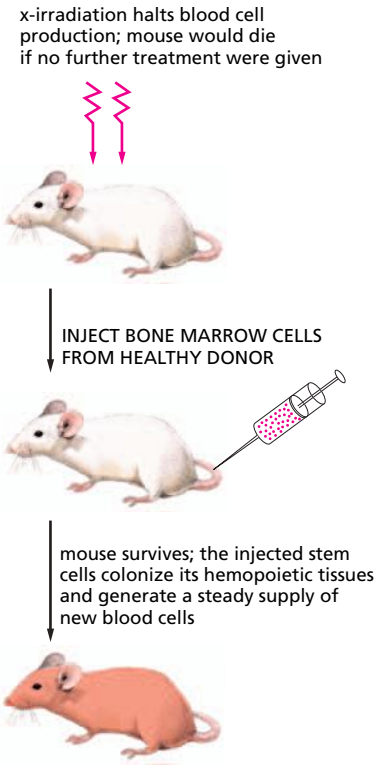


Figure 23–41 Rescue of an irradiated mouse by a transfusion of bone marrow cells. An essentially similar procedure is used in the treatment of leukemia in human patients by bone marrow transplantation.

A Multipotent Stem Cell Gives Rise to All Classes of Blood Cells

To see what range of cell types a single **hemopoietic stem cell** can generate, we need a way to trace the fate of its progeny. This can be done by marking individual stem cells genetically, so that their progeny can be identified even after they have been released into the bloodstream. Although several methods have been used for this, a specially engineered retrovirus (a retroviral vector carrying a marker gene) serves the purpose particularly well. The marker virus, like other retroviruses, can insert its own genome into the chromosomes of the cell it infects, but the genes that would enable it to generate new infectious virus particles have been removed. The marker is therefore confined to the progeny of the cells that were originally infected, and the progeny of one such cell can be distinguished from the progeny of another because the chromosomal sites of insertion of the virus are different. To analyze hemopoietic cell lineages, bone marrow cells are first infected with the retroviral vector *in vitro* and then are transferred into a lethally irradiated recipient; DNA probes can then be used to trace the progeny of individual infected cells in the various hemopoietic and lymphoid tissues of the host. These experiments show that the individual hemopoietic stem cell is *multipotent* and can give rise to the complete range of blood cell types, both myeloid and lymphoid, as well as new stem cells like itself (**Figure 23–42**).

Later in this chapter, we explain how the same methods that were developed for experimentation in mice can now be used for treatment of disease in humans.

Commitment Is a Stepwise Process

Hemopoietic stem cells do not jump directly from a multipotent state into a commitment to just one pathway of differentiation; instead, they go through a series of progressive restrictions. The first step, usually, is commitment to either a myeloid or a lymphoid fate. This is thought to give rise to two kinds of progenitor cells, one capable of generating large numbers of all the different types of myeloid cells, or perhaps of myeloid cells plus B lymphocytes, and the other giving rise to large numbers of all the different types of lymphoid cells, or at least T lymphocytes. Further steps give rise to progenitors committed to the production of just one cell type. The steps of commitment correlate with changes in the expression of specific gene regulatory proteins, needed for the production of different subsets of blood cells. These proteins seem to act in a complicated combinatorial fashion: the *GATA1 protein*, for example, is needed for the maturation of red blood cells, but is active also at much earlier steps in the hemopoietic pathway.

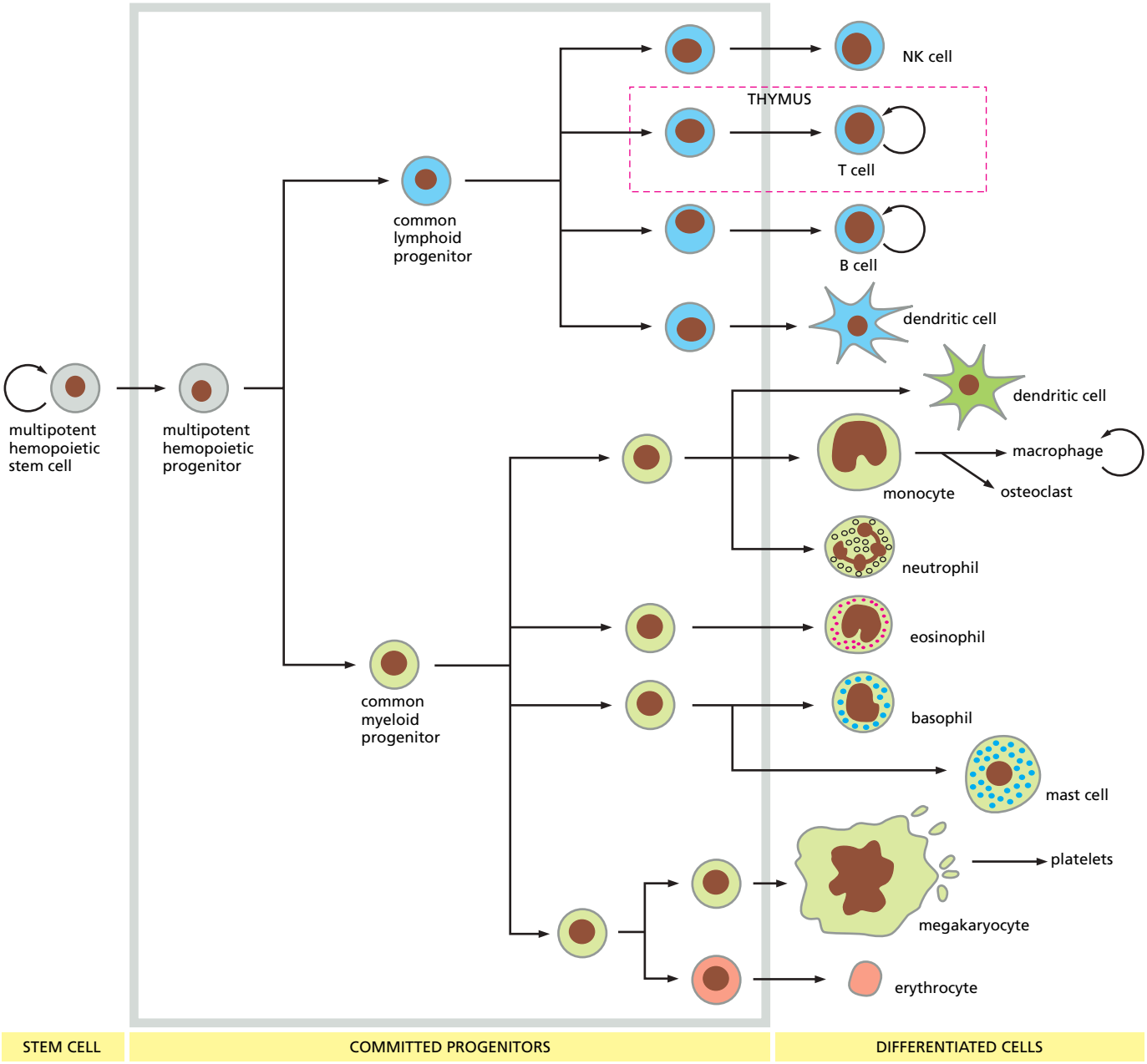


Figure 23–42 A tentative scheme of hemopoiesis. The multipotent stem cell normally divides infrequently to generate either more multipotent stem cells, which are self-renewing, or committed progenitor cells, which are limited in the number of times that they can divide before differentiating to form mature blood cells. As they go through their divisions, the progenitors become progressively more specialized in the range of cell types that they can give rise to, as indicated by the branching of the cell-lineage diagram in the region enclosed in the *gray* box. Many of the details of this part of the lineage diagram are still controversial, however. In adult mammals, all of the cells shown develop mainly in the bone marrow—except for T lymphocytes, which develop in the thymus, and macrophages and osteoclasts, which develop from blood monocytes. Some dendritic cells may also derive from monocytes.

Divisions of Committed Progenitor Cells Amplify the Number of Specialized Blood Cells

Hemopoietic progenitor cells generally become committed to a particular pathway of differentiation long before they cease proliferating and terminally differentiate. The committed progenitors go through many rounds of cell division to amplify the ultimate number of cells of the given specialized type. In this way, a single stem-cell division can lead to the production of thousands of differentiated progeny, which explains why the number of stem cells is such a small fraction of the total population of hemopoietic cells. For the same reason, a high rate

of blood cell production can be maintained even though the stem-cell division rate is low. As noted earlier, infrequent division or quiescence is a common feature of stem cells in several tissues. By reducing the number of division cycles that the stem cells themselves have to undergo in the course of a lifetime, it lowers the risk of generating stem-cell mutations, which would give rise to persistent mutant clones of cells in the body. It also has another effect: it reduces the rate of replicative senescence (discussed in Chapter 17). In fact, hemopoietic stem cells that are forced to keep dividing rapidly (through knockout of a gene called *Gfi1* that restricts their proliferation rate, or by other means) fail to sustain hemopoiesis for a full normal lifespan.

The stepwise nature of commitment means that the hemopoietic system can be viewed as a hierarchical family tree of cells. Multipotent stem cells give rise to *committed progenitor cells*, which are specified to give rise to only one or a few blood cell types. The committed progenitors divide rapidly, but only a limited number of times, before they terminally differentiate into cells that divide no further and die after several days or weeks. Many cells normally die at the earlier steps in the pathway as well. Studies in culture provide a way to find out how the proliferation, differentiation, and death of the hemopoietic cells are regulated.

Stem Cells Depend on Contact Signals From Stromal Cells

Hemopoietic cells can survive, proliferate, and differentiate in culture if, and only if, they are provided with specific signal proteins or are accompanied by cells that produce these proteins. If deprived of such proteins, the cells die. For long-term maintenance, contact with appropriate supporting cells also seems to be necessary: hemopoiesis can be kept going for months or even years *in vitro* by culturing dispersed bone marrow hemopoietic cells on top of a layer of bone-marrow stromal cells, which mimic the environment in intact bone marrow. Such cultures can generate all the types of myeloid cells, and their long-term continuation implies that stem cells, as well as differentiated progeny, are being continually produced.

In the bone marrow, where they normally live, the hemopoietic stem cells are mostly located in close contact with the *osteoblasts* that line the bony surfaces of the marrow cavity—the cells that produce the bone matrix. Treatments and mutations that increase or decrease the number of osteoblasts cause corresponding changes in the numbers of hemopoietic stem cells. This suggests that the osteoblasts provide the signals that the hemopoietic stem cells need to keep them in their uncommitted stem-cell state, just as the intestinal crypt provides the signals needed to maintain stem cells of the gut epithelium. In both systems, stem cells are normally confined to a particular niche, and when they leave this niche they tend to lose their stem-cell potential (**Figure 23–43**). Hemopoietic stem cells in the bone marrow and elsewhere are also often associated with a specialized class of endothelial cells, which may provide them with an alternative niche.

A key feature of the stem-cell niche in the bone marrow, as in the gut, is that it provides stimulation of the Wnt signaling pathway. Artificial activation of this pathway in cultured hemopoietic stem cells helps them to survive, proliferate, and keep their character as stem cells, while blocking Wnt signaling does the opposite. Another interaction that is important for the maintenance of hemopoiesis came to light through the analysis of mouse mutants with a curious combination of defects: a shortage of red blood cells (anemia), of germ cells (sterility), and of pigment cells (white spotting of the skin; see Figure 22–86). As discussed in Chapter 22, this syndrome results from mutations in either of two genes: one, called *Kit*, codes for a receptor tyrosine kinase; the other codes for its ligand. The cell types affected by the mutations all derive from migratory precursors, and it seems that these precursors in each case must express the receptor and be provided with the ligand by their environment if they are to survive and produce progeny in normal numbers. Studies in mutant mice suggest that Kit ligand must be membrane-bound to be fully effective, implying that normal hemopoiesis requires direct cell–cell contact between the hemopoietic

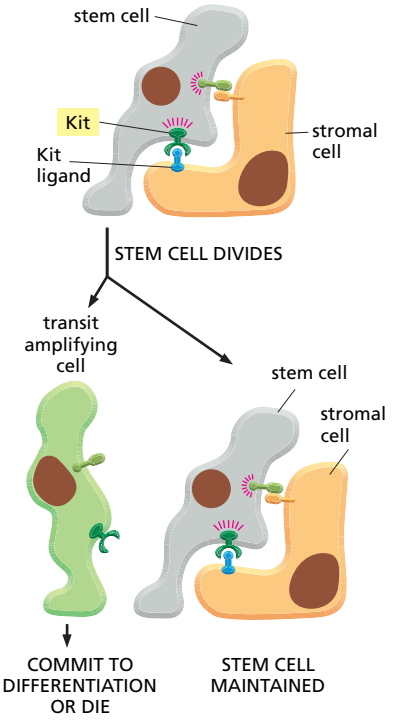


Figure 23–43 Dependence of hemopoietic stem cells on contact with stromal cells. The contact-dependent interaction between Kit and its ligand is one of several signaling mechanisms thought to be involved in hemopoietic stem-cell maintenance. The real system is certainly more complex; the dependence of hemopoietic cells on contact with stromal cells cannot be absolute, since small numbers of the functional stem cells can be found free in the circulation.

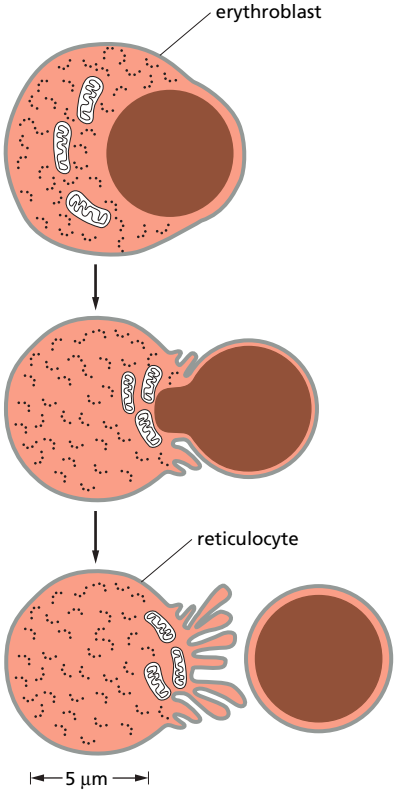


Figure 23–44 A developing red blood cell (erythroblast). The cell is shown extruding its nucleus to become an immature erythrocyte (a reticulocyte), which then leaves the bone marrow and passes into the bloodstream. The reticulocyte will lose its mitochondria and ribosomes within a day or two to become a mature erythrocyte. Erythrocyte clones develop in the bone marrow on the surface of a macrophage, which phagocytoses and digests the nuclei discarded by the erythroblasts.

cells that express Kit receptor protein, and stromal cells (osteoblasts among them) that express Kit ligand.

Factors That Regulate Hemopoiesis Can Be Analyzed in Culture

While stem cells depend on contact with stromal cells for long-term maintenance, their committed progeny do not, or at least not to the same degree. Thus, dispersed bone marrow hemopoietic cells can be cultured in a semisolid matrix of dilute agar or methylcellulose, and factors derived from other cells can be added artificially to the medium. Because cells in the semisolid matrix cannot migrate, the progeny of each isolated precursor cell remain together as an easily distinguishable colony. A single committed neutrophil progenitor, for example, may give rise to a clone of thousands of neutrophils. Such culture systems have provided a way to assay for the factors that support hemopoiesis and hence to purify them and explore their actions. These substances are glycoproteins and are usually called **colony-stimulating factors** (**CSFs**). Of the growing number of CSFs that have been defined and purified, some circulate in the blood and act as hormones, while others act in the bone marrow either as secreted local mediators or, like Kit ligand, as membrane-bound signals that act through cell–cell contact. The best understood of the CSFs that act as hormones is the glycoprotein erythropoietin, which is produced in the kidneys and regulates *erythropoiesis*, the formation of red blood cells.

Erythropoiesis Depends on the Hormone Erythropoietin

The erythrocyte is by far the most common type of cell in the blood (see Table 23–1). When mature, it is packed full of hemoglobin and contains hardly any of the usual cell organelles. In an erythrocyte of an adult mammal, even the nucleus, endoplasmic reticulum, mitochondria, and ribosomes are absent, having been extruded from the cell in the course of its development (**Figure 23–44**). The erythrocyte therefore cannot grow or divide; the only possible way of making more erythrocytes is by means of stem cells. Furthermore, erythrocytes have a limited life-span—about 120 days in humans or 55 days in mice. Worn-out erythrocytes are phagocytosed and digested by macrophages in the liver and spleen, which remove more than 10^{11} senescent erythrocytes in each of us each day. Young erythrocytes actively protect themselves from this fate: they have a protein on their surface that binds to an inhibitory receptor on macrophages and thereby prevents their phagocytosis.

A lack of oxygen or a shortage of erythrocytes stimulates specialized cells in the kidney to synthesize and secrete increased amounts of **erythropoietin** into the bloodstream. The erythropoietin, in turn, stimulates the production of more erythrocytes. Since a change in the rate of release of new erythrocytes into the bloodstream is observed as early as 1–2 days after an increase in erythropoietin levels in the bloodstream, the hormone must act on cells that are very close precursors of the mature erythrocytes.

The cells that respond to erythropoietin can be identified by culturing bone marrow cells in a semisolid matrix in the presence of erythropoietin. In a few days, colonies of about 60 erythrocytes appear, each founded by a single committed erythroid progenitor cell. This progenitor depends on erythropoietin for its survival as well as its proliferation. It does not yet contain hemoglobin, and it is derived from an earlier type of committed erythroid progenitor that does not depend on erythropoietin.

Table 23–2 Some Colony-stimulating Factors (CSFs) That Influence Blood Cell Formation

FACTOR	TARGET CELLS	PRODUCING CELLS	RECEPTORS
Erythropoietin	CFC-E	kidney cells	cytokine family
Interleukin 3 (IL3)	multipotent stem cell, most progenitor cells, many terminally differentiated cells	T lymphocytes, epidermal cells	cytokine family
Granulocyte/macrophage CSF (GMCSF)	GM progenitor cells	T lymphocytes, endothelial cells, fibroblasts	cytokine family
Granulocyte CSF (GCSF)	GM progenitor cells and neutrophils	macrophages, fibroblasts	cytokine family
Macrophage CSF (MCSF)	GM progenitor cells and macrophages	fibroblasts, macrophages, endothelial cells	receptor tyrosine kinase family
Kit ligand	hemopoietic stem cells	stromal cells in bone marrow and many other cells	receptor tyrosine kinase family

A second CSF, called **interleukin-3** (**IL3**), promotes the survival and proliferation of the earlier erythroid progenitor cells. In its presence, much larger erythroid colonies, each comprising up to 5000 erythrocytes, develop from cultured bone marrow cells in a process requiring a week or 10 days. Evidently the descendants of the hemopoietic stem cells, after they have become committed to an erythroid fate, have to step their way through a further long program of cell divisions, changing their character and their dependence on environmental signals as they progress toward the final differentiated state.

Multiple CSFs Influence Neutrophil and Macrophage Production

The two classes of cells dedicated to phagocytosis, neutrophils and macrophages, develop from a common progenitor cell called a **granulocyte/macrophage** (**GM**) **progenitor cell**. Like the other granulocytes (eosinophils and basophils), neutrophils circulate in the blood for only a few hours before migrating out of capillaries into the connective tissues or other specific sites, where they survive for only a few days. They then die by apoptosis and are phagocytosed by macrophages. Macrophages, in contrast, can persist for months or perhaps even years outside the bloodstream, where they can be activated by local signals to resume proliferation.

At least seven distinct CSFs that stimulate neutrophil and macrophage colony formation in culture have been defined, and some or all of these are thought to act in different combinations to regulate the selective production of these cells *in vivo*. These CSFs are synthesized by various cell types—including endothelial cells, fibroblasts, macrophages, and lymphocytes—and their concentration in the blood typically increases rapidly in response to bacterial infection in a tissue, thereby increasing the number of phagocytic cells released from the bone marrow into the bloodstream. IL3 is one of the least specific of the factors, acting on multipotent stem cells as well as on most classes of committed progenitor cells, including GM progenitor cells. Various other factors act more selectively on committed GM progenitor cells and their differentiated progeny (**Table 23–2**), although in many cases they act on certain other branches of the hemopoietic family tree as well.

All of these CSFs, like erythropoietin, are glycoproteins that act at low concentrations (about 10^{-12} M) by binding to specific cell-surface receptors, as discussed in Chapter 15. A few of these receptors are transmembrane tyrosine kinases but most belong to the large cytokine receptor family, whose members are usually composed of two or more subunits, one of which is frequently shared among several receptor types (**Figure 23–45**). The CSFs not only operate on the precursor cells to promote the production of differentiated progeny, they also activate the specialized functions (such as phagocytosis and target-cell killing) of the terminally differentiated cells. Proteins produced artificially from the cloned genes for these factors are strong stimulators of hemopoiesis in experimental animals. They are now widely used in human patients to stimulate the

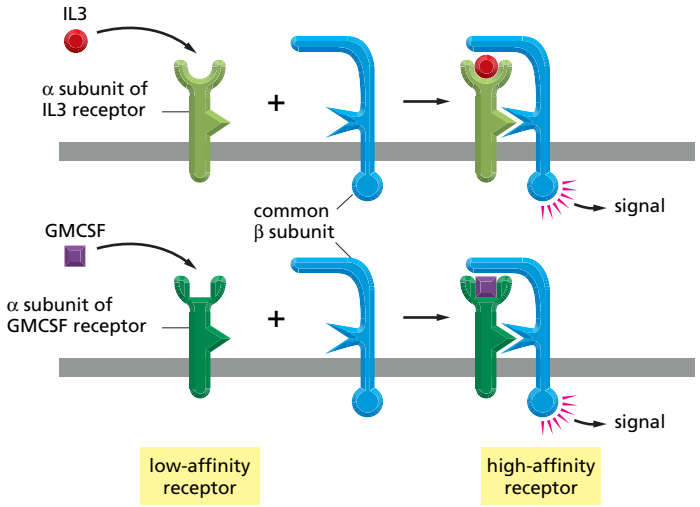


Figure 23–45 Sharing of subunits among CSF receptors. Human IL3 receptors and GMCSF receptors have different α subunits and a common β subunit. Their ligands are thought to bind to the free α subunit with low affinity, and this triggers the assembly of the heterodimer that binds the ligand with high affinity.

regeneration of hemopoietic tissue and to boost resistance to infection—an impressive demonstration of how basic cell biological research and animal experiments can lead to better medical treatment.

The Behavior of a Hemopoietic Cell Depends Partly on Chance

CSFs are defined as factors that promote the production of colonies of differentiated blood cells. But precisely what effect does a CSF have on an individual hemopoietic cell? The factor might control the rate of cell division or the number of division cycles that the progenitor cell undergoes before differentiating; it might act late in the hemopoietic lineage to facilitate differentiation; it might act early to influence commitment; or it might simply increase the probability of cell survival (**Figure 23–46**). By monitoring the fate of isolated individual hemopoietic cells in culture, it has been possible to show that a single CSF, such as GMCSF, can exert all these effects, although it is still not clear which are most important *in vivo*.

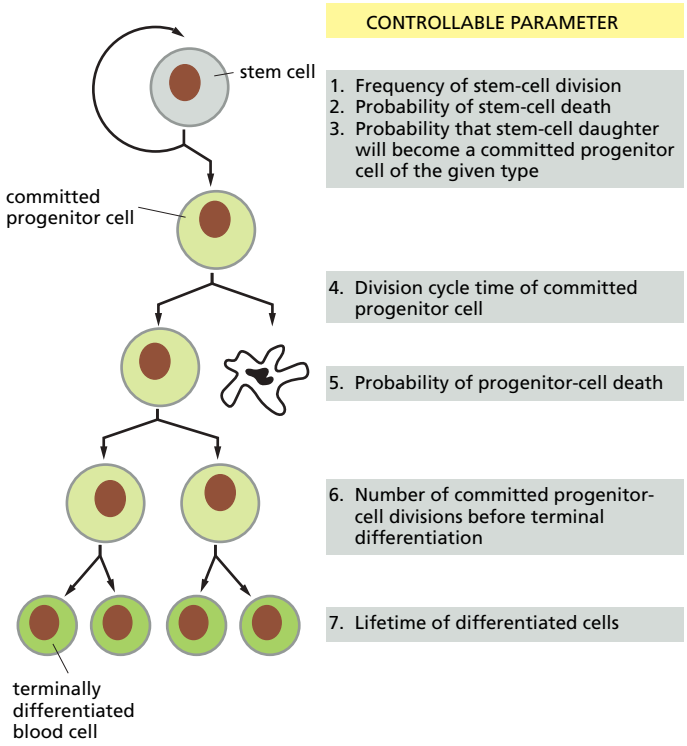


Figure 23–46 Some of the parameters through which the production of blood cells of a specific type might be regulated. Studies in culture suggest that colony-stimulating factors (CSFs) can affect all of these aspects of hemopoiesis.

Studies *in vitro* indicate, moreover, that there is a large element of chance in the way a hemopoietic cell behaves—a reflection, presumably, of "noise" in the genetic control system, as discussed in Chapter 7. At least some of the CSFs seem to act by regulating probabilities, not by dictating directly what the cell shall do. In hemopoietic cell cultures, even if the cells have been selected to be as homogeneous a population as possible, there is a remarkable variability in the sizes and often in the characters of the colonies that develop. And if two sister cells are taken immediately after a cell division and cultured apart under identical conditions, they frequently give rise to colonies that contain different types of blood cells or the same types of blood cells in different numbers. Thus, both the programming of cell division and the process of commitment to a particular path of differentiation seem to involve random events at the level of the individual cell, even though the behavior of the multicellular system as a whole is regulated in a reliable way. The sequence of cell fate restrictions shown in Figure 23–42 conveys the impression of a program executed with computer-like logic and precision. Individual cells may be more quirky and erratic, and may sometimes progress by other decision pathways from the stem-cell state toward terminal differentiation.

Regulation of Cell Survival Is as Important as Regulation of Cell Proliferation

The default behavior of hemopoietic cells in the absence of CSFs is death by apoptosis (discussed in Chapter 18). Thus, in principle, the CSFs could regulate the numbers of the various types of blood cells entirely through selective control of cell survival in this way. There is evidence that the control of cell survival does indeed play a central part in regulating the numbers of blood cells, just as it does for hepatocytes and many other cell types, as we have already seen. The amount of apoptosis in the vertebrate hemopoietic system is enormous: billions of neutrophils die in this way each day in an adult human, for example. In fact, most neutrophils produced in the bone marrow die there without ever functioning. This futile cycle of production and destruction presumably serves to maintain a reserve supply of cells that can be promptly mobilized to fight infection whenever it flares up, or phagocytosed and digested for recycling when all is quiet. Compared with the life of the organism, the lives of cells are cheap.

Too little cell death can be as dangerous to the health of a multicellular organism as too much proliferation. In the hemopoietic system, mutations that inhibit cell death by causing excessive production of the intracellular apoptosis inhibitor Bcl2 promote the development of cancer in B lymphocytes. Indeed, the capacity for unlimited self-renewal is a dangerous property for any cell to possess, and many cases of leukemia arise through mutations that confer this capacity on committed hemopoietic precursor cells that would normally be fated to differentiate and die after a limited number of division cycles.

Summary

The many types of blood cells, including erythrocytes, lymphocytes, granulocytes, and macrophages, all derive from a common multipotent stem cell. In the adult, hemopoietic stem cells are found mainly in bone marrow, and they depend on signals from the marrow stromal (connective-tissue) cells, especially osteoblasts, to maintain their stem-cell character. As in some other stem-cell systems, the Wnt signaling pathway appears to be critical for stem-cell maintenance, though it is not the only one involved. The stem cells normally divide infrequently to produce more stem cells (self-renewal) and various committed progenitor cells (transit amplifying cells), each able to give rise to only one or a few types of blood cells. The committed progenitor cells divide extensively under the influence of various protein signal molecules (colony-stimulating factors, or CSFs) and then terminally differentiate into mature blood cells, which usually die after several days or weeks.

Studies of hemopoiesis have been greatly aided by in vitro *assays in which stem cells or committed progenitor cells form clonal colonies when cultured in a semisolid*

matrix. The progeny of stem cells seem to make their choices between alternative developmental pathways in a partly random manner. Cell death by apoptosis,controlled by the availability of CSFs, also plays a central part in regulating the numbers of mature differentiated blood cells.

GENESIS, MODULATION, AND REGENERATION OF SKELETAL MUSCLE

The term "muscle" includes many cell types, all specialized for contraction but in other respects dissimilar. As noted in Chapter 16, all eucaryotic cells possess a contractile system involving actin and myosin, but muscle cells have developed this apparatus to a high degree. Mammals possess four main categories of cells specialized for contraction: skeletal muscle cells, heart (cardiac) muscle cells, smooth muscle cells, and myoepithelial cells (**Figure 23–47**). These differ in function, structure, and development. Although all of them generate contractile forces by using organized filament systems based on actin and myosin, the actin and myosin molecules employed have somewhat different amino acid sequences, are differently arranged in the cell, and are associated with different sets of proteins to control contraction.

Skeletal muscle cells are responsible for practically all movements that are under voluntary control. These cells can be very large (2–3 cm long and 100 μm

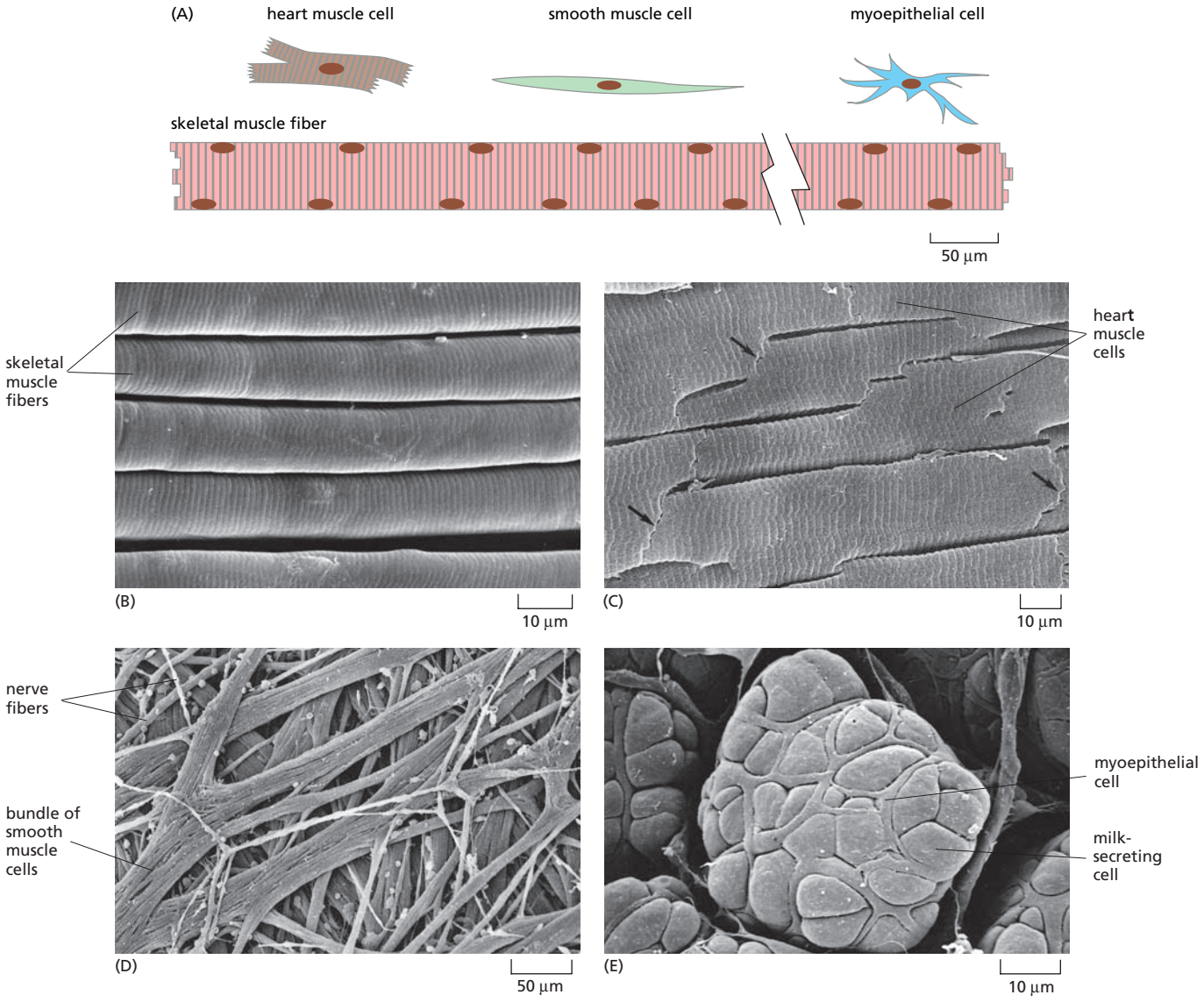


Figure 23–47 The four classes of muscle cells of a mammal. (A) Schematic drawings (to scale). (B–E) Scanning electron micrographs, showing (B) skeletal muscle from the neck of a hamster, (C) heart muscle from a rat, (D) smooth muscle from the urinary bladder of a guinea pig, and (E) myoepithelial cells in a secretory alveolus from a lactating rat mammary gland. The *arrows* in (C) point to intercalated discs—end-to-end junctions between the heart muscle cells; skeletal muscle cells in long muscles are joined end to end in a similar way. Note that the smooth muscle is shown at a lower magnification than the others. (B, courtesy of Junzo Desaki; C, from T. Fujiwara, in Cardiac Muscle in Handbook of Microscopic Anatomy [E.D. Canal, ed.]. Berlin: Springer-Verlag, 1986; D, courtesy of Satoshi Nakasiro; E, from T. Nagato et al., *Cell Tiss. Res.* 209:1–10, 1980. With permission from Springer-Verlag.)

in diameter in an adult human) and are often called *muscle fibers* because of their highly elongated shape. Each one is a syncytium, containing many nuclei within a common cytoplasm. The other types of muscle cells are more conventional, generally having only a single nucleus. **Heart muscle cells** resemble skeletal muscle fibers in that their actin and myosin filaments are aligned in very orderly arrays to form a series of contractile units called *sarcomeres*, so that the cells have a striated (striped) appearance. **Smooth muscle cells** are so named because they do not appear striated. The functions of smooth muscle vary greatly, from propelling food along the digestive tract to erecting hairs in response to cold or fear. **Myoepithelial cells** also have no striations, but unlike all other muscle cells they lie in epithelia and are derived from the ectoderm. They form the dilator muscle of the eye's iris and serve to expel saliva, sweat, and milk from the corresponding glands, as discussed earlier (see Figure 23–11). The four main categories of muscle cells can be further divided into distinctive subtypes, each with its own characteristic features.

The mechanisms of muscle contraction are discussed in Chapter 16. Here we consider how muscle tissue is generated and maintained. We focus on the skeletal muscle fiber, which has a curious mode of development, a striking ability to modulate its differentiated character, and an unusual strategy for repair.

Myoblasts Fuse to Form New Skeletal Muscle Fibers

Chapter 22 described how certain cells, originating from the somites of a vertebrate embryo at a very early stage, become determined as **myoblasts**, the precursors of skeletal muscle fibers. The commitment to be a myoblast depends on gene regulatory proteins of at least two families—a pair of homeodomain proteins called Pax3 and Pax7, and the *MyoD family* of basic helix–loop–helix proteins (discussed in Chapter 7). These act in combination to give the myoblast a memory of its committed state, and, eventually, to regulate the expression of other genes that give the mature muscle cell its specialized character (see Figure 7–75). After a period of proliferation, the myoblasts undergo a dramatic change of state: they stop dividing, switch on the expression of a whole battery of muscle-specific genes required for terminal differentiation, and fuse with one another to form multinucleate skeletal muscle fibers (**Figure 23–48**). Fusion involves specific cell–cell adhesion molecules that mediate recognition between newly differentiating myoblasts and fibers. Once differentiation has occurred, the cells do not divide and the nuclei never again replicate their DNA.

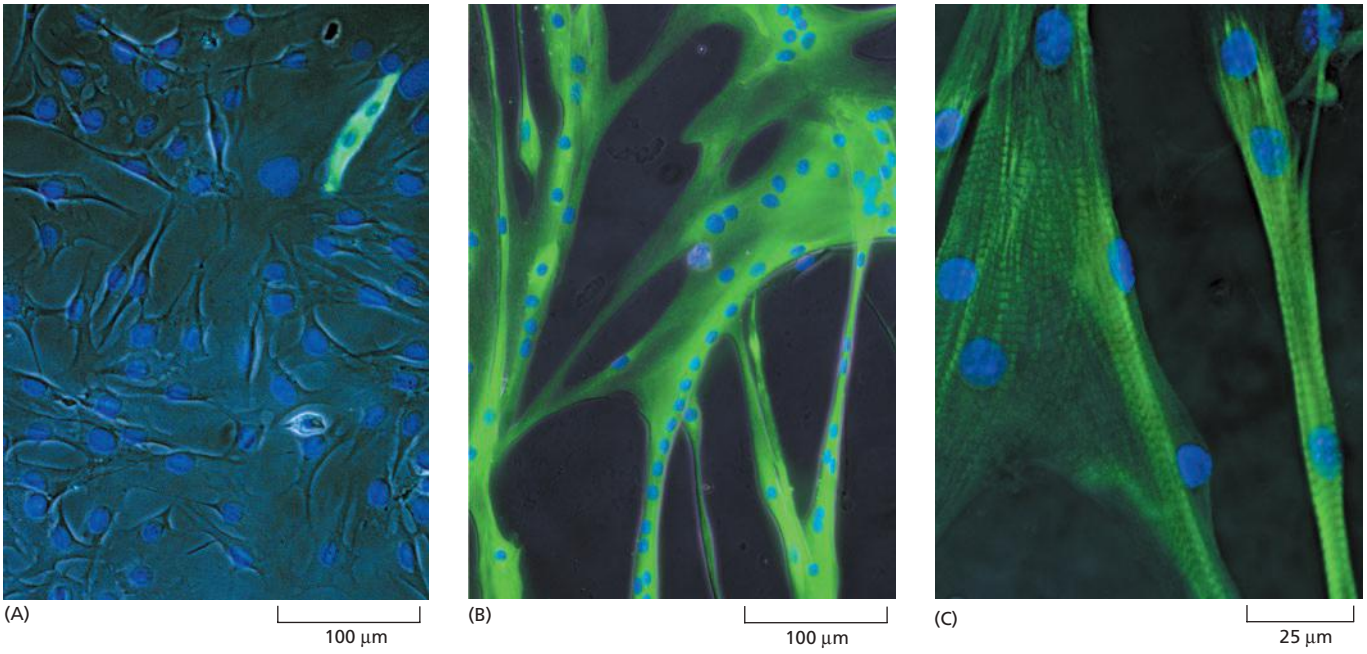


Figure 23–48 Myoblast fusion in culture. The culture is stained with a fluorescent antibody *(green)* against skeletal muscle myosin, which marks differentiated muscle cells, and with a DNA-specific dye *(blue)* to show cell nuclei. (A) A short time after a change to a culture medium that favors differentiation, just two of the many myoblasts in the field of view have switched on myosin production and have fused to form a muscle cell with two nuclei *(upper right)*. (B) Somewhat later, almost all the cells have differentiated and fused. (C) High-magnification view, showing characteristic striations (fine transverse stripes) in two of the multinucleate muscle cells. (Courtesy of Jacqueline Gross and Terence Partridge.)

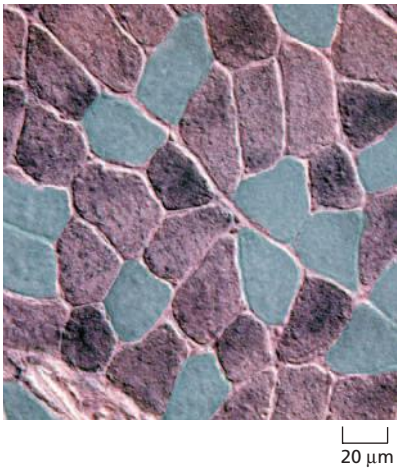


Figure 23–49 Fast and slow muscle fibers. Two consecutive cross sections of the same piece of adult mouse leg muscle were stained with different antibodies, each specific for a different isoform of myosin heavy chain protein, and images of the two sections were overlaid in false color to show the pattern of muscle fiber types. Fibers stained with antibodies against "fast" myosin *(gray)* are specialized to produce fast-twitch contractions; fibers stained with antibodies against "slow" myosin *(pink)* are specialized to produce slow, sustained contractions. The fast-twitch fibers are known as white muscle fibers because they contain relatively little of the colored oxygen-binding protein myoglobin. The slow muscle fibers are called red muscle fibers because they contain much more of it. (Courtesy of Simon Hughes.)

Myoblasts that have been kept proliferating in culture for as long as two years still retain the ability to differentiate and can fuse to form muscle cells in response to a suitable change in culture conditions. Appropriate signal proteins such as fibroblast or hepatocyte growth factor (FGF or HGF) in the culture medium can maintain myoblasts in the proliferative, undifferentiated state: if these soluble factors are removed, the cells rapidly stop dividing, differentiate, and fuse. The system of controls is complex, however, and attachment to the extracellular matrix is also important for myoblast differentiation. Moreover, the process of differentiation is cooperative: differentiating myoblasts secrete factors that apparently encourage other myoblasts to differentiate.

Muscle Cells Can Vary Their Properties by Changing the Protein Isoforms They Contain

Once formed, a skeletal muscle fiber grows, matures, and modulates its character. The genome contains multiple variant copies of the genes encoding many of the characteristic proteins of the skeletal muscle cell, and the RNA transcripts of many of these genes can be spliced in several ways. As a result, muscle fibers produce many variant forms (isoforms) of the proteins of the contractile apparatus. As the muscle fiber matures, it synthesizes different isoforms, satisfying the changing demands for speed, strength, and endurance in the fetus, the newborn, and the adult. Within a single adult muscle, several distinct types of skeletal muscle fibers, each with different sets of protein isoforms and different functional properties, can be found side by side (**Figure 23–49**). The characteristics of the different fiber types are determined partly before birth by the genetic program of development, partly in later life by activity and training. Different classes of motor neurons innervate slow muscle fibers (for sustained contraction) and fast muscle fibers (for rapid twitch), and the innervation can regulate muscle-fiber gene expression and size through the different patterns of electrical stimulation that these neurons deliver.

Skeletal Muscle Fibers Secrete Myostatin to Limit Their Own Growth

A muscle can grow in three ways: its fibers can increase in number, in length, or in girth. Because skeletal muscle fibers are unable to divide, more of them can be made only by the fusion of myoblasts, and the adult number of multinucleated skeletal muscle fibers is in fact attained early—before birth, in humans. Once formed, a skeletal muscle fiber generally survives for the entire lifetime of the animal. However, individual muscle nuclei can be added or lost. The enormous postnatal increase in muscle bulk is achieved by cell enlargement. Growth in length depends on recruitment of more myoblasts into the existing multinucleated fibers, which increases the number of nuclei in each cell. Growth in girth, such as occurs in the muscles of weightlifters, involves both myoblast recruitment and an increase in the size and numbers of the contractile myofibrils that each muscle fiber nucleus supports.

What, then, are the mechanisms that control muscle cell numbers and muscle cell size? One part of the answer lies in an extracellular signal protein called *myostatin*. Mice with a loss-of-function mutation in the myostatin gene have enormous muscles—two to three times larger than normal (**Figure 23–50**). Both the numbers and the size of the muscle cells seem to be increased. Mutations in the same gene are present in so-called "double-muscled" breeds of cattle (see Figure 17–69): in selecting for big muscles, cattle breeders have unwittingly selected for myostatin deficiency. Myostatin belongs to the TGFβ superfamily of signal proteins. It is normally made and secreted by skeletal muscle cells, and it acts powerfully on myoblasts, inhibiting both proliferation and differentiation. Its function, evidently, is to provide negative feedback to limit muscle growth, in adult life as well as during development. The growth of some other organs is similarly controlled by a negative-feedback action of a factor that they themselves produce. We shall encounter another example in a later section.

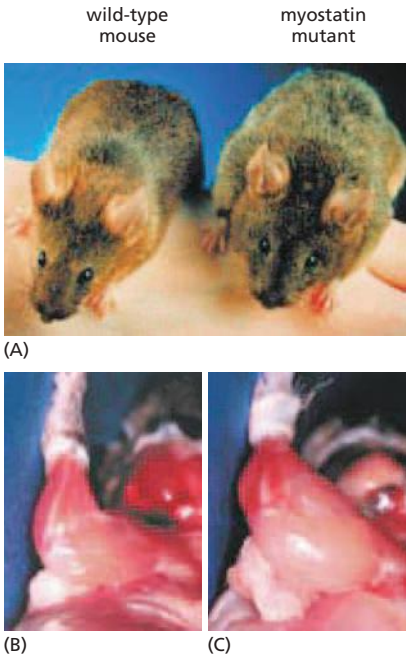


Figure 23–50 Regulation of muscle size by myostatin. (A) A normal mouse compared with a mutant mouse deficient in myostatin. (B) Leg of a normal and (C) of a myostatin-deficient mouse, with skin removed to show the massive enlargement of the musculature in the mutant. (From S.J. Lee and A.C. McPherron, *Curr. Opin. Genet. Devel.* 9:604–607, 1999. With permission from Elsevier.)

Some Myoblasts Persist as Quiescent Stem Cells in the Adult

Even though humans do not normally generate new skeletal muscle fibers in adult life, they still have the capacity to do so, and existing muscle fibers can resume growth when the need arises. Cells capable of serving as myoblasts are retained as small, flattened, and inactive cells lying in close contact with the mature muscle cell and contained within its sheath of basal lamina (**Figure 23–51**). If the muscle is damaged or stimulated to grow, these *satellite cells* are activated to proliferate, and their progeny can fuse to repair the damaged muscle or to allow muscle growth. Like myoblasts, they are regulated by myostatin. Satellite cells, or some subset of the satellite cells, are thus the stem cells of adult skeletal muscle, normally held in reserve in a quiescent state but available when needed as a self-renewing source of terminally differentiated cells. Studies of these cells have provided some of the clearest evidence for the immortal strand hypothesis of asymmetric stem-cell division, as illustrated earlier in Figure 23–10).

The process of muscle repair by means of satellite cells is, nevertheless, limited in what it can achieve. In one form of *muscular dystrophy*, for example, a genetic defect in the cytoskeletal protein dystrophin damages differentiated skeletal muscle cells. As a result, satellite cells proliferate to repair the damaged muscle fibers. This regenerative response is, however, unable to keep pace with the damage, and connective tissue eventually replaces the muscle cells, blocking any further possibility of regeneration. A similar loss of capacity for repair seems to contribute to the weakening of muscle in the elderly.

In muscular dystrophy, where the satellite cells are constantly called upon to proliferate, their capacity to divide may become exhausted as a result of progressive shortening of their telomeres in the course of each cell cycle (discussed in Chapter 17). Stem cells of other tissues seem to be limited in the same way, as we noted earlier in the case of hemopoietic stem cells: they normally divide only at a slow rate, and mutations or exceptional circumstances that cause them to divide more rapidly can lead to premature exhaustion of the stem-cell supply.

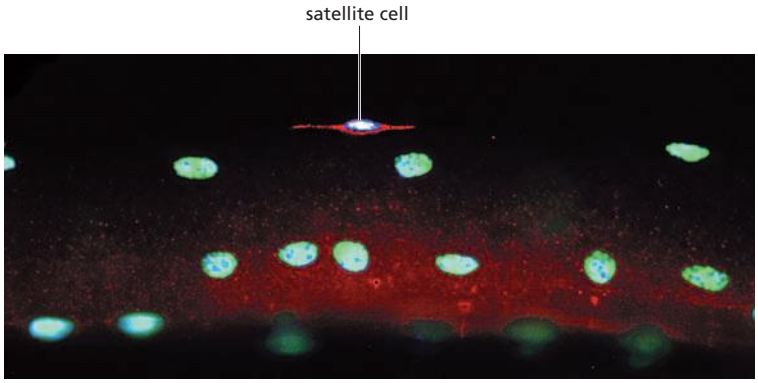


Figure 23–51 A satellite cell on a skeletal muscle fiber. The specimen is stained with an antibody *(red)* against a muscle cadherin, M-cadherin, which is present on both the satellite cell and the muscle fiber and is concentrated at the site where their membranes are in contact. The nuclei of the muscle fiber are stained *green,* and the nucleus of the satellite cell is stained *blue.* (Courtesy of Terence Partridge.)

Summary

Skeletal muscle fibers are one of four main categories of vertebrate cells specialized for contraction, and they are responsible for all voluntary movement. Each skeletal muscle fiber is a syncytium and develops by the fusion of many myoblasts. Myoblasts proliferate extensively, but once they have fused, they can no longer divide. Fusion generally follows the onset of myoblast differentiation, in which many genes encoding muscle-specific proteins are switched on coordinately. Some myoblasts persist in a quiescent state as satellite cells in adult muscle; when a muscle is damaged, these cells are reactivated to proliferate and to fuse to replace the muscle cells that have been lost. They are the stem cells of skeletal muscle. Muscle bulk is regulated homeostatically by a negative-feedback mechanism, in which existing muscle secretes myostatin, which inhibits further muscle growth.

FIBROBLASTS AND THEIR TRANSFORMATIONS: THE CONNECTIVE-TISSUE CELL FAMILY

Many of the differentiated cells in the adult body can be grouped into families whose members are closely related by origin and by character. An important example is the family of **connective-tissue cells**, whose members are not only related but also unusually interconvertible. The family includes *fibroblasts*, *cartilage cells*, and *bone cells*, all of which are specialized for the secretion of collagenous extracellular matrix and are jointly responsible for the architectural framework of the body. The connective-tissue family also includes *fat cells* and *smooth muscle cells*. **Figure 23–52** illustrates these cell types and the interconversions that are thought to occur between them. Connective-tissue cells contribute to the support and repair of almost every tissue and organ, and the adaptability of their differentiated character is an important feature of the responses to many types of damage.

Fibroblasts Change Their Character in Response to Chemical Signals

Fibroblasts seem to be the least specialized cells in the connective-tissue family. They are dispersed in connective tissue throughout the body, where they secrete a nonrigid extracellular matrix that is rich in type I or type III collagen, or both, as discussed in Chapter 19. When a tissue is injured, the fibroblasts nearby proliferate, migrate into the wound <TGAT>, and produce large amounts of collagenous matrix, which helps to isolate and repair the damaged tissue. Their ability to thrive in the face of injury, together with their solitary lifestyle, may explain why fibroblasts are the easiest of cells to grow in culture—a feature that has made them a favorite subject for cell biological studies (**Figure 23–53**).

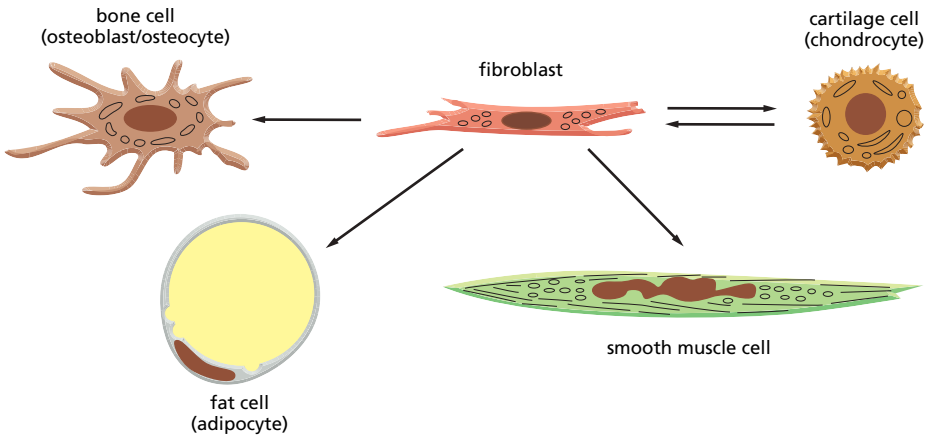


Figure 23–52 The family of connective-tissue cells. *Arrows* show the interconversions that are thought to occur within the family. For simplicity, the fibroblast is shown as a single cell type, but it is uncertain how many types of fibroblasts exist in fact and whether the differentiation potential of different types is restricted in different ways.

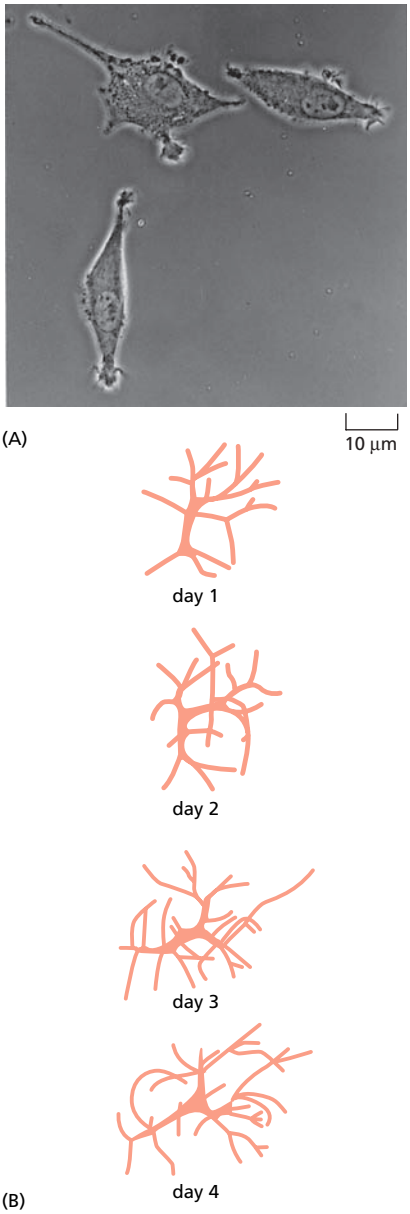


Figure 23–53 The fibroblast. (A) A phase-contrast micrograph of fibroblasts in culture. (B) These drawings of a living fibroblastlike cell in the transparent tail of a tadpole show the changes in its shape and position on successive days. Note that while fibroblasts flatten out in culture, they can have more complex, process-bearing morphologies in tissues. See also Figure 19–54. (A, from E. Pokorna et al., *Cell Motil. Cytoskeleton* 28:25–33, 1994; B, redrawn from E. Clark, *Am. J. Anat.* 13:351–379, 1912. Both with permission from Wiley-Liss.)

As indicated in Figure 23–52, fibroblasts also seem to be the most versatile of connective-tissue cells, displaying a remarkable capacity to differentiate into other members of the family. There are uncertainties about their interconversions, however. Fibroblasts in different parts of the body are intrinsically different, and there may be differences between them even in a single region. "Mature" fibroblasts with a lesser capacity for transformation may, for example, exist side by side with "immature" fibroblasts (often called mesenchymal cells) that can develop into a variety of mature cell types.

The stromal cells of bone marrow, mentioned earlier, provide a good example of connective-tissue versatility. These cells, which can be regarded as a kind of fibroblast, can be isolated from the bone marrow and propagated in culture. Large clones of progeny can be generated in this way from single ancestral stromal cells. According to the signal proteins that are added to the culture medium, the members of such a clone can either continue proliferating to produce more cells of the same type, or can differentiate as fat cells, cartilage cells, or bone cells. Because of their self-renewing, multipotent character, they are referred to as *mesenchymal stem cells*.

Fibroblasts from the dermal layer of the skin are different. When placed in the same culture conditions, they do not show the same plasticity. Yet they, too, can be induced to change their character. At a healing wound, for example, they change their actin gene expression and take on some of the contractile properties of smooth muscle cells, thereby helping to pull the wound margins together; such cells are called *myofibroblasts*. More dramatically, if a preparation of bone matrix, made by grinding bone into a fine powder and dissolving away the hard mineral component, is implanted in the dermal layer of the skin, some of the cells there (probably fibroblasts) become transformed into cartilage cells, and a little later, others transform into bone cells, thereby creating a small lump of bone. These experiments suggest that components in the extracellular matrix can dramatically influence the differentiation of connective-tissue cells.

We shall see that similar cell transformations occur in the natural repair of broken bones. In fact, bone matrix contains high concentrations of several signal proteins that can affect the behavior of connective-tissue cells. These include members of the TGFβ superfamily, including BMPs and TGFβ itself. These factors regulate growth, differentiation, and matrix synthesis by connective-tissue cells, exerting a variety of actions depending on the target cell type and the combination of other factors and matrix components that are present. When injected into a living animal, they can induce the formation of cartilage, bone, or fibrous matrix, according to the site and circumstances of injection. TGFβ is especially important in wound healing, where it stimulates the conversion of fibroblasts into myofibroblasts and promotes the formation of the collagen-rich scar tissue that gives a healed wound its strength.

The Extracellular Matrix May Influence Connective-Tissue Cell Differentiation by Affecting Cell Shape and Attachment

The extracellular matrix may influence the differentiated state of connective-tissue cells through physical as well as chemical effects. This has been shown in studies on cultured cartilage cells, or **chondrocytes**. Under appropriate culture conditions, these cells proliferate and maintain their differentiated character, continuing for many cell generations to synthesize large quantities of highly distinctive cartilage matrix, with which they surround themselves. If, however, the cells are kept at relatively low density and remain as a monolayer on the culture

dish, a transformation occurs. They lose their characteristic rounded shape, flatten down on the substratum, and stop making cartilage matrix: they stop producing type II collagen, which is characteristic of cartilage, and start producing type I collagen, which is characteristic of fibroblasts. By the end of a month in culture, almost all the cartilage cells have switched their collagen gene expression and taken on the appearance of fibroblasts. The biochemical change must occur abruptly, since very few cells are observed to make both types of collagen simultaneously.

The biochemical change seems to be induced, at least in part, by the change in cell shape and attachment. Cartilage cells that have made the transition to a fibroblast-like character, for example, can be gently detached from the culture dish and transferred to a dish of agarose. By forming a gel around them, the agarose holds the cells suspended without any attachment to a substratum, forcing them to adopt a rounded shape. In these circumstances, the cells promptly revert to the character of chondrocytes and start making type II collagen again. Cell shape and anchorage may control gene expression through intracellular signals generated at focal contacts by integrins acting as matrix receptors, as discussed in Chapter 19.

For most types of cells, and especially for a connective-tissue cell, the opportunities for anchorage and attachment depend on the surrounding matrix, which is usually made by the cell itself. Thus, a cell can create an environment that then acts back on the cell to reinforce its differentiated state. Furthermore, the extracellular matrix that a cell secretes forms part of the environment for its neighbors as well as for the cell itself, and thus tends to make neighboring cells differentiate in the same way. A group of chondrocytes forming a nodule of cartilage, for example, either in the developing body or in a culture dish, can be seen to enlarge by the conversion of neighboring fibroblasts into chondrocytes.

Osteoblasts Make Bone Matrix

Cartilage and bone are tissues of very different character; but they are closely related in origin, and the formation of the skeleton depends on an intimate partnership between them.

Cartilage tissue is structurally simple, consisting of cells of a single type—chondrocytes—embedded in a more or less uniform highly hydrated matrix consisting of proteoglycans and type II collagen, whose remarkable properties we have already discussed in Chapter 19. The cartilage matrix is deformable, and the tissue grows by expanding as the chondrocytes divide and secrete more matrix (**Figure 23–54**). **Bone**, by contrast, is dense and rigid; it grows by apposition—that is, by deposition of additional matrix on free surfaces. Like reinforced concrete, the bone matrix is predominantly a mixture of tough fibers (type I collagen fibrils), which resist pulling forces, and solid particles (calcium phosphate as *hydroxylapatite* crystals), which resist compression. The collagen fibrils in adult bone are arranged in regular plywoodlike layers, with the fibrils in each layer lying parallel to one another but at right angles to the fibrils in the layers on either side. They occupy a volume nearly equal to that occupied by the calcium phosphate. The bone matrix is secreted by **osteoblasts** that lie at the surface of the existing matrix and deposit fresh layers of bone onto it. Some of the osteoblasts remain free at the surface, while others gradually become embedded in their own secretion. This freshly formed material (consisting chiefly of type I collagen) is

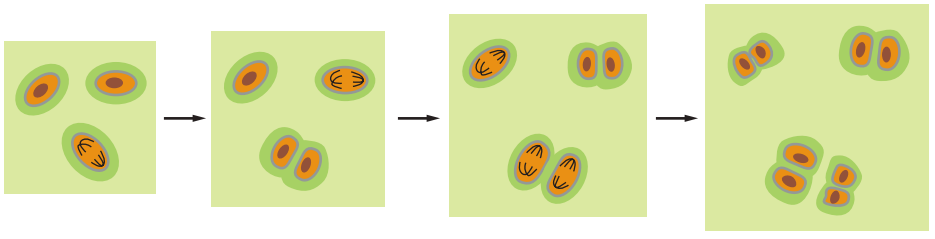


Figure 23–54 The growth of cartilage. The tissue expands as the chondrocytes divide and make more matrix. The freshly synthesized matrix with which each cell surrounds itself is shaded *dark green*. Cartilage may also grow by recruiting fibroblasts from the surrounding tissue and converting them into chondrocytes.

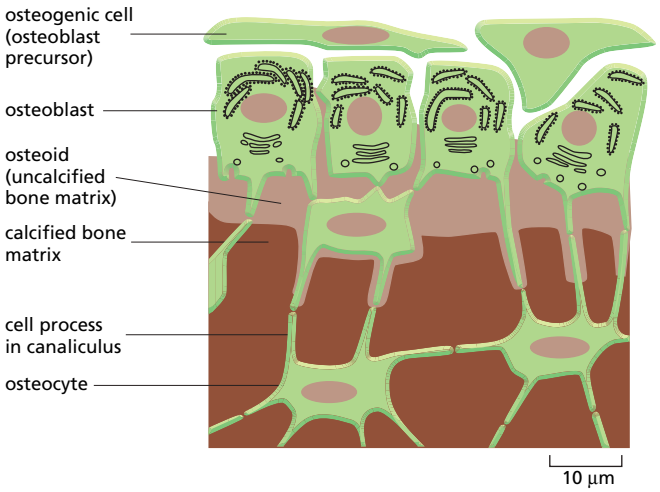


Figure 23–55 Deposition of bone matrix by osteoblasts. Osteoblasts lining the surface of bone secrete the organic matrix of bone (osteoid) and are converted into osteocytes as they become embedded in this matrix. The matrix calcifies soon after it has been deposited. The osteoblasts themselves are thought to derive from osteogenic stem cells that are closely related to fibroblasts.

called *osteoid*. It is rapidly converted into hard bone matrix by the deposition of calcium phosphate crystals in it. Once imprisoned in hard matrix, the original bone-forming cell, now called an **osteocyte**, has no opportunity to divide, although it continues to secrete further matrix in small quantities around itself. The osteocyte, like the chondrocyte, occupies a small cavity, or *lacuna*, in the matrix, but unlike the chondrocyte it is not isolated from its fellows. Tiny channels, or *canaliculi*, radiate from each lacuna and contain cell processes from the resident osteocyte, enabling it to form gap junctions with adjacent osteocytes (**Figure 23–55**). Although the networks of osteocytes do not themselves secrete or erode substantial quantities of matrix, they probably play a part in controlling the activities of the cells that do. Blood vessels and nerves run through the tissue, keeping the bone cells alive and reacting when the bone is damaged.

A mature bone has a complex and beautiful architecture, in which dense plates of *compact bone* tissue enclose spaces spanned by light frameworks of *trabecular bone*—a filigree of delicate shafts and flying buttresses of bone tissue, with soft marrow in the interstices (**Figure 23–56**). The creation, maintenance, and repair of this structure depend not only on the cells of the connective-tissue family that synthesize matrix, but also on a separate class of cells called *osteoclasts* that degrade it, as we shall discuss below.

Most Bones Are Built Around Cartilage Models

Most bones, and in particular the long bones of the limbs and trunk, originate from minute "scale models" formed out of cartilage in the embryo. Each scale model grows, and as new cartilage forms, the older cartilage is replaced by bone. The process is known as *endochondral* bone formation. Cartilage growth and erosion and bone deposition are so ingeniously coordinated that the adult bone, though it may be half a meter long, is almost the same shape as the initial cartilaginous model, which was no more than a few millimeters long.

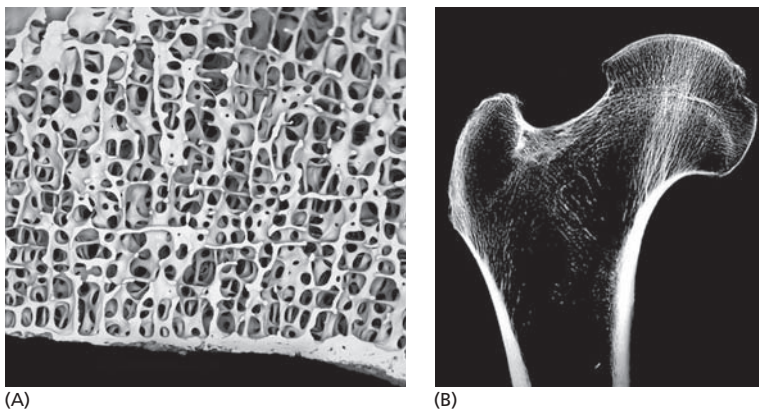

(A) (B)

Figure 23–56 Trabecular and compact bone. (A) Low-magnification scanning electron micrograph of trabecular bone in a vertebra of an adult man. The soft marrow tissue has been dissolved away. (B) A slice through the head of the femur, with bone marrow and other soft tissue likewise dissolved away, reveals the compact bone of the shaft and the trabecular bone in the interior. Because of the way in which bone tissue remodels itself in response to mechanical load, the trabeculae become oriented along the principle axes of stress within the bone. (A, courtesy of Alan Boyde; B, from J.B. Kerr, Atlas of Functional Histology. Mosby, 1999.)

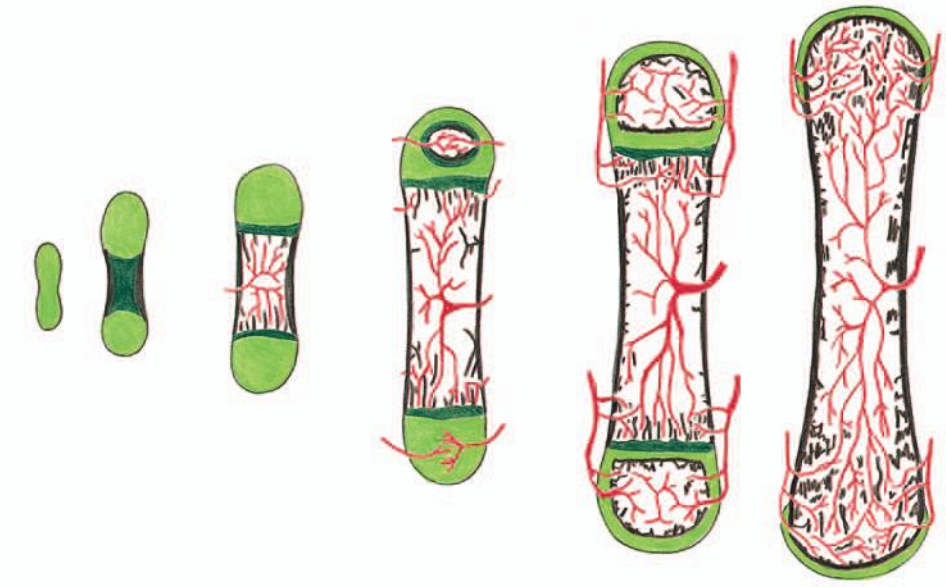


Figure 23–57 The development of a long bone. Long bones, such as the femur or the humerus, develop from a miniature cartilage model. Uncalcified cartilage is shown in *light green,* calcified cartilage in *dark green,* bone in *black,* and blood vessels in *red*. The cartilage is not converted to bone but is gradually replaced by it through the action of osteoclasts and osteoblasts, which invade the cartilage in association with blood vessels. Osteoclasts erode cartilage and bone matrix, while osteoblasts secrete bone matrix. The process of ossification begins in the embryo and is not completed until the end of puberty. The resulting bone consists of a thick-walled hollow cylinder of compact bone enclosing a large central cavity occupied by the bone marrow. Note that not all bones develop in this way. The membrane bones of the skull, for example, are formed directly as bony plates, not from a prior cartilage model. (Adapted from D.W. Fawcett, A Textbook of Histology, 12th ed. New York: Chapman and Hall, 1994.)

The process begins in the embryo with the appearance of hazily defined "condensations"—groups of embryonic connective tissue cells that become more closely packed than their neighbors and begin to express a characteristic set of genes—including, in particular, *Sox9* and, after a slight delay, *Runx2*. These two genes code for gene regulatory proteins that are critical for cartilage and bone development, respectively. Mutant cells lacking *Sox9* are unable to differentiate as cartilage but can form bone (and in some parts of the body will make bone where cartilage should be). Conversely, animals lacking functional *Runx2* make no bone and are born with a skeleton consisting solely of cartilage.

Soon after expression of *Sox9* has begun, the cells in the core of the condensation begin to secrete cartilage matrix, dividing and enlarging individually as they do so. In this way, they form an expanding rod of cartilage surrounded by more densely packed non-cartilage cells. The cartilage cells in the middle segment of the rod become hypertrophied (grossly enlarged) and cease dividing; and at the same time, they start to secrete Indian Hedgehog—a signal molecule of the Hedgehog family. This in turn provokes increased production of certain Wnt proteins, which activate the Wnt pathway in cells surrounding the cartilage rod. As a result, they switch off expression of *Sox9,* maintain expression of *Runx2*, and begin to differentiate as osteoblasts, creating a collar of bone around the shaft of the cartilage model. Artificial overactivation of the Wnt pathway tips a larger proportion of cells into making bone rather than cartilage; an artificial block in the Wnt signaling pathway does the opposite. In this system, therefore, Wnt signaling controls the choice between alternative paths of differentiation, with *Sox9* expression leading the way toward cartilage, and *Runx2* expression leading the way toward bone.

The hypertrophied cartilage cells in the shaft of the cartilage model soon die, leaving large cavities in the matrix, and the matrix itself becomes mineralized, like bone, by the deposition of calcium phosphate crystals. Osteoclasts and blood vessels invade the cavities and erode the residual cartilage matrix, creating a space for bone marrow, and osteoblasts following in their wake begin to deposit trabecular bone in parts of the cavity where strands of cartilage matrix remain as a template. The cartilage tissue at the ends of the bone is replaced by bone tissue at a much later stage, by a somewhat similar process, as shown in **Figure 23–57**. Continuing elongation of the bone, up to the time of puberty, depends on a plate of growing cartilage between the shaft and the head of the bone. Defective growth of the cartilage in this plate, as a result of a dominant mutation in the gene that codes for an FGF receptor (FGFR3), is responsible for the commonest form of dwarfism, known as *achondroplasia* (**Figure 23–58**).

The cartilage growth plate is eventually replaced by bone and disappears. The only surviving remnant of cartilage in the adult long bone is a thin but

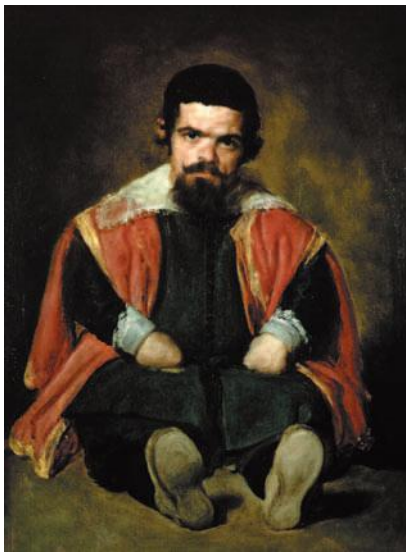


Figure 23–58 Achondroplasia. This type of dwarfism occurs in one of 10,000–100,000 births; in more than 99% of cases it results from a mutation at an identical site in the genome, corresponding to amino acid 380 in the FGF receptor FGFR3 (a glycine in the transmembrane domain). The mutation is dominant, and almost all cases are due to new, independently occurring mutations, implying an extraordinarily high mutation rate at this particular site in the genome. The defect in FGF signaling causes dwarfism by interfering with the growth of cartilage in developing long bones. (From Velasquez's painting of Sebastian de Morra. © Museo del Prado, Madrid.)

important layer that forms a smooth, slippery covering on the bone surfaces at joints, where one bone articulates with another (see Figure 23–57). Erosion of this layer of cartilage, through aging, mechanical damage, or autoimmune attack, leads to *arthritis*, one of the commonest and most painful afflictions of old age.

Bone Is Continually Remodeled by the Cells Within It

For all its rigidity, bone is by no means a permanent and immutable tissue. Running through the hard extracellular matrix are channels and cavities occupied by living cells, which account for about 15% of the weight of compact bone. These cells are engaged in an unceasing process of remodeling: while osteoblasts deposit new bone matrix, osteoclasts demolish old bone matrix. This mechanism provides for continuous turnover and replacement of the matrix in the interior of the bone.

Osteoclasts (**Figure 23–59**) are large multinucleated cells that originate, like macrophages, from hemopoietic stem cells in the bone marrow. The precursor cells are released as monocytes into the bloodstream and collect at sites of bone resorption, where they fuse to form the multinucleated osteoclasts, which cling to surfaces of the bone matrix and eat it away. Osteoclasts are capable of tunneling deep into the substance of compact bone, forming cavities that are then invaded by other cells. A blood capillary grows down the center of such a tunnel,

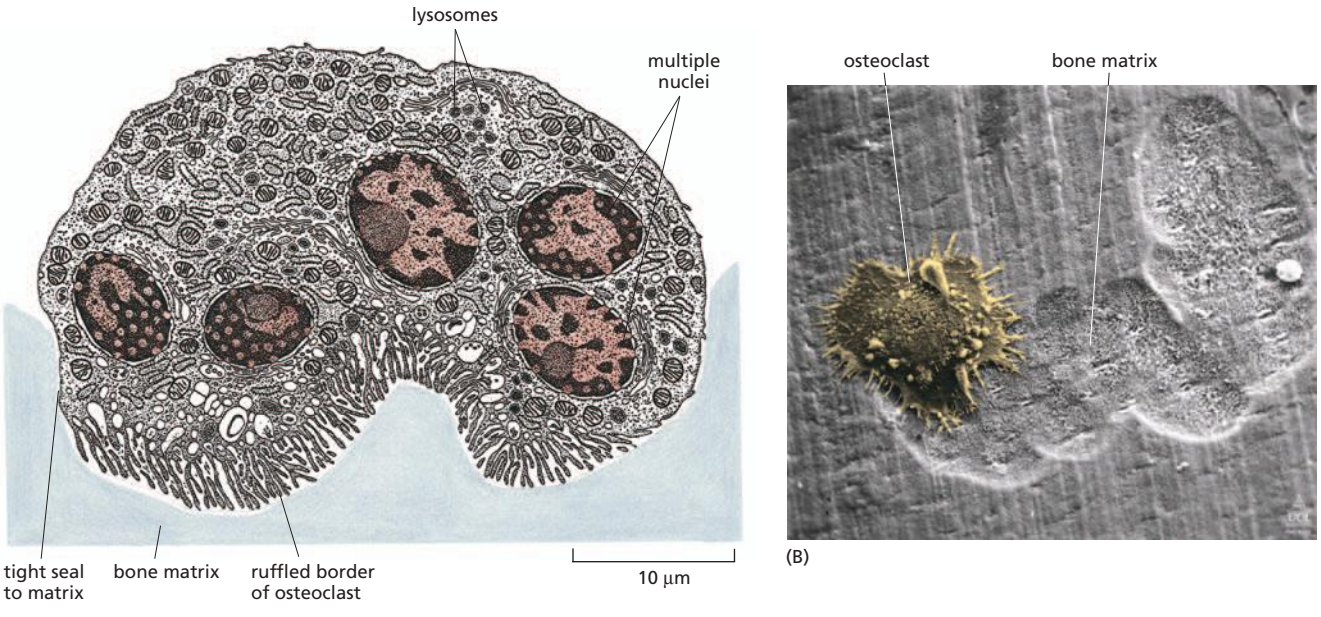


Figure 23–59 Osteoclasts. (A) Drawing of an osteoclast in cross section. This giant, multinucleated cell erodes bone matrix. The "ruffled border" is a site of secretion of acids (to dissolve the bone minerals) and hydrolases (to digest the organic components of the matrix). Osteoclasts vary in shape, are motile, and often send out processes to resorb bone at multiple sites. They develop from monocytes and can be viewed as specialized macrophages. (B) An osteoclast on bone matrix, seen by scanning electron microscopy. The osteoclast has been crawling over the matrix, eating it away, and leaving a trail of pits where it has done so. (A, from R.V. Krstić, Ultrastructure of the Mammalian Cell: An Atlas. Berlin: Springer-Verlag, 1979; B, courtesy of Alan Boyde.)

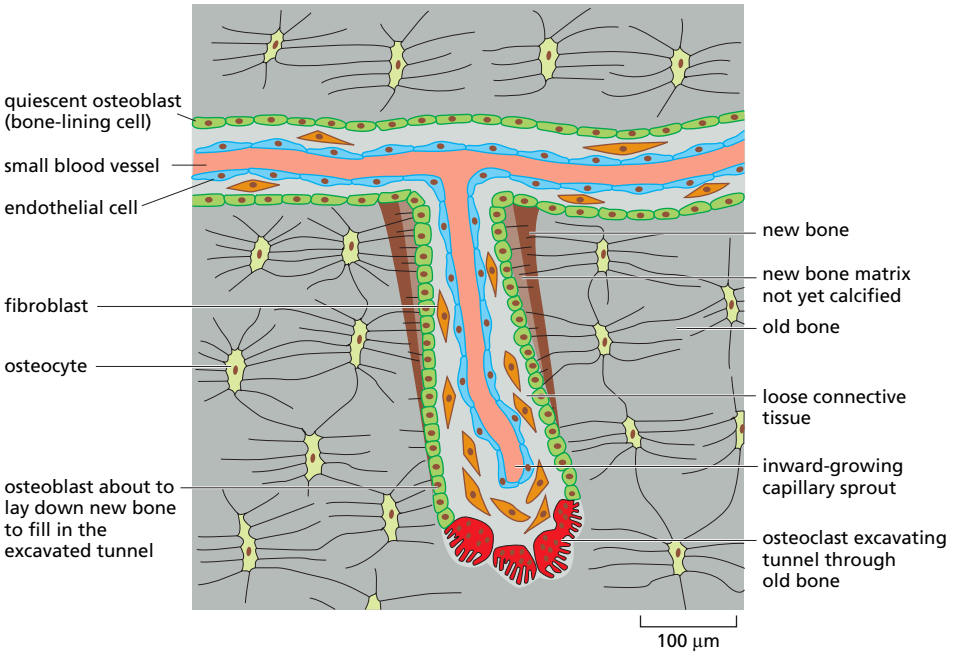


Figure 23–60 The remodeling of compact bone. Osteoclasts acting together in a small group excavate a tunnel through the old bone, advancing at a rate of about 50 μm per day. Osteoblasts enter the tunnel behind them, line its walls, and begin to form new bone, depositing layers of matrix at a rate of 1–2 μm per day. At the same time, a capillary sprouts down the center of the tunnel. The tunnel eventually becomes filled with concentric layers of new bone, with only a narrow central canal remaining. Each such canal, besides providing a route of access for osteoclasts and osteoblasts, contains one or more blood vessels that transport the nutrients the bone cells require for survival. Typically, about 5–10% of the bone in a healthy adult mammal is replaced in this way each year. (After Z.F.G. Jaworski, B. Duck and G. Sekaly, *J. Anat.* 133:397–405, 1981. With permission from Blackwell Publishing.)

and the walls of the tunnel become lined with a layer of osteoblasts (**Figure 23–60**). To produce the plywoodlike structure of compact bone, these osteoblasts lay down concentric layers of new matrix, which gradually fill the cavity, leaving only a narrow canal surrounding the new blood vessel. Many of the osteoblasts become trapped in the bone matrix and survive as concentric rings of osteocytes. At the same time as some tunnels are filling up with bone, others are being bored by osteoclasts, cutting through older concentric systems. The consequences of this perpetual remodeling are beautifully displayed in the layered patterns of matrix observed in compact bone (**Figure 23–61**).

Osteoclasts Are Controlled by Signals From Osteoblasts

The osteoblasts that make the matrix also produce the signals that recruit and activate the osteoclasts to degrade it. Two proteins appear to have this role: one is Macrophage-CSF (MCSF), which we already encountered in our account of hemopoiesis (see Table 23–2); the other is TNF11, a member of the TNF family

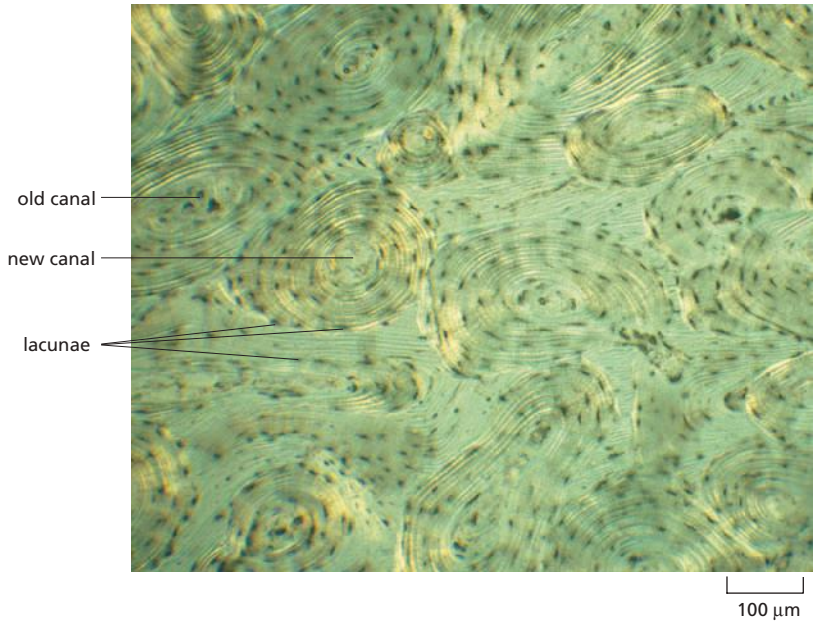


Figure 23–61 A transverse section through a compact outer portion of a long bone. The micrograph shows the outlines of tunnels that have been formed by osteoclasts and then filled in by osteoblasts during successive rounds of bone remodeling. The section has been prepared by grinding. The hard matrix has been preserved, but not the cells. Lacunae and canaliculi that were occupied by osteocytes are clearly visible, however. The alternating bright and dark concentric rings correspond to an alternating orientation of the collagen fibers in the successive layers of bone matrix laid down by the osteoblasts that lined the wall of the canal during life. (This pattern is revealed here by viewing the specimen between partly crossed polarizing filters.) Note how older systems of concentric layers of bone have been partly cut through and replaced by newer systems.

(also called RANKL). The behavior of the osteoblasts in attracting their opponents may seem self-defeating, but it has the useful function of localizing osteoclasts in the tissue where they are needed.

To prevent excessive degradation of matrix, the osteoblasts secrete, along with MCSF and TNF11, another protein, *osteoprotegerin*, that tends to block the action of TNF11. The higher the level of Wnt activation in the osteoblasts, the more osteoproteregin they secrete and, consequently, the lower the level of osteoclast activation and the lower the rate of bone matrix degradation. The Wnt signaling pathway thus seems to have two distinct functions in bone formation: at early stages, it controls the initial commitment of cells to an osteoblast fate; later, it acts in the differentiated osteoblasts to help govern the balance between matrix deposition and matrix erosion.

Disturbance of this balance can lead to *osteoporosis*, where there is excessive erosion of the bone matrix and weakening of the bone, or to the opposite condition, *osteopetrosis,* where the bone becomes excessively thick and dense. Hormonal signals, including estrogen, androgens, and the peptide hormone *leptin*, famous for its role in the control of appetite (discussed below), have powerful effects on this balance. At least some of these effects are mediated through influences on the osteoblasts' production of TNF11 and osteoprotegerin.

Circulating hormones affect bones throughout the body. No less important are local controls that allow bone to be deposited in one place while it is resorbed in another. Through such controls over the process of remodeling, bones are endowed with a remarkable ability to adjust their structure in response to long-term variations in the load imposed on them. It is this that makes orthodontics possible, for example: a steady force applied to a tooth with a brace will cause it to move gradually, over many months, through the bone of the jaw, through remodeling of the bone tissue ahead of it and behind it. The adaptive behavior of bone implies that the deposition and erosion of the matrix are in some way governed by local mechanical stresses (see Figure 23–56). Some evidence suggests that this is because mechanical stress on the bone tissue activates the Wnt pathway in the osteoblasts or osteocytes, thereby regulating their production of the signals that regulate osteoclast activity.

Bone can also undergo much more rapid and dramatic reconstruction when the need arises. Some cells capable of forming new cartilage persist in the connective tissue that surrounds a bone. If the bone is broken, the cells in the neighborhood of the fracture repair it by a sort of recapitulation of the original embryonic process: cartilage is first laid down to bridge the gap and is then replaced by bone. The capacity for self-repair, so strikingly illustrated by the tissues of the skeleton, is a property of living structures that has no parallel among present-day man-made objects.

Fat Cells Can Develop From Fibroblasts

Fat cells, or **adipocytes**, also derive from fibroblastlike cells, both during normal mammalian development and in various pathological circumstances. In muscular dystrophy, for example, where the muscle cells die, they are gradually replaced by fatty connective tissue, probably by conversion of local fibroblasts. Fat-cell differentiation (whether normal or pathological) begins with the expression of two families of gene regulatory proteins: the *CEBP* (CCAAT/enhancer binding protein) family and the *PPAR* (peroxisome proliferator-activated receptor) family, especially PPARγ. Like the MyoD and MEF2 families in skeletal muscle development, the CEBP and PPARγ proteins drive and maintain one another's expression, through various cross-regulatory and autoregulatory control loops. They work together to control the expression of the other genes characteristic of adipocytes.

The production of enzymes for import of fatty acids and glucose and for fat synthesis leads to an accumulation of fat droplets, consisting mainly of triacylglycerol (see Figure 2–81). These then coalesce and enlarge until the cell is hugely distended (up to 120 μm in diameter), with only a thin rim of cytoplasm around the mass of lipid (**Figure 23–62** and **Figure 23–63**). Lipases are also made

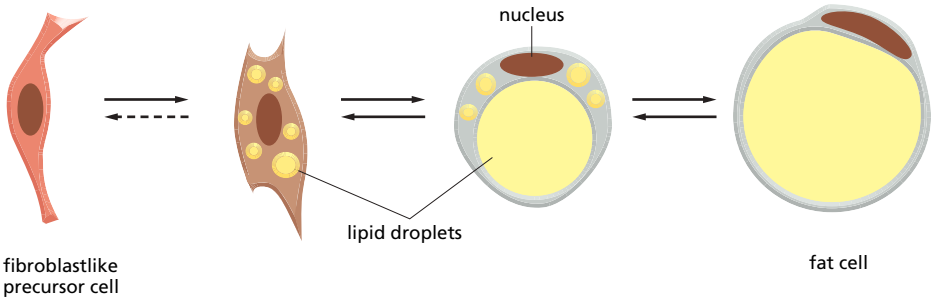


Figure 23–62 The development of a fat cell. A fibroblastlike precursor cell is converted into a mature fat cell by the accumulation and coalescence of lipid droplets. The process is at least partly reversible, as indicated by the *arrows;* the *dashed arrow* indicates uncertainty as to whether a differentiated fat cell can ever revert to the state of a pluripotent fibroblast. The cells in the early and intermediate stages can divide, but the mature fat cell cannot.

in the fat cell, giving it the capacity to reverse the process of lipid accumulation, by breaking down the triacylglycerols into fatty acids that can be secreted for consumption by other cells. The fat cell can change its volume by a factor of a thousand as it accumulates and releases lipid.

Leptin Secreted by Fat Cells Provides Feedback to Regulate Eating

Almost all animals under natural circumstances have to cope with food supplies that are are variable and unpredictable. Fat cells have the vital role of storing reserves of nourishment in times of plenty and releasing them in times of dearth. It is thus essential to the function of adipose tissue that its quantity should be adjustable throughout life, according to the supply of nutrients. For our ancestors, this was a blessing; in the well-fed half of the modern world, it has become also a curse. In the United States, for example, approximately 30% of the population suffers from obesity, defined as a body mass index (weight/height^2) more than 30 kg/m^2, equivalent to about 30% above ideal weight.

It is not easy to determine to what extent the changes in the quantity of adipose tissue depend on changes in the numbers of fat cells, as opposed to changes in fat-cell size. Changes in cell size are probably the main factor in normal nonobese adults, but in severe obesity, at least, the number of fat cells also increases. The factors that drive the recruitment of new fat cells are not well understood, although they are thought to include growth hormone and IGF1 (insulinlike growth factor-1). It is clear, however, that the increase or decrease of fat cell size is regulated directly by levels of circulating nutrients and by hormones, such as insulin, that reflect nutrient levels. The surplus of food intake over energy expenditure thus directly governs the accumulation of adipose tissue.

But how are food intake and energy expenditure themselves regulated? Factors such as cholecystokinin, secreted by gut cells in response to food in the gut lumen as discussed earlier, are responsible for short-term control, over the course of a meal or a day. But we also need long-term controls, if we are not to get steadily fatter and fatter or thinner and thinner over the course of a lifetime. Most important, from an evolutionary point of view and for our ancestors coping with food supplies that were often scanty and uncertain, starvation must provoke hunger and the pursuit of food. Those who have known real prolonged hunger testify to the overwhelming force of this compulsion. The key signal appears to be a protein hormone called **leptin**, which normally circulates in the bloodstream when fat reserves are adequate, and disappears, producing chronic hunger, when they are not. Mutant mice that lack leptin or the appropriate leptin receptor are extremely fat (**Figure 23–64**). Mutations in the same genes sometimes occur in humans, although very rarely. The consequences are similar: constant hunger, overeating, and crippling obesity.

Leptin is normally made by fat cells; the bigger they are, the more they make. Leptin acts on many tissues, and in particular in the brain, on cells in those regions of the hypothalamus that regulate eating behavior. Absence of leptin is a signal of starvation, driving the behavior that will restore fat reserves to their proper level. Thus, leptin, like myostatin released from muscle cells, provides a feedback mechanism to regulate the growth of the tissue that secretes it.

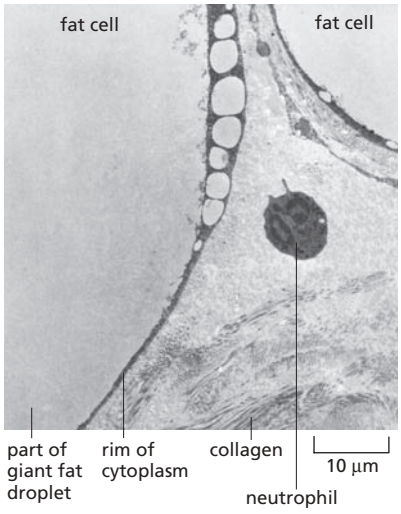


Figure 23–63 Fat cells. This low-magnification electron micrograph shows parts of two fat cells. A neutrophil cell that happens to be present in the adjacent connective tissue provides a sense of scale; each of the fat cells is more than 10 times larger than the neutrophil in diameter and is almost entirely filled with a single large fat droplet. The small fat droplets *(pale oval shapes)* in the remaining rim of cytoplasm are destined to fuse with the central droplet. The nucleus is not visible in either of the fat cells in the picture. (Courtesy of Don Fawcett, from D.W. Fawcett, A Textbook of Histology, 12th ed. New York: Chapman and Hall, 1994.)

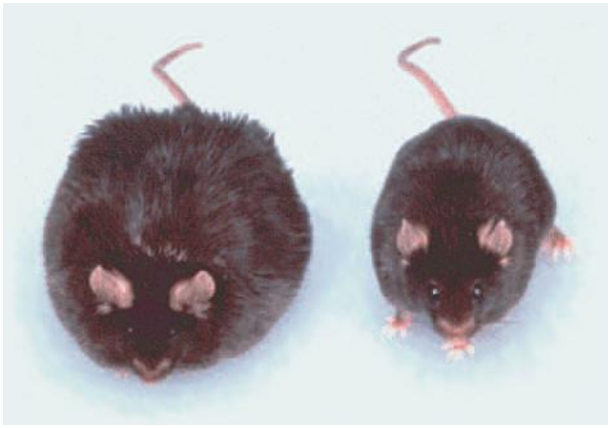


Figure 23–64 Effects of leptin deficiency. A normal mouse *(right)* compared with a mouse that has a mutation in the *Obese* gene, which codes for leptin *(left)*. The leptin-deficient mutant fails to limit its eating and becomes grotesquely fat (three times the weight of a normal mouse). (Courtesy of Jeffrey M. Friedman.)

In most obese people, leptin levels in the bloodstream are persistently high, and yet appetite is not suppressed, even though leptin receptors are also present and functional. The leptin feedback control evolved, it seems, to save us from death by starvation, rather than from obesity through overeating. In the well-fed regions of the world, we depend on a complex of other mechanisms, many of them still poorly understood, to keep us from getting too fat.

Summary

The family of connective-tissue cells includes fibroblasts, cartilage cells, bone cells, fat cells, and smooth muscle cells. Some classes of fibroblasts, such as the mesenchymal stem cells of bone marrow, seem to be able to transform into any of the other members of the family. These transformations of connective-tissue cell type are regulated by the composition of the surrounding extracellular matrix, by cell shape, and by hormones and growth factors.

Cartilage and bone both consist of cells and solid matrix that the cells secrete around themselves—chondrocytes in cartilage, osteoblasts in bone (osteocytes being osteoblasts that have become trapped within the bone matrix). The matrix of cartilage is deformable so that the tissue can grow by swelling, whereas bone is rigid and can grow only by apposition. The two tissues have related origins and collaborate closely. Thus, most long bones develop from miniature cartilage "models," which, as they grow, serve as templates for the deposition of bone. Wnt signaling regulates the choice between the two pathways of cell differentiation—as chondrocyte (requiring Sox9 expression) or as osteoblast (requiring Runx2 expression). While osteoblasts secrete bone matrix, they also produce signals that recruit monocytes from the circulation to become osteoclasts, which degrade bone matrix. Osteoblasts and osteocytes control the balance of deposition and degradation of matrix by adjusting the signals they send to the osteoclasts. Through the activities of these antagonistic classes of cells, bone undergoes perpetual remodeling through which it can adapt to the load it bears and alter its density in response to hormonal signals. Moreover, adult bone retains an ability to repair itself if fractured, by reactivation of the mechanisms that governed its embryonic development: cells in the neighborhood of the break convert into cartilage, which is later replaced by bone.

While the chief function of most members of the connective-tissue family is to secrete extracellular matrix, fat cells serve as storage sites for fat. Feedback control keeps the quantity of fat tissue from falling too low: fat cells release a hormone, leptin, which acts in the brain, and disappearance of leptin acts as a starvation danger signal, driving the behavior that will restore fat reserves to an adequate level.

STEM-CELL ENGINEERING

As we have seen, many of the tissues of the body are not only self-renewing but also self-repairing, and this is largely thanks to stem cells and the feedback

controls that regulate their behavior. But where Nature's own mechanisms fail, can we intervene and do better? Can we find ways of getting cells to reconstruct living tissues that have been lost or damaged by disease or injury and are incapable of spontaneous repair? An obvious strategy is to exploit the special developmental capabilities of the stem cells or progenitors from which the missing tissue components normally derive. But how are such cells to be obtained, and how can we put them to use? That is the topic of this final section.

Hemopoietic Stem Cells Can Be Used to Replace Diseased Blood Cells with Healthy Ones

Earlier in this chapter, we saw how mice can be irradiated to kill off their hemopoietic cells, and then rescued by a transfusion of new stem cells, which repopulate the bone marrow and restore blood-cell production. In the same way, patients with leukemia, for example, can be irradiated or chemically treated to destroy their cancerous cells along with the rest of their hemopoietic tissue, and then can be rescued by a transfusion of healthy, non-cancerous hemopoietic stem cells, which can be harvested from the bone marrow of a suitable donor. This creates problems of immune rejection if the bone marrow donor and the recipient differ genetically, but careful tissue matching and the use of immunosuppressive drugs can reduce these difficulties to a tolerable level. In some cases, where the leukemia arises from a mutation in a specialized type of blood cell progenitor rather than in the hemopoietic stem cell itself, it is possible to rescue the patient with his or her own cells. A sample of bone marrow is taken before the irradiation and sorted to obtain a preparation of hemopoietic stem cells that is free from leukemic cells. This purified preparation is then transfused back into the patient after the irradiation.

The same technology also opens the way, in principle, to one form of gene therapy: hemopoietic stem cells can be isolated in culture, genetically modified by DNA transfection or some other technique to introduce a desired gene, and then transfused back into a patient in whom the gene was lacking, to provide a self-renewing source of the missing genetic component. A version of this approach is under trial for the treatment of AIDS. Hemopoietic stem cells can be taken from the patient infected with HIV, genetically modified by transfection with genetic material that makes the stem cells and their progeny resistant to HIV infection, and transfused back into the same patient.

Epidermal Stem-Cell Populations Can Be Expanded in Culture for Tissue Repair

Another simple example of the use of stem cells is in the repair of the skin after extensive burns. By culturing cells from undamaged regions of the burned patient's skin, it is possible to obtain epidermal stem cells quite rapidly in large numbers. These can then be used to repopulate the damaged body surface. For good results after a third-degree burn, however, it is essential to provide first an immediate replacement for the lost dermis. For this, dermis taken from a human cadaver can be used, or an artificial dermis substitute. This is still an area of active experimentation. In one technique, an artificial matrix of collagen mixed with a glycosaminoglycan is formed into a sheet, with a thin membrane of silicone rubber covering its external surface as a barrier to water loss, and this skin substitute (called Integra) is laid on the burned body surface after the damaged tissue has been cleaned away. Fibroblasts and blood capillaries from the patient's surviving deep tissues migrate into the artificial matrix and gradually replace it with new connective tissue. Meanwhile, the epidermal cells are cultivated until there are enough to form a thin sheet of adequate extent. Two or more weeks after the original operation, the silicone rubber membrane is carefully removed and replaced with this cultured epidermis, so as to reconstruct a complete skin.

Neural Stem Cells Can Be Manipulated in Culture

While the epidermis is one of the simplest and most easily regenerated tissues, the central nervous system (the CNS) is the most complex and seems the most difficult to reconstruct in adult life. The adult mammalian brain and spinal cord have very little capacity for self-repair. Stem cells capable of generating new neurons are hard to find in adult mammals—so hard to find, indeed, that until recently they were thought to be absent.

We now know, however, that CNS neural stem cells capable of giving rise to both neurons and glial cells do persist in the adult mammalian brain. Moreover, in certain parts of the brain they continually produce new neurons to replace those that die (**Figure 23–65**). Neuronal turnover occurs on a more dramatic scale in certain songbirds, where large numbers of neurons die each year and are replaced by newborn neurons as part of the process by which the bird learns a new song in each breeding season.

The proof that the adult mammalian brain contains neural stem cells came from experiments in which pieces of brain tissue were dissociated and used to establish cell cultures. In suitable culture conditions, cells derived from an appropriate region of the brain will form floating "neurospheres"—clusters consisting of a mixture of neural stem cells with neurons and glial cells derived from the stem cells. These neurospheres can be propagated through many cell generations, or their cells can be taken at any time and implanted back into the brain of an intact animal. Here they will produce differentiated progeny, in the form of neurons and glial cells.

Using slightly different culture conditions, with the right combination of growth factors in the medium, the neural stem cells can be grown as a monolayer and induced to proliferate as an almost pure stem-cell population without attendant differentiated progeny. By a further change in the culture conditions, these cells can be induced at any time to differentiate to give a mixture of neurons and glial cells (**Figure 23–66**), or just one of these two cell types, according to the composition of the culture medium.

The pure cultures of neural stem cells, dividing to produce more neural stem cells, are valuable as more than just a source of cells for transplantation. They should help in the analysis of the factors that define the stem-cell state and control the switch to differentiation. Since the cells can be manipulated genetically by DNA transfection and other means, they open up new ways to investigate the role of specific genes in these processes and in genetic diseases of the nervous system, such as neurodegenerative diseases. They also create opportunities, in principle at least, for genetic engineering of neural cells to treat disease.

Neural Stem Cells Can Repopulate the Central Nervous System

Neural stem cells grafted into an adult brain show a remarkable ability to adjust their behavior to match their new location. Stem cells from the mouse hippocampus, for example, implanted in the mouse olfactory-bulb-precursor pathway (see Figure 23–65) give rise to neurons that become correctly incorporated into the olfactory bulb. This capacity of neural stem cells and their progeny to

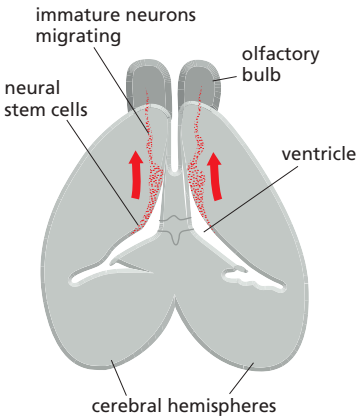


Figure 23–65 The continuing production of neurons in an adult mouse brain. The brain is viewed from above, in a cut-away section, to show the region lining the ventricles of the forebrain where neural stem cells are found. These cells continually produce progeny that migrate to the olfactory bulb, where they differentiate as neurons. The constant turnover of neurons in the olfactory bulb is presumably linked in some way to the turnover of the olfactory receptor neurons that project to it from the olfactory epithelium, as discussed earlier. There is also a continuing turnover of neurons in the adult hippocampus, a region specially concerned with learning and memory, where plasticity of adult function seems to be associated with turnover of a specific subset of neurons. (Adapted from B. Barres, *Cell* 97:667–670, 1999. With permission from Elsevier.)

adapt to a new environment promises to have important clinical applications in the treatment of diseases where neurons degenerate or lose their myelin sheaths, and in injuries of the central nervous system. Thus, neural stem cells (derived from fetal human tissue) have been grafted into the spinal cord of mice that are crippled by a spinal cord injury or by a mutation that leads to defective myelination; the mice chosen were of an immunodeficient strain, and so did not reject the grafted cells. The grafted cells then gave rise both to neurons that connected with the host neurons and to oligodendrocytes that formed new myelin sheaths around demyelinated host axons. As a result, the host mice recovered some of their control over their limbs.

Such findings hold out the hope that, in spite of the extraordinary complexity of nerve cell types and neuronal connections, it may be possible to use neural stem cells to repair at least some types of damage and disease in the central nervous system.

Stem Cells in the Adult Body Are Tissue-Specific

When cells are removed from the body and maintained in culture or are transplanted from one site in the body to another, as in the procedures we have just described, they generally remain broadly faithful to their origins. Keratinocytes continue to behave as keratinocytes, hemopoietic cells as hemopoietic cells, neural cells as neural cells, and so on. Placed in an abnormal environment, differentiated cells may, it is true, cease to display the full normal set of differentiated features, and stem cells may lose their stem-cell character and differentiate; but they do not switch to expressing the characteristics of another radically different cell type. Thus, each type of specialized cell has a memory of its developmental history and seems fixed in its specialized fate. Some limited transformations can certainly occur, as we saw in our account of the connective-tissue cell family, and some stem cells can generate a variety of differentiated cell types, but the possibilities are restricted. Each type of stem cell serves for the renewal of one particular type of tissue.

Obviously, the practical opportunities would be much greater if stem cells were more versatile and not so specialized—if we could take them from one type of tissue where they are easily available, and use them to repair a different tissue

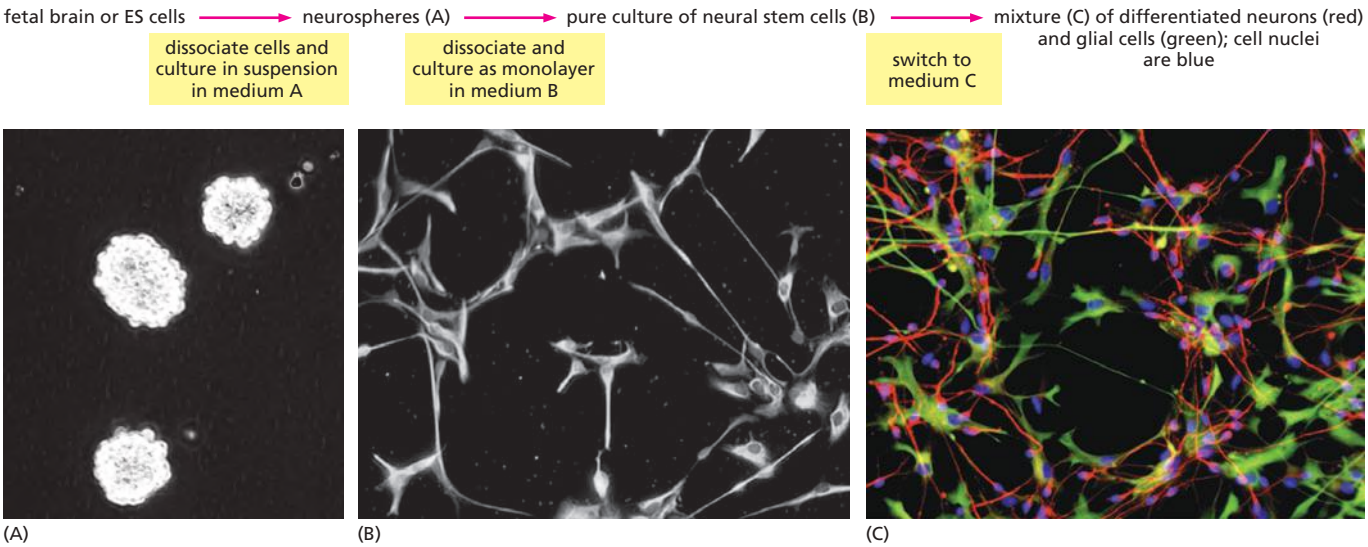


Figure 23–66 Neural stem cells. The photographs show the steps leading from fetal brain tissue, via neurospheres (A), to a pure culture of neural stem cells (B). These stem cells can be kept proliferating as such indefinitely, or, through a change of medium, can be caused to differentiate (C) into neurons *(red)* and glial cells *(green)*. Neural stem cells with the same properties can also be derived, via a similar series of steps, from ES cells. (Micrographs from L. Conti et al., *PLoS* 3:1594–1606, 2005. With permission from Public Library of Science.)

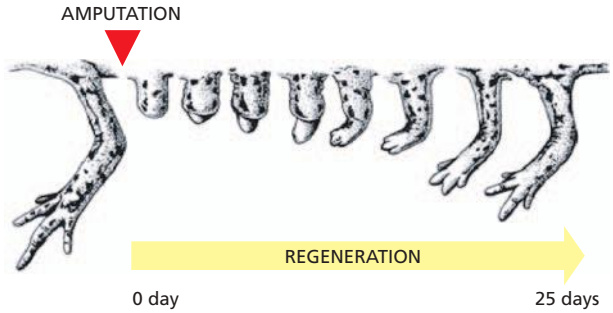


Figure 23–67 Newt limb regeneration. A time-lapse sequence showing the progress of limb regeneration in an axolotl from amputation at the level of the humerus. The sequence show the wound-healing, dedifferentiation, blastema, and redifferentiation stages of regeneration. Total time shown is approximately 20–30 days. (Courtesy of Susan Bryant and David Gardiner.)

where they are needed. Thus, there has been great excitement in the past decade over reports that stem cells of various specialized tissues can, in certain circumstances, show astonishing developmental plasticity, giving rise to cells of radically different types—hemopoietic stem cells to neurons, for example, or neural stem cells to muscle. The validity of these findings is hotly debated, however, and faults have been found in some of the key evidence. For example, many apparent cases of such switches of cell fate are now thought to be actually the result of cell fusion events, through which nuclei from one type of specialized cell are exposed to cytoplasm of another cell type and consequently switch on an altered set of genes. In any case, most reports of interconversions between radically different adult cell lineages agree that these are rare events. While research continues into these extreme forms of stem-cell plasticity, we do not yet know how to make such direct interconversions happen on a large enough scale or reliably enough, if at all, for practical medical application.

This is not to say that the radical transformation of cells from one differentiated character to another is an impossible dream or that efficient ways of bringing it about will never be found. In fact, some non-mammalian species can regenerate lost tissues and organs by just such interconversions. A newt, for example, can regenerate an amputated limb through a process in which differentiated cells seem to revert to an embryonic character and recapitulate embryonic development. Differentiated multinucleate muscle cells in the remaining limb stump reenter the cell cycle, dedifferentiate, and break up into mononucleated cells; these then proliferate to form a bud similar to the limb bud of an embryo, and eventually redifferentiate into the range of cell types needed to reconstruct the missing part of the limb (**Figure 23–67**). Why a newt can manage this—as well as many other extraordinary feats of regeneration—but a mammal cannot is still a profound mystery.

ES Cells Can Make Any Part of the Body

While stem cells of adult mammalian tissues seem to be quite restricted in what they can do, another type of mammalian stem cell is extraordinarily versatile. As described in Chapters 8 and 22, it is possible to take an early mouse embryo, at the blastocyst stage, and through cell culture to derive from it a class of stem cells called **embryonic stem cells**, or **ES cells**. ES cells can be kept proliferating indefinitely in culture and yet retain an unrestricted developmental potential. If ES cells are put back into a blastocyst, they become incorporated into the embryo and can give rise to all the tissues and cell types in the body, including germ cells, integrating perfectly into whatever site they may come to occupy, and adopting the character and behavior that normal cells would show at that site. We can think of development in terms of a series of choices presented to cells as they follow a road that leads from the fertilized egg to terminal differentiation. After their long sojourn in culture, the ES cell and its progeny can evidently still read the signs at each branch in the highway and respond as normal embryonic cells would. If ES cells are implanted directly into an embryo at a later stage or into an adult tissue, however, they fail to receive the appropriate sequence of cues; their differentiation then is not properly controlled, and they will often give rise to a tumor.

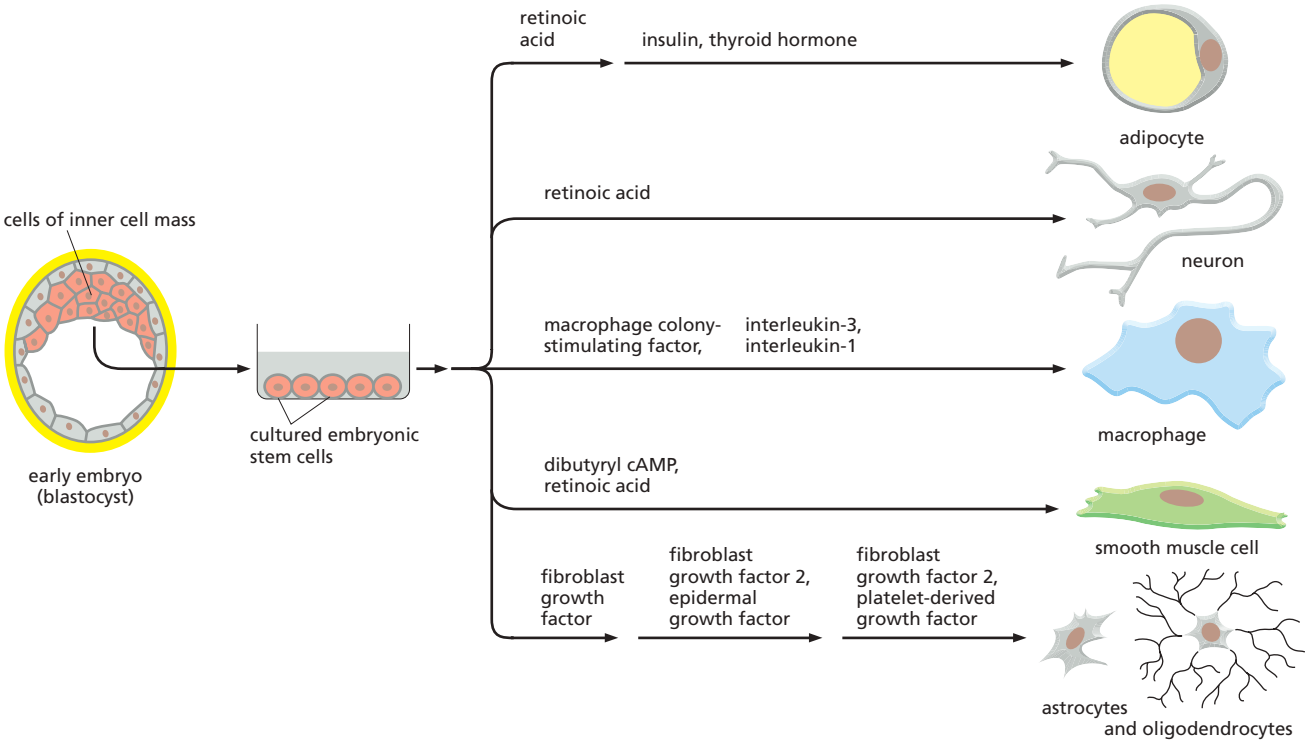


Figure 23–68 Production of differentiated cells from mouse ES cells in culture. ES cells derived from an early mouse embryo can be cultured indefinitely as a monolayer, or allowed to form aggregates called embryoid bodies, in which the cells begin to specialize. Cells from embryoid bodies, cultured in media with different factors added, can then be driven to differentiate in various ways. (Based on E. Fuchs and J.A. Segre, *Cell* 100:143–155, 2000. With permission from Elsevier.)

Cells with properties similar to those of mouse ES cells can now be derived from early human embryos and from human fetal germ cells, creating a potentially inexhaustible supply of cells that might be used for the replacement and repair of mature human tissues that are damaged. Although one may have ethical objections to such use of human embryos, it is worth considering the possibilities that are opened up. Setting aside the dream of growing entire organs from ES cells by a recapitulation of embryonic development, experiments in mice suggest that it should be possible in the future to use ES cells to replace the skeletal muscle fibers that degenerate in victims of muscular dystrophy, the nerve cells that die in patients with Parkinson's disease, the insulin-secreting cells that are lacking in type I diabetics, the heart muscle cells that die in a heart attack, and so on.

If ES cells are to be used for this sort of tissue repair, they first have to be coaxed along the desired pathway of development. ES cells can, in fact, be induced to differentiate into a wide variety of cell types in culture (**Figure 23–68**), by treatment with appropriate combinations of signal proteins and growth factors. <GGAA> They can, for example, be used to generate neurospheres and neural stem cells. Neural stem cells derived from mouse ES cells, like those derived from brain tissue, can be grafted into the brain of an adult host mouse, where they will differentiate to give neurons and glial cells. If the host is deficient in myelin-forming oligodendrocytes, a graft of ES-derived oligodendrocyte precursors can correct the deficiency and supply myelin sheaths for axons that lack them.

Patient-Specific ES Cells Could Solve the Problem of Immune Rejection

There are many problems to be solved before ES cells can be used effectively for tissue repair in human patients. One of the most severe, limiting the use of adult

stem cells also, is immune rejection. If ES-derived cells of a given genotype are grafted into a genetically different individual, the grafted cells are likely to be rejected by the immune system as foreign. Ways of dealing with this problem using immunosuppressive drugs have been developed for the transplantation of organs such as kidneys and hearts, but they are far from perfect.

To avoid immunological problems altogether, we need grafted cells that are genetically identical to those of the host. How, then, can ES cells be produced to order, with the same genotype as an adult human patient who needs a transplant? As discussed in Chapter 8, one possible route is via *somatic cell nuclear transfer*. In this procedure—not yet achieved with human cells, despite some false hopes—the nucleus would be taken from a somatic cell of the patient, and injected into an oocyte provided by a donor (in general, a woman other than the patient), replacing the original oocyte nucleus. From this hybrid oocyte, a blastocyst could be obtained, and from the blastocyst, ES cells. These and their progeny would contain the nuclear genome of the patient, and should in principle be transplantable without risk of immune rejection. But the whole procedure involves many difficulties, and is a long way from the stage where it could be used for treatment.

It would be far preferable if we could take cells from the adult patient and convert them to an ES-like character by manipulating gene expression more directly. A first step along this road is to identify the key determinants of ES cell character—the master regulatory proteins that specify that character, if they exist. Biochemical comparisons of ES cells with other cell types suggest a set of candidates for this role. These candidates can be tested by introducing the appropriate DNA expression constructs into differentiated cells, such as fibroblasts, that can be grown in culture. A combination of such transgenes, coding for a set of four gene regulatory proteins (Oct3/4, Sox2, Myc, and Klf4), seems in fact to be able to convert fibroblasts into cells with ES-like properties, including the ability to differentiate in diverse ways. The conversion rate is low—only a small proportion of fibroblasts containing the transgenes make the switch—and the converted cells are different from true ES cells in significant respects. Nevertheless, these experiments show a possible way toward the production of cells with ES-like versatility from adult somatic cells.

ES Cells Are Useful for Drug Discovery and Analysis of Disease

Although transplantation of ES-derived cells for the treatment of human diseases still seems to be far in the future, there are other ways in which ES cells promise to be more immediately valuable. They can be used to generate large homogeneous populations of differentiated cells of a specific type in culture; and these can serve for testing the effects of large numbers of chemical compounds in the search for new drugs with useful actions on a given human cell type. By techniques such as those we have just described, it may be possible, furthermore, to create ES-like cells containing the genomes of patients who suffer from a given genetic disease, and to use these patient-specific stem cells for the discovery of drugs useful in the treatment of that disease. Such cells should be valuable also for analysis of the disease mechanism. And at a basic level, manipulations of ES cells in culture should help us to fathom some of the many unsolved mysteries of stem-cell biology.

Serious ethical issues to need be resolved and enormous technical problems overcome before stem-cell technology can yield all the benefits that we dream of. But by one route or another, it seems that cell biology is beginning to open up new opportunities for improving on Nature's mechanisms of tissue repair, remarkable as those mechanisms are.

Summary

Stem cells can be manipulated artificially and used both for the treatment of disease and for other purposes such as drug discovery. Hemopoietic stem cells, for example,

can be transfused into leukemia patients to replace a diseased hemopoietic system, and epidermal stem cells taken from undamaged skin of a badly burned patient can be rapidly grown in large numbers in culture and grafted back to reconstruct an epidermis to cover the burns. Neural stem cells can be derived from some regions of the fetal or adult brain, and when grafted into a brain that is damaged can differentiate into neurons and glial cells that become integrated into the host tissue and may help to bring about a partial repair, at least in experimental studies in animals.

In the normal adult body, each type of stem cell gives rise to a restricted range of differentiated cell types. Although there have been many reports of stem-cell plasticity that violates these restrictions, the evidence is still contentious. Embryonic stem cells (ES cells), however, are able to differentiate into any cell type in the body, and they can be induced to differentiate into many different cell types in culture. From ES cells it is possible, for example, to generate neural stem cell lines that will proliferate indefinitely as pure stem-cell cultures but can respond to an appropriate change of culture conditions at any time by differentiating into neurons and glia. Methods to derive ES-like cells from cells of adult tissues are under development. In principle, such ES-like cells, carrying the genome of a specific patient, could be used for tissue repair, avoiding the problem of immune rejection. More immediately, they provide an in vitro testing ground for the investigation of the physiology and pharmacology of cells of any normal or pathological genotype, and for the discovery of drugs with useful effects on these cells.

REFERENCES

General

Fawcett DW (1994) Bloom and Fawcett: A Textbook of Histology, 12th ed. New York/London: Arnold/Chapman & Hall.

Kerr JB (1999) Atlas of Functional Histology. London: Mosby.

Lanza R, Gearhart J, Hogan B et al (eds) (2004). Handbook of Stem Cells. Amsterdam: Elsevier.

Young B, Lowe JS, Stevens A & Heath JW (2006) Wheater's Functional Histology: A Text and Colour Atlas, 5th ed. Edinburgh: Churchill Livingstone/Elsevier.

Epidermis and Its Renewal by Stem Cells

Fuchs E (2007) Scratching the surface of skin development. *Nature* 445:834–842.

Imagawa W, Yang J, Guzman R & Nandi S (1994) Control of mammary gland development. In The Physiology of Reproduction (Knobil E & Neill JD eds), 2nd ed, pp 1033–1063. New York: Raven Press.

Ito M, Yang Z, Andl T et al (2007) Wnt-dependent *de novo* hair follicle regeneration in adult mouse skin after wounding. *Nature* 447:316–320.

Jacinto A, Martinez-Arias A & Martin P (2001) Mechanisms of epithelial fusion and repair. *Nature Cell Biol* 3:E117–123.

Jensen UB, Lowell S & Watt FM (1999) The spatial relationship between stem cells and their progeny in the basal layer of human epidermis: a new view based on whole-mount labelling and lineage analysis. *Development* 126:2409–2418.

Prince JM, Klinowska TC, Marshman E et al (2002) Cell-matrix interactions during development and apoptosis of the mouse mammary gland *in vivo*. *Dev Dyn* 223:497–516.

Shackleton M, Vaillant F, Simpson KJ et al (2006) Generation of a functional mammary gland from a single stem cell. *Nature* 439:84–88.

Shinin V, Gayraud-Morel B, Gomes D & Tajbakhsh S (2006) Asymmetric division and cosegregation of template DNA strands in adult muscle satellite cells. *Nature Cell Biol* 8:677–687.

Stanger BZ, Tanaka AJ & Melton DA (2007) Organ size is limited by the number of embryonic progenitor cells in the pancreas but not the liver. *Nature* 445:886–891.

Steinert PM (2000) The complexity and redundancy of epithelial barrier function. *J Cell Biol* 151:F5–F8.

Watt FM, Lo Celso C & Silva-Vargas V (2006) Epidermal stem cells: an update. *Curr Opin Genet Dev* 16:518–524.

Sensory Epithelia

Axel R (2005) Scents and sensibility: a molecular logic of olfactory perception (Nobel lecture). *Angew Chem Int Ed Engl* 44:6110–6127.

Buck LB (2000) The molecular architecture of odor and pheromone sensing in mammals. *Cell* 100:611–618.

Howard J & Hudspeth AJ (1988) Compliance of the hair bundle associated with gating of mechanoelectrical transduction channels in the bullfrog's saccular hair cell. *Neuron* 1:189–199.

Izumikawa M, Minoda R, Kawamoto K et al (2005) Auditory hair cell replacement and hearing improvement by Atoh1 gene therapy in deaf mammals. *Nature Med* 11:271–276.

Masland RH (2001) The fundamental plan of the retina. *Nature Neurosci* 4:877–886.

Mombaerts P (2006) Axonal wiring in the mouse olfactory system. *Annu Rev Cell Dev Biol* 22:713–737.

Mombaerts P,Wang F, Dulac C et al (1996) Visualizing an olfactory sensory map. *Cell* 87:675–686.

Morrow EM, Furukawa T & Cepko CL (1998) Vertebrate photoreceptor cell development and disease. *Trends Cell Biol* 8:353–358.

Pazour GJ, Baker SA, Deane JA et al (2002) The intraflagellar transport protein, IFT88, is essential for vertebrate photoreceptor assembly and maintenance. *J Cell Biol* 157:103–113.

Stone JS & Rubel EW (2000) Cellular studies of auditory hair cell regeneration in birds. *Proc Natl Acad Sci USA* 97:11714–11721.

Vollrath MA, Kwan KY & Corey DP (2007) The micromachinery of mechanotransduction in hair cells. *Annu Rev Neurosci* 30:339–365.

The Airways and the Gut

Batlle E, Henderson JT, Beghtel H et al (2002) Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. *Cell* 111:251–263.

Bjerknes M & Cheng H (1999) Clonal analysis of mouse intestinal epithelial progenitors. *Gastroenterology* 116:7–14.

Crosnier C, Stamataki D & Lewis J (2006) Organizing cell renewal in the intestine: stem cells, signals and combinatorial control. *Nature Rev Genet* 7:349–359.

Dor Y, Brown J, Martinez OI & Melton DA (2004) Adult pancreatic beta-cells are formed by self-duplication rather than stem-cell differentiation. *Nature* 429:41–46.

Fre S, Huyghe M, Mourikis P et al (2005) Notch signals control the fate of immature progenitor cells in the intestine. *Nature* 435:964–968.

Haramis AP, Begthel H, van den Born M et al (2004) *De novo* crypt formation and juvenile polyposis on BMP inhibition in mouse intestine. *Science* 303:1684–1686.

Kim CF, Jackson EL, Woolfenden AE et al (2005) Identification of bronchioalveolar stem cells in normal lung and lung cancer. *Cell* 121:823–835.
Li L & Xie T (2005) Stem cell niche: structure and function. *Annu Rev Cell Dev Biol* 21:605–631.
Sancho E, Batlle E & Clevers H (2004) Signaling pathways in intestinal development and cancer. *Annu Rev Cell Dev Biol* 20:695–723.
Sansom OJ, Reed KR, Hayes AJ et al (2004) Loss of Apc *in vivo* immediately perturbs Wnt signaling, differentiation, and migration. *Genes Dev* 18:1385–1390.
Taub R (2004) Liver regeneration: from myth to mechanism. *Nature Rev Mol Cell Biol* 5:836–847.
van Es JH, van Gijn ME, Riccio O et al (2005) Notch/gamma-secretase inhibition turns proliferative cells in intestinal crypts and adenomas into goblet cells. *Nature* 435:959–963.

Blood Vessels, Lymphatics, and Endothelial Cells

Adams RH (2003) Molecular control of arterial-venous blood vessel identity. *J Anat* 202:105–112.
Carmeliet P & Tessier-Lavigne M (2005) Common mechanisms of nerve and blood vessel wiring. *Nature* 436:193–200.
Folkman J & Haudenschild C (1980) Angiogenesis in vitro. *Nature* 288:551–556.
Folkman J (1996) Fighting cancer by attacking its blood supply. *Sci Am* 275:150–154.
Gerhardt H, Golding M, Fruttiger M et al (2003) VEGF guides angiogenic sprouting utilizing endothelial tip cell filopodia. *J Cell Biol* 161:1163–1177.
Hellstrom M, Phng LK, Hofmann JJ et al (2007) Dll4 signalling through Notch1 regulates formation of tip cells during angiogenesis. *Nature* 445:776–780.
Lawson ND & Weinstein BM (2002) In vivo imaging of embryonic vascular development using transgenic zebrafish. *Dev Biol* 248:307–318.
Lindahl P, Johansson BR, Leveen P & Betsholtz C (1997) Pericyte loss and microaneurysm formation in PDGF-B-deficient mice. *Science* 277:242–245.
Oliver G & Alitalo K (2005) The lymphatic vasculature: recent progress and paradigms. *Annu Rev Cell Dev Biol* 21:457–483.
Pugh CW & Ratcliffe PJ (2003) Regulation of angiogenesis by hypoxia: role of the HIF system. *Nature Med* 9:677–684.

Renewal by Multipotent Stem Cells: Blood Cell Formation

Allsopp RC, Morin GB, DePinho R, Harley CB & Weissman IL (2003) Telomerase is required to slow telomere shortening and extend replicative lifespan of HSCs during serial transplantation. *Blood* 102:517–520.
Calvi LM, Adams GB, Weibrecht KW et al (2003) Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature* 425:841–846.
Hock H, Hamblen MJ, Rooke HM et al (2004) Gfi-1 restricts proliferation and preserves functional integrity of haematopoietic stem cells. *Nature* 431:1002–1007.
Metcalf D (1980) Clonal analysis of proliferation and differentiation of paired daughter cells: action of granulocyte-macrophage colony-stimulating factor on granulocyte-macrophage precursors. *Proc Natl Acad Sci USA* 77:5327–5330.
Metcalf D (1999) Stem cells, pre-progenitor cells and lineage-committed cells: are our dogmas correct? *Annu NY Acad Sci* 872:289–303.
Orkin SH (2000) Diversification of haematopoietic stem cells to specific lineages. *Nature Rev Genet* 1:57–64.
Reya T, Duncan AW, Ailles L et al (2003) A role for Wnt signalling in self-renewal of haematopoietic stem cells. *Nature* 423:409–414.
Shizuru JA, Negrin RS & Weissman IL (2005) Hematopoietic stem and progenitor cells: clinical and preclinical regeneration of the hematolymphoid system. *Annu Rev Med* 56:509–538.
Wintrobe MM (1980) Blood, Pure and Eloquent. New York: McGraw-Hill.

Genesis, Modulation, and Regeneration of Skeletal Muscle

Andersen JL, Schjerling P & Saltin B (2000) Muscle, genes and athletic performance. *Sci Am* 283:48–55.
Bassel-Duby R & Olson EN (2006) Signaling pathways in skeletal muscle remodeling. *Annu Rev Biochem* 75:19–37.
Buckingham M (2006) Myogenic progenitor cells and skeletal myogenesis in vertebrates. *Curr Opin Genet Dev* 16:525–532.
Collins CA, Olsen I, Zammit PS et al (2005) Stem cell function, self-renewal, and behavioral heterogeneity of cells from the adult muscle satellite cell niche. *Cell* 122:289–301.
Lee SJ (2004) Regulation of muscle mass by myostatin. *Annu Rev Cell Dev Biol* 20:61–86.
Weintraub H, Davis R, Tapscott S et al (1991) The myoD gene family: nodal point during specification of the muscle cell lineage. *Science* 251:761–766.

Fibroblasts and Their Transformations: the Connective-tissue Cell Family

Benya PD & Shaffer JD (1982) Dedifferentiated chondrocytes reexpress the differentiated collagen phenotype when cultured in agarose gels. *Cell* 30:215–224.
Day TF, Guo X, Garrett-Beal L & Yang Y (2005) Wnt/beta-catenin signaling in mesenchymal progenitors controls osteoblast and chondrocyte differentiation during vertebrate skeletogenesis. *Dev Cell* 8:739–750.
Flier JS (2004) Obesity wars: molecular progress confronts an expanding epidemic. *Cell* 116:337–350.
Glass DA, Bialek P, Ahn JD et al (2005) Canonical Wnt signaling in differentiated osteoblasts controls osteoclast differentiation. *Dev Cell* 8:751–764.
Karsenty G & Wagner EF (2002) Reaching a genetic and molecular understanding of skeletal development. *Dev Cell* 2:389–406.
Kronenberg HM (2003) Developmental regulation of the growth plate. *Nature* 423:332–336.
Pittenger MF, Mackay AM, Beck SC et al (1999) Multilineage potential of adult human mesenchymal stem cells. *Science* 284:143–147.
Rinn JL, Bondre C, Gladstone HB, Brown PO & Chang HY (2006) Anatomic demarcation by positional variation in fibroblast gene expression programs. *PLoS Genet* 2:e119.
Rosen ED & Spiegelman BM (2006) Adipocytes as regulators of energy balance and glucose homeostasis. *Nature* 444:847–853.
Schafer M & Werner S (2007) Transcriptional control of wound repair. *Annu Rev Cell Dev Biol* in press.
Seeman E & Delmas PD (2006) Bone quality—the material and structural basis of bone strength and fragility. *N Engl J Med* 354:2250–2261.
Zelzer E & Olsen BR (2003) The genetic basis for skeletal diseases. *Nature* 423:343–348.

Stem-cell Engineering

Brockes JP & Kumar A (2005) Appendage regeneration in adult vertebrates and implications for regenerative medicine. *Science* 310:1919–1923.
Brustle O, Jones KN, Learish RD et al (1999) Embryonic stem cell-derived glial precursors: a source of myelinating transplants. *Science* 285:754–756.
Conti L, Pollard SM, Gorba T et al (2005) Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biol* 3:e283.
Eggan K, Baldwin K, Tackett M et al (2004) Mice cloned from olfactory sensory neurons. *Nature* 428:44–49.
Lee TI, Jenner RG, Boyer LA et al (2006) Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 125:301–313.
Ming GL & Song H (2005) Adult neurogenesis in the mammalian central nervous system. *Annu Rev Neurosci* 28:223–250.
Okita K, Ichisaka T & Yamanaka S (2007) Generation of germline-competent induced pluripotent stem cells. *Nature* in press.
Raff M (2003) Adult stem cell plasticity: fact or artifact? *Annu Rev Cell Dev Biol* 19:1–22.
Schulz JT, 3rd,Tompkins RG & Burke JF (2000) Artificial skin. *Annu Rev Med* 51:231–244.
Suhonen JO, Peterson DA, Ray J & Gage FH (1996) Differentiation of adult hippocampus-derived progenitors into olfactory neurons *in vivo*. *Nature* 383:624–627.
Wagers AJ & Weissman IL (2004) Plasticity of adult stem cells. *Cell* 116:639–648.

Pathogens, Infection, and Innate Immunity

24

Infectious diseases currently cause about one-third of all human deaths in the world, more than all forms of cancer combined. In addition to the continuing heavy burden of ancient diseases such as tuberculosis and malaria, new infectious diseases are continually emerging, including the current pandemic (worldwide epidemic) of *AIDS (acquired immune deficiency syndrome),* which has already caused more than 25 million deaths worldwide. Moreover, some diseases long thought to result from other causes are now turning out to be associated with infections. Most gastric ulcers, for example, are caused not by stress or spicy food, as was once believed, but by a bacterial infection of the stomach caused by *Helicobacter pylori.*

The burden of infectious diseases is not spread equally across the planet. Poorer countries and communities suffer disproportionately. Frequently, the prevalence of infectious diseases correlates with poor public sanitation and public health systems, which are often further compromised by natural disasters or political upheavals. Some infectious diseases, however, occur primarily or exclusively among industrialized communities: Legionnaire's disease, commonly spread through air-conditioning systems, is a recent example.

Humans have long been both troubled and fascinated with infectious diseases. The earliest written descriptions of how to limit the spread of rabies date back more than 3,000 years. Since the mid-1800s, physicians and scientists have struggled to identify the agents that cause infectious diseases, collectively called **pathogens**. More recently, the advent of microbial genetics and molecular cell biology has greatly enhanced our understanding of the causes and mechanisms of infectious diseases. We now know that pathogens frequently exploit the biological attributes of their host's cells in order to infect them. This understanding can give us new insights into normal cell biology, as well as strategies for treating and preventing infectious diseases.

In a world teeming with hostile, subtle, and rapidly evolving pathogens, how does a fragile and slowly evolving human survive? Like all other multicellular organisms, we have evolved several mechanisms to resist infection by pathogens. First, physical barriers, such as our tough outer layers of skin, and associated chemical defenses, such as acid in the stomach, prevent most microorganisms (microbes) from coming into contact with sterile tissues in our body. Second, individual human cells possess some intrinsic defensive capabilities; for example, cells aggressively degrade double-stranded RNA molecules, which are a hallmark of certain kinds of viral infections. To combat especially powerful pathogens that breach these barricades, vertebrates use two types of immune defense, which are carried out by specialized proteins and cells: **innate immune responses** spring into action immediately after an infection begins and do not depend on the host's prior exposure to the pathogen, while more powerful **adaptive immune responses** operate later in an infection and are highly specific for the pathogen that induced them.

In this chapter, we begin with an overview of the different kinds of organisms that cause disease. We then discuss the cell biology of infection and, finally, consider barriers to infection and innate immunity. Adaptive immunity is the subject of Chapter 25.

In This Chapter

INTRODUCTION TO PATHOGENS 1486

CELL BIOLOGY OF INFECTION 1501

BARRIERS TO INFECTION AND THE INNATE IMMUNE SYSTEM 1524

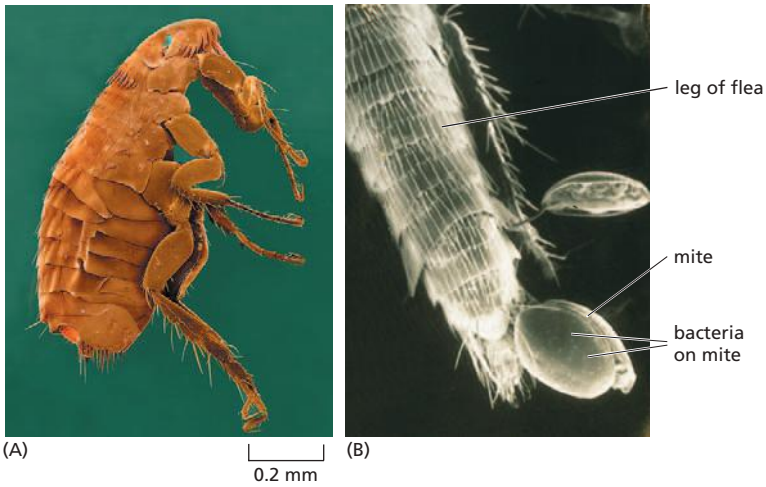


Figure 24–1 Parasitism at many levels. (A) Scanning electron micrograph of a flea. The flea is a common parasite of mammals—including dogs, cats, rats, and humans. It drinks the blood of its host. Flea bites spread bubonic plague by passing the pathogenic bacterium *Yersinia pestis* from the bloodstream of one infected host to that of another. (B) A close-up view of this flea's leg reveals that it also has a parasite, a type of mite. The mite, in turn, is covered with bacteria. It is likely that *bacteriophages*, which are bacterial viruses, parasitize these bacteria.

Jonathan Swift reported a similar observation in 1733:

So, naturalists observe, a flea
Has smaller fleas that on him prey;
And these have smaller still to bite 'em;
And so proceed ad infinitum.

(A, courtesy of Tina Carvalho/MicroAngela; B, courtesy of Stanley Falkow.)

INTRODUCTION TO PATHOGENS

We normally think of pathogens as hostile invaders that attack our bodies. But a pathogen, like any other organism, is simply fulfilling its biological imperative to live and procreate. Living at the expense of a host organism is a very effective strategy, and it is possible that every living organism on earth is subject to some type of infection (**Figure 24–1**). A human host is a nutrient-rich, warm, and moist environment, which remains at a uniform temperature and constantly renews itself. It is not surprising that many microorganisms have evolved the ability to survive and reproduce in this desirable niche. In this section, we discuss some of the common features that microorganisms must have in order to be infectious. We then explore the wide variety of organisms that are known to cause disease in humans.

Pathogens Have Evolved Specific Mechanisms for Interacting with Their Hosts

The human body is a complex and thriving ecosystem. It contains about 10^{13} human cells and also about 10^{14} bacterial, fungal, and protozoan cells, which represent thousands of microbial species. These commensal microbes, called the **normal flora**, are usually confined to certain areas of the body, including the skin, mouth, large intestine, and vagina. The normal flora are not just free-loading inhabitants of the ecosystem that is the human body, they can also affect our health. The anaerobic bacteria that inhabit our intestines contribute to the digestion of food, and they are also essential for proper development of the gastrointestinal tract in infants. Normal flora on the skin and elsewhere also help us by competing with disease-causing microorganisms. In addition, humans are always infected with viruses, the vast majority of which rarely become noticeable.

If it is normal for us to live in such close intimacy with a wide variety of microbes, why are some of them capable of causing us illness or death? As we shall see, this question has several answers, and the ability of a particular microorganism to cause obvious damage and disease in a host can depend greatly on external influences. **Primary pathogens**, which can cause overt disease in most healthy people, are usually distinct from the normal flora. They differ from commensal organisms in their abilities to breach barriers and survive in host locations where other microorganisms cannot. Our normal microbial inhabitants only cause trouble if our immune systems are weakened or if they gain access to a normally sterile part of the body, as when a bowel perforation enables gut flora to enter the peritoneal cavity of the abdomen, causing *peritonitis*; occasionally they cause disease if our immune response to them is inappropriately strong. In contrast, primary pathogens do not require an immunocompromised or injured host. Primary pathogens have evolved highly specialized

mechanisms for crossing cellular and biochemical barriers and for eliciting specific responses from the host organism that contribute to the pathogen's survival and multiplication. For some pathogens, these mechanisms are adapted to a unique host species, whereas for others they are sufficiently general that the pathogen can invade, survive, and thrive in a wide variety of hosts.

Some pathogens cause acute epidemic infections and are forced to spread rapidly from one sick or dying host to another; historically important examples include bubonic plague and smallpox. Others cause persistent infections that may last for years in a single individual without necessarily leading to overt disease; examples include Epstein–Barr virus (which can cause the severe flu-like illness mononucleosis in some people), the bacterium *Mycobacterium tuberculosis* (which can cause the life-threatening lung infection tuberculosis), and the intestinal worm *Ascaris*. Although each of these pathogens can make some people critically ill, billions of people who are mostly unaware that they are infected carry each of them in an asymptomatic way. It is hard to draw a line between persistent infection and commensalism. Throughout this chapter, we shall continue to acknowledge the diversity of pathogens and infections while focusing on the cell biological principles that are common to many of them.

In order to survive and multiply in a host, a successful pathogen must be able to: (1) colonize the host; (2) find a nutritionally compatible niche in the host's body; (3) avoid, subvert, or circumvent the host's innate and adaptive immune responses; (4) replicate, using host resources; and (5) exit and spread to a new host. Under severe selective pressure to induce host responses that help to accomplish these tasks, pathogens have evolved mechanisms that maximally exploit the biology of their host organisms. Many of them are therefore skillful and practical cell biologists, metaphorically speaking, and we can learn a great deal of cell biology by observing them.

At the same time, our constant exposure to pathogens has strongly influenced human evolution. The development of the exquisitely precise adaptive immune system in vertebrates, described in Chapter 25, was an important escalation in the arms race that has always existed between pathogens and their hosts. In modern times, humans have upped the ante by deliberately altering our behavior to limit the ability of pathogens to infect us. Improvements in public health measures, including the construction of working sewer systems and clean water supplies, have contributed to the gradual decline in the frequency of total deaths due to infectious disease over the past several centuries. Societies that have dedicated resources to improving childhood nutrition have benefited from generally improved health, including greatly reduced death rates from early childhood infections. Medical interventions such as vaccinations, antimicrobial drugs, and routine testing of blood before using it for transfusion, have also substantially reduced the burden of infectious disease for many humans. As we learn more about the mechanisms by which pathogens cause disease (pathogenesis), our brains will continue to serve as an important extension of our immune systems in fighting infectious diseases.

The Signs and Symptoms of Infection May Be Caused by the Pathogen or by the Host's Responses

Although we can easily understand why infectious microorganisms would evolve to reproduce in a host, it is less clear why they would evolve to cause disease, that is, to damage their hosts. One explanation may be that, in some cases, the pathological responses that microorganisms elicit enhance the efficiency of their spread or propagation and hence clearly have a selective advantage for the pathogen. The virus-containing lesions on the genitalia caused by *herpes simplex* infection, for example, facilitate direct spread of the virus from an infected host to an uninfected partner during sexual contact. Similarly, diarrheal infections are efficiently spread from patient to caretaker. In many cases, however, the induction of disease has no apparent advantage for the pathogen. Some host responses to infection such as lethargy and withdrawal from social interactions would instead seem to inhibit pathogen spread. Infected humans may altruistically try

to avoid infecting their friends and family members, and healthy humans may likewise consciously avoid the sick; similar behavior has also been documented in spiny lobsters. These may be adaptive responses that have been selected in host populations that tend to live in groups.

Many of the symptoms and signs that we associate with infectious disease are actually direct manifestations of the host's immune responses in action. Some hallmark signs at the site of a bacterial infection, including swelling, redness, and the production of pus (mainly dead white blood cells), result from immune system cells attempting to destroy the invading microorganisms. Fever, too, is a defensive response, as the increase in body temperature can inhibit the proliferation of some microorganisms. In extreme cases, the most severe and damaging consequences of an infectious disease are directly caused by an overzealous immune response: the massive tissue destruction seen in cases of leishmaniasis (an infection caused by eucaryotic pathogens that are members of the genus *Leishmania*) is an example. Thus, understanding the biology of an infectious disease requires an appreciation of the contributions of both pathogen and host.

To understand the relative contributions of the infecting microorganism and the host in causing the signs and symptoms of disease, it is useful to consider the cause and extent of damage done to host tissues during an infection. Each interaction between a particular microorganism and a particular host is unique, and the outcome depends on a constantly changing landscape of microbial activity and host immune function. The extent of damage caused to the host depends on the interplay of these factors. In some cases, a particular microorganism may act as a harmless or even beneficial commensal in most people at most times but may cause invasive disease in people with weakened immune systems; this is true for the common skin inhabitant *Staphylococcus epidermidis*, for example (**Figure 24–2**A). Other microorganisms, such as the virus that causes mumps, will cause severe damage only in the presence of strong immune responses (Figure 24–2B). A very interesting category, nicely illustrating the importance of the interplay between host and microbial factors in causing damage, are the many pathogens that cause severe disease in people with either very weak or very strong immune responses but little if any damage in people with intermediate immune responses (Figure 24–2C). An excellent example is tuberculosis, which currently infects between 1 and 2 billion people on Earth (usually in their lungs), although most are unaware of it because their immune systems have effectively contained the infection. When, however, a person with such a latent *M. tuberculosis* infection becomes immunosupressed, through drug therapy or infection with the human immunodeficiency virus (HIV), for example, the delicate balance between the bacterium and the immune system is tipped in favor of the bacterium, which now replicates in an uncontrolled manner, leading to serious disease, often with painful cough producing bloody sputum. Conversely, when the immune response to *M. tuberculosis* is overzealous, it can destroy an extensive amount of lung tissue.

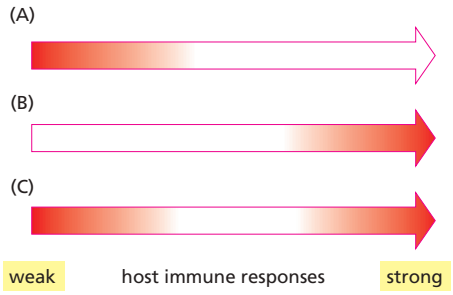


Figure 24–2 Interaction between microbes and immune responses in microbial pathogenesis. The strength of the immune responses increases in the direction of the arrows, while the amount of host damage is indicated by the amount of *red shading* in the arrows. (A) Normal flora such as the bacterium *Staphylococcus epidermidis* and the fungus *Pneumocystis carinii* only cause disease when the immune system is abnormally weak. (B) For some pathogens, such as the paramyxovirus that causes mumps and the bacterium *Chlamydia trachomatis* that causes a common sexually transmitted disease, it is the immune responses rather than the pathogen that is primarily responsible for the tissue damage. (C) For many pathogens, hosts with normal immune responses may be colonized asymptomatically, but hosts with very strong or very weak immune responses may suffer severe damage or even death. Many microorganisms behave in this way, including the bacterium *Mycobacterium tuberculosis*, the fungus *Aspergillus*, and the herpes simplex virus.

Pathogens Are Phylogenetically Diverse

Many types of pathogens cause disease in humans. The most familiar are viruses and bacteria. Viral infections cause diseases ranging from AIDS and smallpox to the common cold. They are essentially fragments of nucleic acid (DNA or RNA) encoding a relatively small number of gene products, wrapped in a protective shell of proteins and (in some cases) membrane (**Figure 24–3**A). They have no capacity for independent metabolic activity and therefore depend absolutely on metabolic energy supplied by the host. They all use the basic protein synthesis machinery of their host cells for their replication, and many rely on host cell transcription machinery as well.

Of all the bacteria we encounter in our lives, only a small minority are primary pathogens. Much larger and more complex than viruses, bacteria are usually free-living cells, which perform most of their basic metabolic functions themselves, relying on the host primarily for nutrition (Figure 24–3B).

Some other infectious agents are eucaryotic organisms. These range from single-celled fungi and protozoa (Figure 24–3C), through large complex metazoa such as parasitic worms. One of the most common infectious diseases on the planet, shared by about a billion people at present, is an infestation in the gut by the nematode worm *Ascaris lumbricoides*. Ascaris closely resembles its cousin *Caenorhabditis elegans*, which is widely used as a model organism for genetic and developmental biological research (discussed in Chapter 22). *C. elegans*, however, is only about 1 mm in length, whereas *Ascaris* can reach 30 cm (Figure 24–3D).

Some rare neurodegenerative diseases, including "mad cow" disease, are caused by an unusual type of infectious particle called a *prion*, which is made only of protein. Although the prion contains no genome, it can nevertheless replicate and kill the host.

Even within each class of pathogen, there is striking diversity. Viruses vary tremendously in their size, shape, and content (DNA versus RNA, enveloped or not, and so on), and the same is true for the other pathogens. The ability to cause disease is an evolutionary niche, not a legacy shared only among close relatives.

Each individual pathogen causes disease in a different way, and the same pathogen may cause different diseases in different hosts, making it challenging to understand the basic biology of infection. But, when considering the interactions of infectious agents with their hosts, some common themes of pathogenesis emerge, which are the focus of this chapter.

We now introduce the basic features of each of the major types of pathogens, before we examine the mechanisms that pathogens use to control their hosts and the innate immune responses that hosts use to control pathogens.

Bacterial Pathogens Carry Specialized Virulence Genes

Bacteria are small and appear structurally simple. Most can be classified broadly by their shape as rods, spheres, or spirals (**Figure 24–4**A) and by their so-called **Gram-staining** properties (Figure 24–4B and C). Their relatively small size and simple range of shapes belies their extraordinary molecular, metabolic, and ecological diversity. At the molecular level, bacteria are far more diverse than eucaryotes, and they can successfully occupy ecological niches in extremes of temperature, salt, and nutrient limitation that would daunt even the most

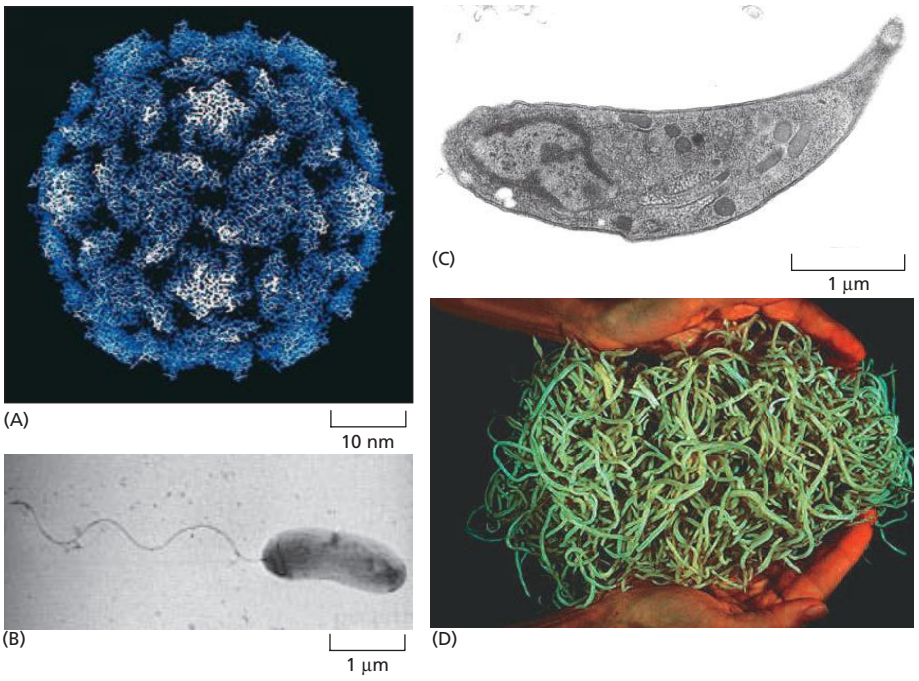


Figure 24–3 Pathogens in many forms. (A) The structure of the protein coat, or *capsid*, of poliovirus. This virus was once a common cause of paralysis, but the disease (poliomyelitis) has been nearly eradicated by widespread vaccination. (B) The bacterium *Vibrio cholerae*, the causative agent of the epidemic, diarrheal disease cholera. (C) The protozoan parasite *Toxoplasma gondii*. The definitive hosts for this organism are cats, ranging in size from housecats to tigers, but it also can cause serious infections in the muscles and brains of immunocompromised people with AIDS. (D) This clump of *Ascaris* nematodes was removed from the obstructed intestine of a two-year-old boy. (A, courtesy of Robert Grant, Stephan Crainic, and James M. Hogle; B, all attempts have been made to contact the copyright holder and we would be pleased to hear from them; C, courtesy of John Boothroyd and David Ferguson; D, from J.K. Baird et al., *Am. J. Trop. Med. Hyg.* 35:314–318, 1986. Photograph by Daniel H. Connor.)

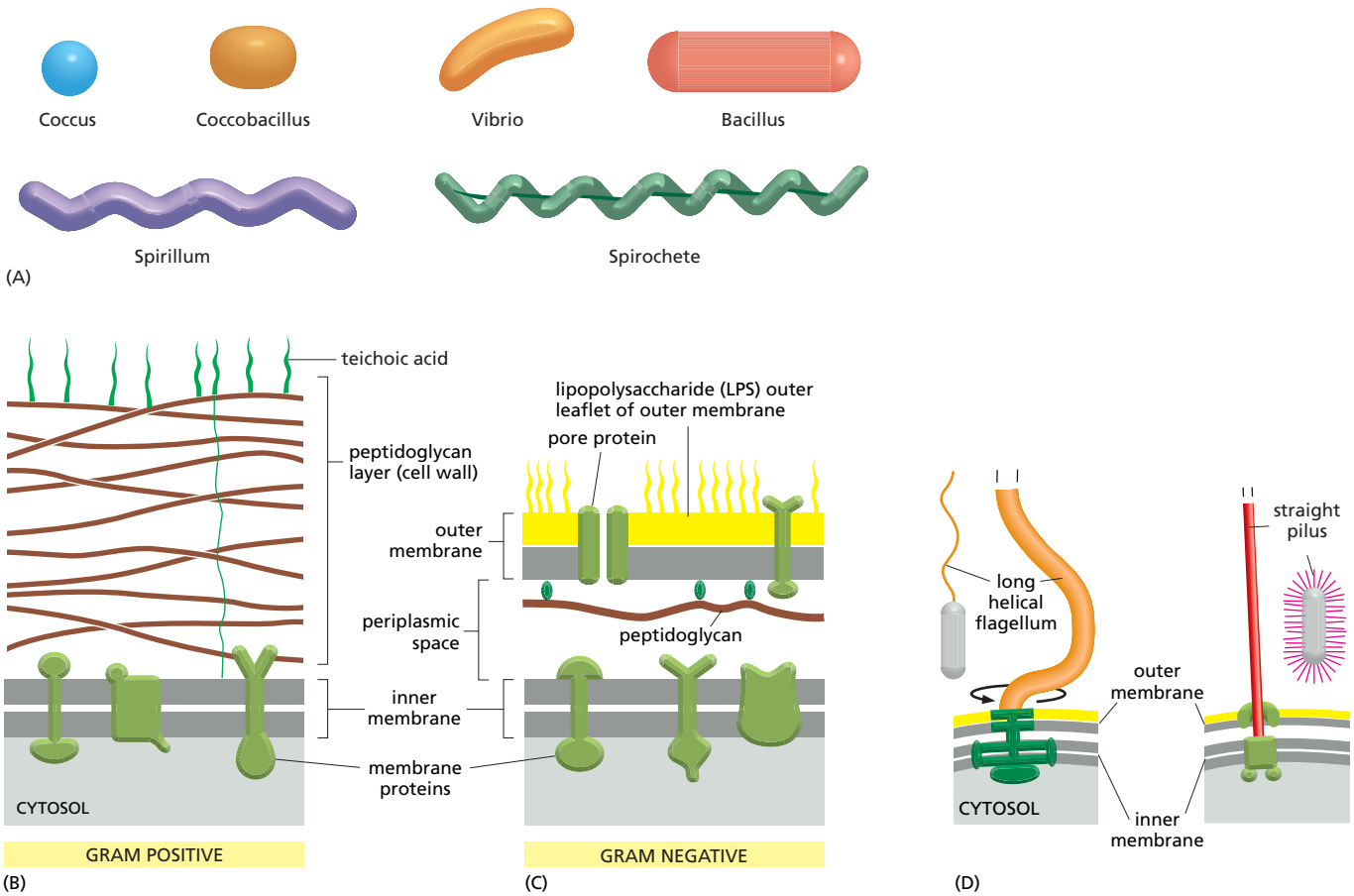


Figure 24–4 Bacterial shapes and cell-surface structures. (A) Bacteria are classified by shape. (B and C) They are also classified as *Gram-positive* or *Gram-negative.* (B) Bacteria such as *Streptococcus* and *Staphylococcus* have a single membrane and a thick cell wall made of cross-linked *peptidoglycan*. They retain the violet dye used in the Gram staining procedure and are thus called Gram-positive. (C) Gram-negative bacteria such as *Escherichia coli (E. coli)* and *Salmonella* have two membranes, separated by a *periplasmic space* (see Figure 11–18). The peptidoglycan layer in the cell wall of these organisms is located in the periplasmic space and is thinner than in Gram-positive bacteria; they therefore fail to retain the dye in the Gram staining procedure. The inner membrane of Gram-negative bacteria is a phospholipid bilayer, and the inner leaflet of the outer membrane is also made primarily of phospholipids; the outer leaflet of the outer membrane, however, is composed of a unique glycosylated lipid called *lipopolysaccharide (LPS)* (see Figure 24–47). (D) Cell-surface appendages are important for bacterial behavior. Many bacteria swim using the rotation of helical flagella (see Figure 15–71). The bacterium illustrated has only a single flagellum at one pole; others such as *E. coli* have multiple flagella. Straight *pili* (also called *fimbriae*) are used to adhere to various surfaces in the host, as well as to facilitate genetic exchange between bacteria. Some kinds of pili can retract to generate force and thereby help bacteria move across surfaces. Both flagella and pili are anchored to the cell surface by large multiprotein complexes.

intrepid eucaryote. Although they lack the elaborate morphological variety of eucaryotic cells, bacteria display a surprising array of surface appendages, which enable the cells to swim or adhere to desirable surfaces (Figure 24–4D). Their genomes are also small, typically between 1,000,000 and 5,000,000 nucleotide pairs (compared to 12,000,000 for yeast and more than 3,000,000,000 for humans).

As already emphasized, only a minority of bacterial species have the ability to cause disease in humans. Some of those that do cause disease can only replicate inside the body of their host and are called *obligate pathogens.* Others replicate in an environmental reservoir such as water or soil and only cause disease if they happen to encounter a susceptible host; these are called *facultative pathogens.* Many bacteria are normally harmless but have a latent ability to cause disease in an injured or immunocompromised host; these are called *opportunistic pathogens.* As discussed previously, whether or not a particular bacterium causes disease in a particular host depends on a wide variety of factors, including the overall health of the host; many normal flora, for example, can cause severe infections in people with AIDS.

Some bacterial pathogens are fastidious in their choice of host and will infect only a single species or a group of related species, whereas others are generalists. *Shigella flexneri*, for example, which causes epidemic dysentery (bloody diarrhea) in areas of the world lacking a clean water supply, will infect only humans and other primates. By contrast, the closely related bacterium *Salmonella enterica*, which is a common cause of food poisoning in humans, can also infect many other vertebrates, including chickens and turtles. A champion generalist is the opportunistic pathogen *Pseudomonas aeruginosa*, which is capable of causing disease in plants as well as animals.

A relatively small number of genes causes the significant differences between a virulent pathogenic bacterium and its closest nonpathogenic relative. Genes that contribute to the ability of an organism to cause disease are called **virulence genes**, and the proteins they encode are called **virulence factors**. Virulence genes are frequently clustered together, either in groups on the bacterial chromosome called *pathogenicity islands* or on extrachromosomal *virulence plasmids* (**Figure 24–5**). These genes may also be carried on mobile *bacteriophages* (bacterial viruses). It seems therefore that a new pathogen may arise when groups of virulence genes are transferred together into a previously avirulent bacterium. As more genomes of pathogenic and nonpathogenic bacteria are being completely sequenced, it is becoming clear that the acquisition of large chunks of DNA and other gross chromosomal changes have contributed to bacterial evolution, enabling bacterial species to inhabit new ecological and nutritional niches, as well as to cause disease. Even within a single bacterial species, the amount of chromosomal variation is astonishing; different strains of *E. coli* may differ by as much as 25% in their genomes.

Acquisition of genes and gene clusters can drive the rapid evolution of pathogens and turn nonpathogens into pathogens. Consider, for example, *Vibrio cholerae*—the Gram-negative bacterium that causes the epidemic diarrheal disease cholera. The genes encoding the two subunits of the toxin that cause the diarrhea are carried on a mobile bacteriophage (**Figure 24–6**A and B). Of the hundreds of strains of *Vibrio cholerae* found in lakes in the wild, the only ones that cause pandemic human disease are those infected with this bacterial virus. As summarized in Figure 24–6C, there have been eight identified pandemics of *V. cholerae* since 1817. The first six were caused by the periodic reemergence of similar strains, called Classical strains. Besides the toxins encoded by the bacteriophage and pathogenicity islands, the Classical strains also shared a similar primary carbohydrate surface antigen, called O1, which is part of the lipopolysaccharide that makes up the outer leaflet of the outer membrane (see Figure 25–4C). In 1961, the seventh pandemic began, caused by a new strain (named "El Tor"), which was markedly different from the Classical strains and appeared to have arisen when an O1-expressing strain in the wild acquired two bacteriophages, as well as at least two new pathogenicity islands not found in Classical strains. El Tor eventually displaced the Classical strains all over the world. In 1991, an eighth pandemic began, this time with the frightening twist that even people who had suffered cholera previously were not immune, as the new strain had a different type of O antigen, rendering the anti-O1 antibodies present in the blood of survivors of previous cholera epidemics ineffective against the new strain. In other respects, the new strain was very similar to El Tor; it had apparently simply acquired a new cassette for synthesis of a different type of O antigen.

What are the genes that can enable a bacterium to cause disease in a healthy host? Many virulence genes encode proteins that interact directly with host cells.

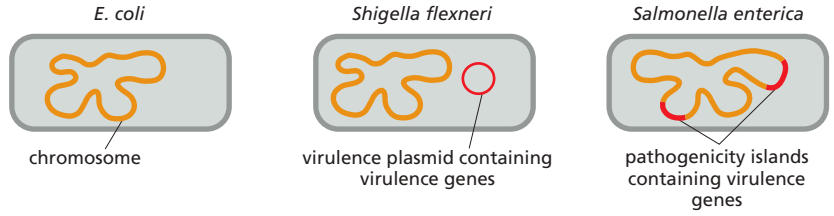


Figure 24–5 Genetic differences between pathogenic and nonpathogenic bacteria. Nonpathogenic *E. coli* has a single circular chromosome. *E. coli* is very closely related to two types of food-borne pathogens—*Shigella flexneri*, which causes dysentery, and *Salmonella enterica*, a common cause of food poisoning. If these three organisms were being named today based on molecular techniques, they would be classified in the same genus, if not in the same species. The chromosome of *S. flexneri* differs from that of *E. coli* at only a few loci; most of the genes required for pathogenesis (virulence genes) are carried on an extrachromosomal virulence plasmid. The chromosome of *S. enterica* carries two large inserts (pathogenicity islands) not found in the *E. coli* chromosome; these inserts each contain many virulence genes.

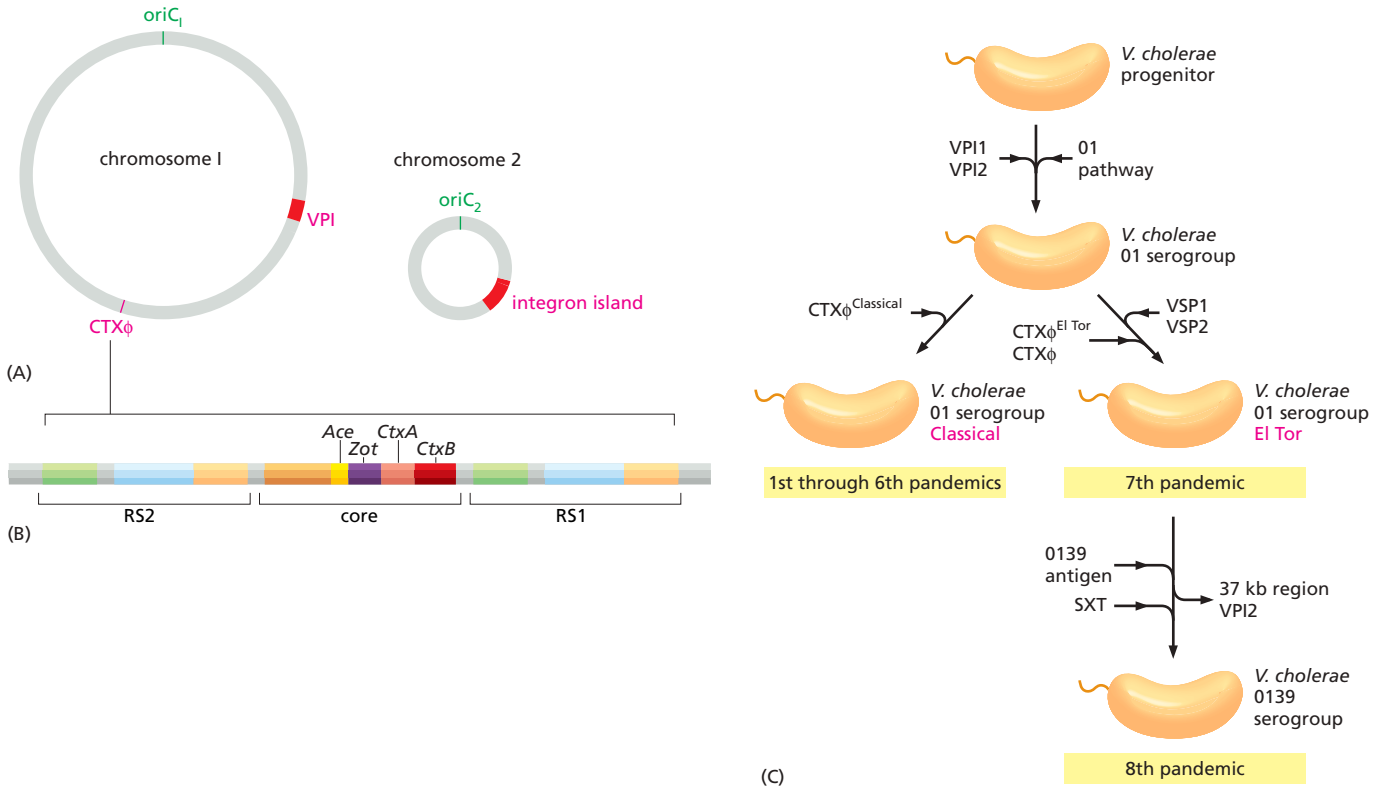


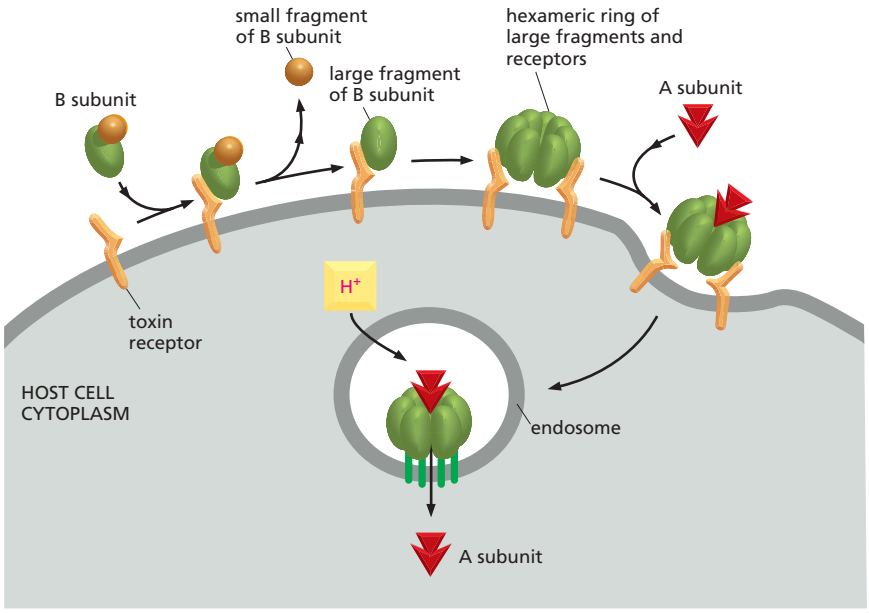
Figure 24–6 Genetic organization of *Vibrio cholerae*. An electron micrograph of *Vibrio cholerae* (*V. cholerae*) is shown in Figure 24–3B. (A) The bacterium is unusual in having two circular chromosomes rather than one. Each chromosome has its own origin of replication ($oriC_1$ and $oriC_2$). Three loci in pathogenic strains of *V. cholerae* are absent in nonpathogenic strains and appear to have been acquired relatively recently. CTXϕ on chromosome 1 is an integrated bacteriophage genome that carries the genes encoding cholera toxin. The pathogenicity island VPI on chromosome 1 includes genes encoding factors required for intestinal colonization. The *integron island* on chromosome 2 facilitates the insertion of newly acquired DNA fragments downstream of a strong transcriptional promoter and thereby promotes the sequential acquisition of novel genes. Although this integron island has not yet been shown to be required for virulence in *V. cholerae*, similar integron islands in many other pathogens contain virulence genes, as well as genes involved in antibiotic resistance. (B) Map of the CTXϕ locus. The genes encoding the two subunits of cholera toxin are *CtxA* and *CtxB*. Other genes in the core region *(Ace* and *Zot)* are also involved in virulence. The two repeated flanking sequences RS2 and RS1 were involved in the insertion of the bacteriophage genome into chromosome 1. (C) Comparative-genomics-based model for the evolution of pathogenic *V. cholerae* strains. Progenitor strains in the wild first acquired the biosynthetic pathway necessary to make the O1 antigen type of carbohydrate chain on the outer-membrane lipopolysaccharide (see Figures 24–4C and 24–47), and perhaps also acquired one or two of the *Vibrio* pathogenicity islands (VPI1 and VPI2). Incorporation of the CTXϕ bacteriophage created the Classical pathogenic strains responsible for the first six worldwide epidemics of cholera between 1817 and 1923. Sometime in the twentieth century, an O1 strain in the environment picked up the CTXϕ bacteriophage again, along with an associated bacteriophage RS1ϕ and two new pathogenicity islands (VSP1 and VSP2), creating the El Tor strain that emerged as the seventh worldwide pandemic in 1961. Before 1991, an El Tor strain picked up a new DNA cassette enabling it to produce the O139 antigen type of carbohydrate chain rather than O1. This altered its interaction with the human immune system, without diminishing its virulence, and triggered the start of the eighth pandemic; this bacterium also picked up a new pathogenicity island (SXT) and lost most of the VPI2 island.

Two carried by the *Vibrio cholerae* phage, for example, encode two subunits of **cholera toxin** (see Figure 24–6B). The B subunit of this secreted, toxic protein binds to a glycolipid component of the plasma membrane of the epithelial cells in the gut of a person who has consumed *Vibrio cholerae* in contaminated water. The B subunit transfers the A subunit through the plasma membrane into the epithelial cell cytoplasm. The A subunit is an enzyme that catalyzes the transfer of an ADP-ribose moiety from NAD^+ to the trimeric G protein G_s, which normally activates adenylyl cyclase to make cyclic AMP (discussed in Chapter 15). ADP-ribosylation of the G protein results in an overaccumulation of cyclic AMP and an ion imbalance, leading to the massive watery diarrhea associated with cholera. The infection is then spread to new hosts by the fecal–oral route via contaminated food and water.

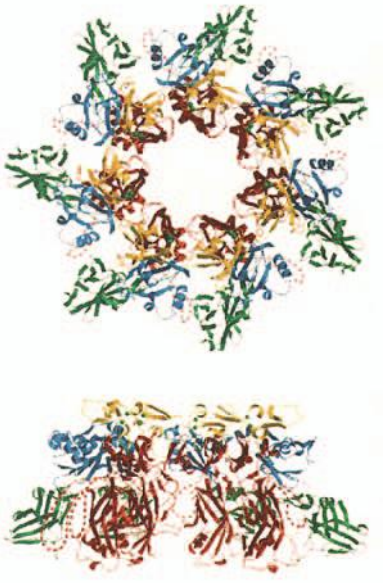
Some pathogenic bacteria use several independent mechanisms to cause toxicity to the cells of their host. *Anthrax*, for example, is an acute infectious disease

of sheep, cattle, other herbivores, and occasionally humans. It is usually caused by contact with spores of the Gram-positive bacterium *Bacillus anthracis*. Unlike cholera, anthrax does not spread directly from person to person. Dormant spores can survive in soil for long periods and are highly resistant to adverse environmental conditions, including heat, ultraviolet and ionizing radiation, pressure, and chemical agents. After the spores are inhaled, ingested, or rubbed into breaks in the skin, the spores germinate, and the bacteria begin to replicate. The bacteria secrete two toxins, called **lethal toxin** and **edema toxin**, either of which is sufficient to cause signs of infection. Like cholera toxin, both anthrax toxins are made of two subunits. The B subunit is identical in the two anthrax toxins, and it binds to a host cell-surface receptor protein to transfer the two different A subunits into host cells (**Figure 24–7**). The A subunit of edema toxin is an adenylyl cyclase that directly converts host-cell ATP into cyclic AMP, leading to an ion imbalance that can cause an accumulation of extracellular fluid *(edema)* in the infected skin or lung. The A subunit of lethal toxin is a protease that cleaves several members of the MAP kinase kinase family (see Figure 15–60). Injection of lethal toxin into the bloodstream of an animal causes shock (a fall in blood pressure) and death. The molecular mechanisms leading to death in anthrax remain uncertain.

These examples illustrate a common theme among virulence factors. The factors are frequently either toxic proteins *(toxins)* that interact directly with important host structural or intracellular signaling proteins to elicit a host cell response that is beneficial to pathogen colonization or replication, or they are proteins that are needed to deliver such toxins to the host cell targets. One common and particularly efficient delivery mechanism found in several Gram-negative pathogens, called the **type III secretion system**, acts like a tiny syringe that injects toxic proteins from the cytoplasm of an extracellular bacterium directly into the cytoplasm of an adjacent host cell (**Figure 24–8**). The effector proteins that these injection devices deliver into the host cell cytoplasm can elicit a variety of host cell responses that enable the bacterium to invade or survive. There is a remarkable degree of structural similarity between the type III syringe and the base of a bacterial flagellum (see Figure 15–71), and many of the proteins in the two structures are homologous. Because flagella are found in a wider range of bacteria than are type III secretion systems and the secretion systems appear to be adaptations specific for pathogenesis, it seems likely that the type III secretion systems evolved from flagella rather than the other way around. Other types of specialized toxin delivery systems found in pathogens appear to have evolved independently. For example, *type IV secretion systems*, used by several pathogens



(A)



(B)

Figure 24–7 Anthrax toxin entry into host cells. (A) The B subunits of the toxin form a pore to transport the enzymatically active A subunit into the cytoplasm of host cells. The B subunit binds to a receptor protein on the host cell surface and is cleaved autocatalytically, releasing a soluble small fragment and a large fragment that remains associated with the receptor. Seven copies of the large fragment and their associated receptors aggregate on the cell surface to form a heptameric ring. The toxin A subunit then binds to the heptameric ring and is endocytosed along with the ring. As the pH drops in the endosome, the ring undergoes a dramatic conformational change that generates a pore in the endosomal membrane, and the A subunit is delivered into the cytosol of the host cell through the pore. (B) X-ray crystal structure of the B subunit heptameric ring, viewed from the top and from the side. (B, from C. Petosa et al., *Nature* 385:833–838, 1997, with permission from Macmillan Publishers Ltd.)

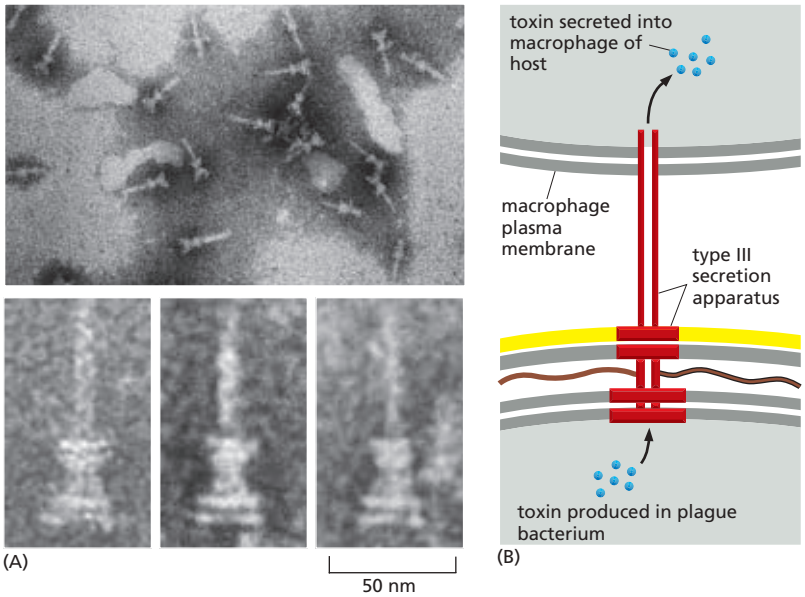


Figure 24–8 Type III secretion systems that can deliver virulence factors into the cytosol of a host cell. (A) Electron micrographs of purified type III secretion systems. About two dozen proteins form the complete structure, which is seen in the three enlarged electron micrographs below. (B) The large lower ring is embedded in the bacterial plasma membrane, and the smaller upper ring is embedded in the bacterial outer membrane. During infection, contact of the hollow tube tip with the plasma membrane of a host cell triggers secretion into the host cell. Here, the plague bacterium, *Yersinia pestis,* delivers toxins to a macrophage. (A, from K. Tamano et al., *EMBO J.* 19:3876–3887, 2000. With permission from Macmillan Publishers Ltd.)

to deliver toxins to the cytoplasm of host cells in a manner analogous to the type III systems, are closely related to the conjugative apparatus that many bacteria use to exchange genetic material.

Fungal and Protozoan Parasites Have Complex Life Cycles with Multiple Forms

Pathogenic fungi and protozoan parasites are eucaryotes. It is therefore more difficult to find drugs that will kill them without killing the host. Consequently, antifungal and antiparasitic drugs are often less effective and more toxic than antibiotics. A second characteristic of fungal and parasitic infections that makes them difficult to treat is the tendency of the infecting organisms to switch among several different forms during their life cycles. A drug that is effective at killing one form is often ineffective at killing another form, which therefore survives the treatment.

The **fungal** branch of the eucaryotic kingdom includes both unicellular *yeasts* (such as *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*) and filamentous, multicellular *molds* (like those found on moldy fruit or bread). Most of the important pathogenic fungi exhibit *dimorphism*—the ability to grow in either yeast or mold form. The yeast-to-mold or mold-to-yeast transition is frequently associated with infection. *Histoplasma capsulatum*, for example, grows as a mold at low temperature in the soil, but it switches to a yeast form when inhaled into the lung, where it can cause the disease histoplasmosis (**Figure 24–9**).

Protozoan parasites are single-celled eucaryotes with more elaborate life cycles than fungi, and they frequently require the services of more than one host. **Malaria** is the most common protozoal disease, infecting 200–300 million people every year and killing 1–3 million of them. It is caused by four species of *Plasmodium*, which are transmitted to humans by the bite of the female of any of 60 species of *Anopheles* mosquito. *Plasmodium falciparum*—the most intensively studied of the malaria-causing parasites—exists in no fewer than eight distinct forms, and it requires both the human and mosquito hosts to complete its sexual cycle (**Figure 24–10**A). Gametocytes are formed in the bloodstream of infected humans, but they can only differentiate into gametes and fuse to form a zygote in the gut of the mosquito. Three of the *Plasmodium* forms are highly specialized to invade and replicate in specific tissues—the insect gut lining, the human liver, and the human red blood cell. Even within a single host cell type, the red blood cell, the *Plasmodium* parasite undergoes a complex sequence of developmental events, reflected in striking morphological changes (Figure

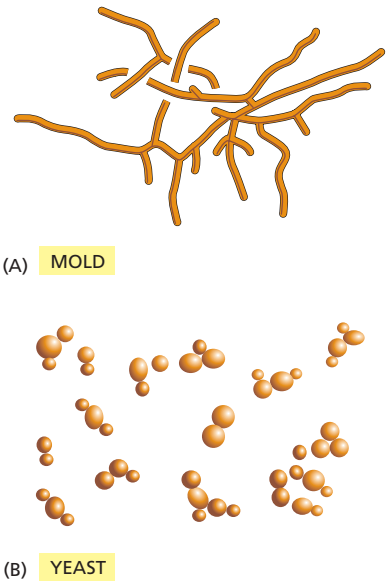
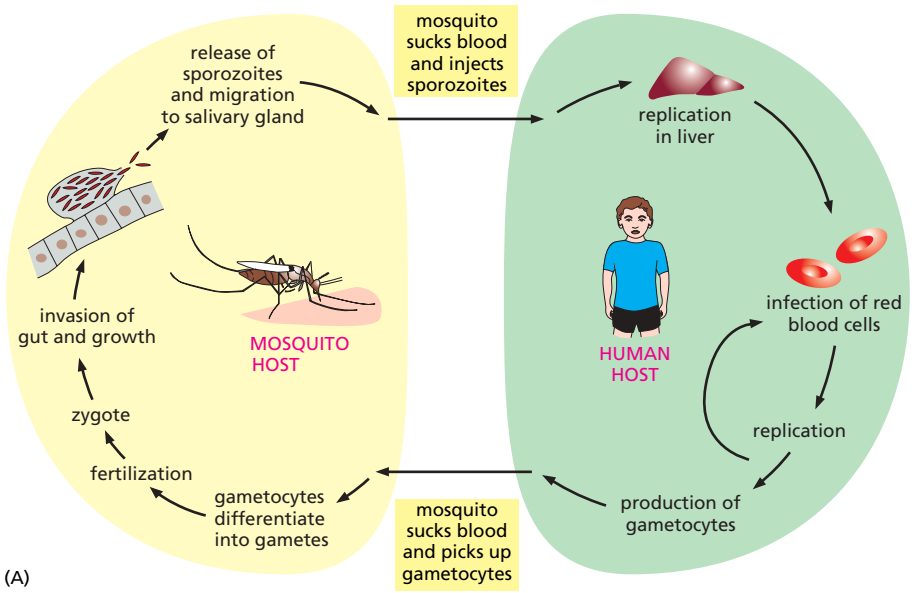


Figure 24–9 Dimorphism in the pathogenic fungus *Histoplasma capsulatum*. (A) At low temperature in the soil, *Histoplasma* grows as a multicellular filamentous mold. (B) After it is inhaled into the lung of a mammal, *Histoplasma* undergoes a morphological switch triggered by the change in temperature. In this unicellular yeast-like form, it closely resembles *Saccharomyces cerevisiae*.



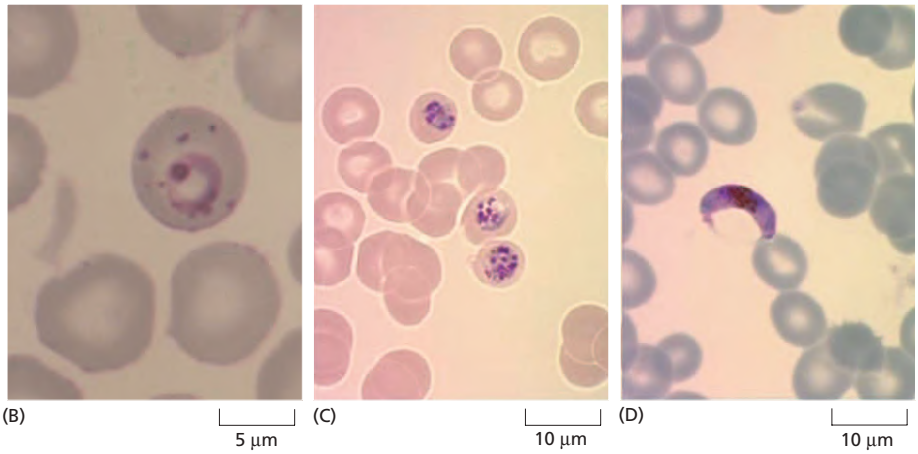


Figure 24–10 The complex life cycle of malaria parasites. (A) The sexual cycle of *Plasmodium falciparum* requires passage between a human host and an insect host. (B)–(D) Blood smears from people with malaria, showing three different forms of the parasite that appear in red blood cells: (B) ring stage; (C) schizont; (D) gametocyte. (Micrographs courtesy of the Centers for Disease Control, Division of Parasitic Diseases, DPDx.)

24–10B, C, D) as well as in stage-specific regulation of a majority of its transcripts (**Figure 24–11**).

Because malaria is so widespread and devastating, it has acted as a strong selective pressure on human populations in areas of the world that harbor the *Anopheles* mosquito. *Sickle-cell anemia*, for example, is a recessive genetic disorder caused by a point mutation in the gene that encodes the hemoglobin β chain, and it is common in areas of Africa with a high incidence of the most serious form of malaria (caused by *Plasmodium falciparum)*. The malarial parasites grow poorly in red blood cells from either homozygous sickle-cell patients or healthy heterozygous carriers, and, as a result, malaria is seldom found among carriers of this mutation. For this reason, malaria has served to maintain the otherwise deleterious sickle-cell mutation at high frequency in these regions of Africa.

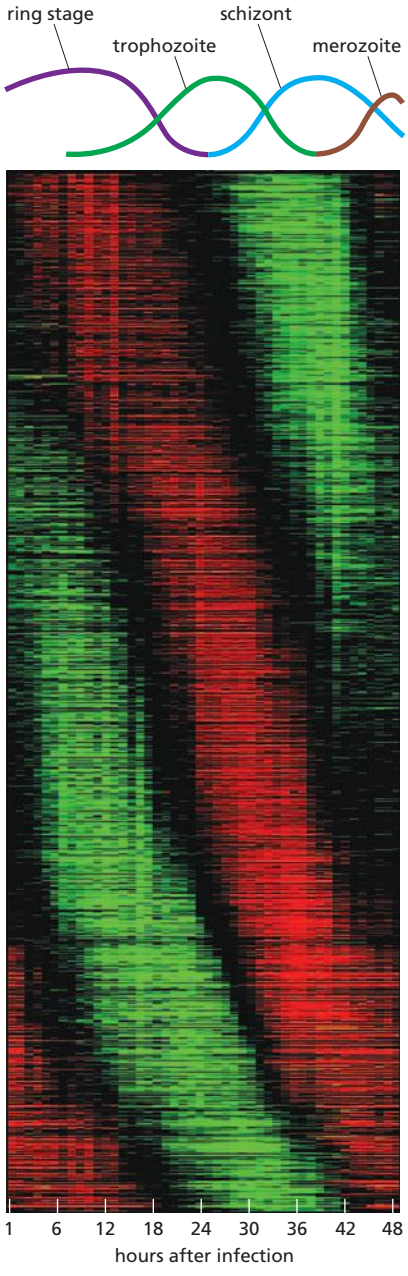


Figure 24–11 Time-dependent transcriptional program in malaria parasites developing in red blood cells. RNA was isolated from red blood cells infected with *Plasmodium falciparum* at 1-hour intervals over a total of 48 hours. In this image, each horizontal line represents one of the ~2700 genes in which the transcriptional level changed significantly during the course of infection. *Red* indicates an increase in mRNA abundance relative to the average, and *green* indicates a decrease. The genes were arranged in order from top to bottom according to the relative phases of their transcriptional activation. This regular, orderly progression of gene expression parallels the morphological differentiation of the *Plasmodium* parasite through the ring stage, trophozoite stage, schizont stage, and merozoite stage, all observed within red blood cells in infected humans (see Figure 24–10B, C, D). (Adapted from Z. Bozdech et al., *PLoS Biol.* 1:E5, 2003. With permission from Public Library of Science.)

All Aspects of Viral Propagation Depend on Host Cell Machinery

Bacteria, yeast, and protozoan pathogens are cells themselves. Even as intracellular pathogens, they use their own machinery for DNA replication, transcription, and translation, and they provide their own sources of metabolic energy. **Viruses**, by contrast, are the ultimate hitchhikers, carrying little more than information in the form of nucleic acid. The information is largely replicated, packaged, and preserved by the host cells (**Figure 24–12**). Viruses have a small genome, made up of a single nucleic acid type—either DNA or RNA—which, in either case, may be single-stranded or double-stranded. The genome is packaged in a protein coat, which in some viruses is further enclosed by a lipid envelope.

Viruses replicate in various ways. In general, replication involves (1) disassembly of the infectious virus particle, (2) replication of the viral genome, (3) synthesis of the viral proteins by the host cell translation machinery, and (4) reassembly of these components into progeny virus particles. A single virus particle (a *virion*) that infects a single host cell can produce thousands of progeny in the infected cell. Such prodigious multiplication often kills the host cell: the infected cell breaks open (lyses) and thereby allows the progeny virions access to nearby host cells. Many of the clinical manifestations of some kinds of viral infection reflect this *cytolytic effect* of the virus. Both the cold sores formed by *herpes simplex* virus and the lesions caused by the *smallpox* virus, for example, reflect the killing of the epidermal cells in a local area of infected skin. As discussed earlier and again later, some host cell death is also caused by the immune responses to the virus.

Virions come in a wide variety of shapes and sizes, and, unlike cellular life forms, they cannot be systematically classified by their relatedness into a single phylogenetic tree. Because of their tiny sizes, we now have complete genome sequences for nearly all clinically important viruses. The virions of *poxvirus* are among the largest, up to 450 nm long, which is about the size of some small bacteria. Their genome of double-stranded DNA consists of about 270,000 nucleotide pairs. At the other end of the size scale are the virions of parvovirus, which are less than 20 nm in diameter and have a single-stranded DNA genome of under 5000 nucleotides (**Figure 24–13**). The genetic information in a virus can be carried in a variety of unusual nucleic acid forms (**Figure 24–14**).

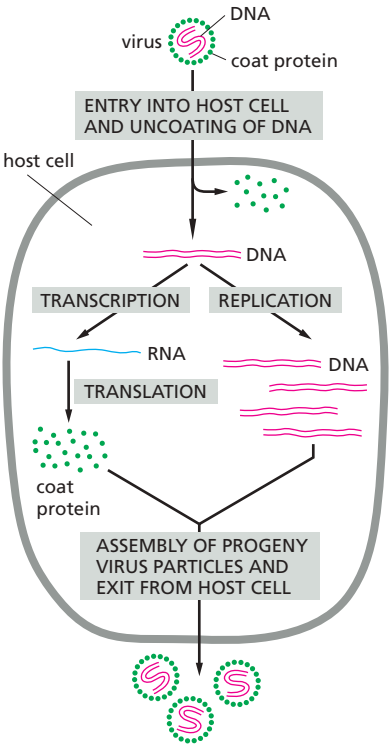


Figure 24–12 A simple viral life cycle. The hypothetical virus shown consists of a small double-stranded DNA molecule that codes for only a single viral capsid protein. No known virus is this simple.

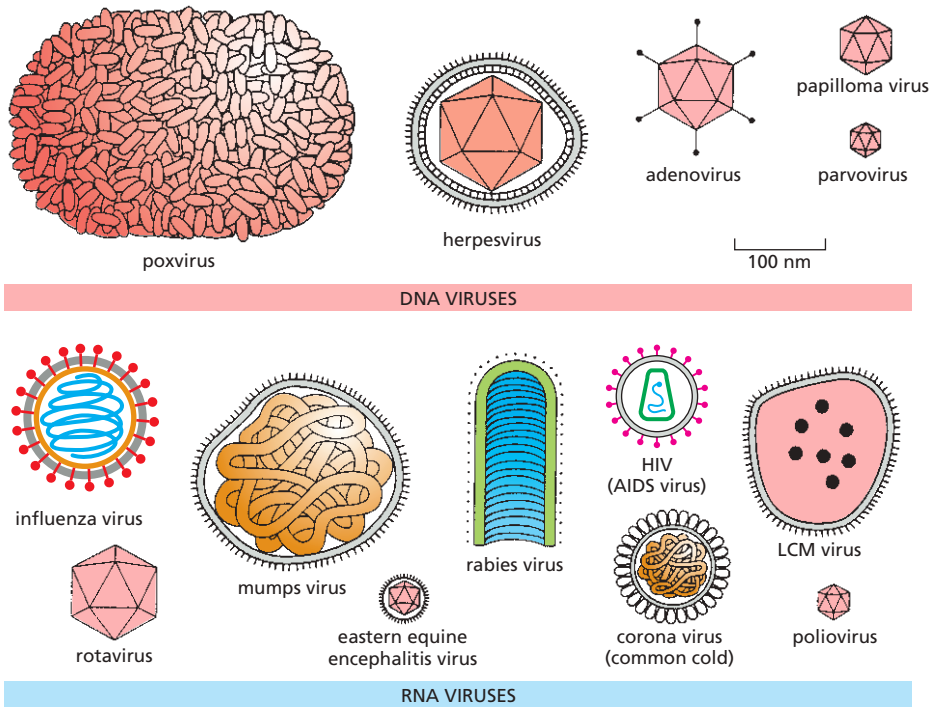


Figure 24–13 Examples of viral morphology. As shown, both DNA and RNA viruses vary greatly in both size and shape.

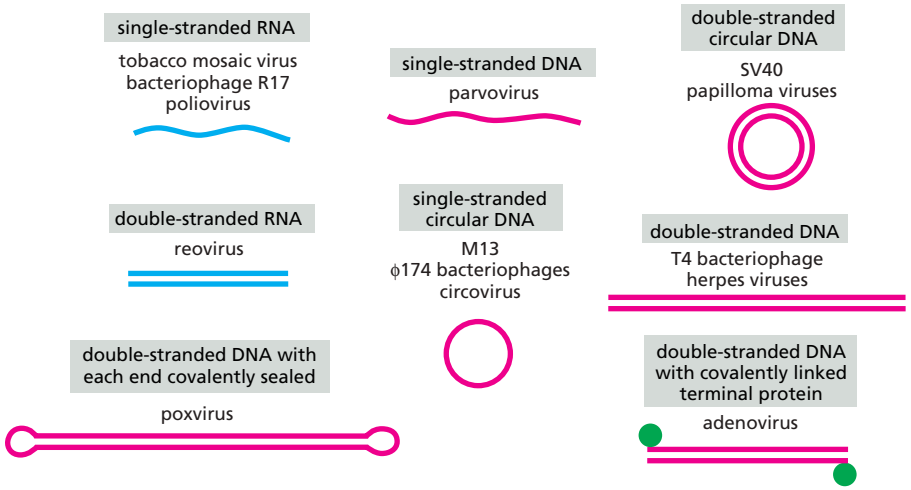
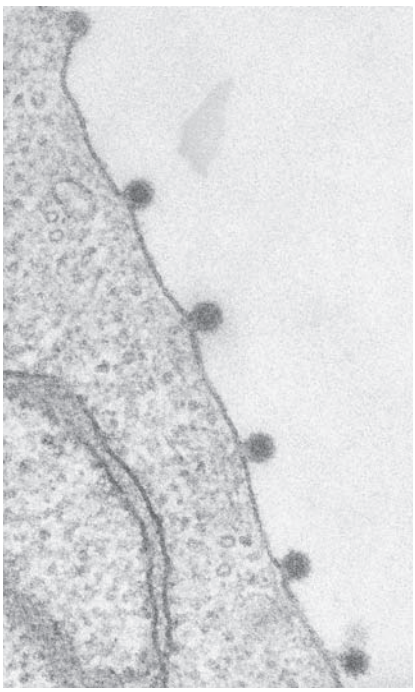


Figure 24–14 Schematic drawings of several types of viral genomes. The smallest viruses contain only a few genes and can have an RNA or a DNA genome. The largest viruses contain hundreds of genes and have a double-stranded DNA genome. The peculiar ends (as well as the circular forms) overcome the difficulty of replicating the last few nucleotides at the end of a DNA strand (discussed in Chapter 5).

The **capsid** that encloses the viral genome is made of one or several proteins, arranged in regularly repeating layers and patterns; the viral genome together with the capsid is called a **nucleocapsid**. In *enveloped viruses*, the nucleocapsid is enclosed by a lipid bilayer membrane that the virus acquires in the process of budding from the host cell plasma membrane (**Figure 24–15**). Whereas *nonenveloped viruses* usually leave an infected cell by lysing it, an enveloped virus can leave the cell by budding, without disrupting the plasma membrane and, therefore, without killing the cell. Enveloped viruses can cause persistent infections that may last for years, often without noticeable deleterious effects on the host.

Despite this variety, all viral genomes encode three types of proteins: proteins for replicating the genome, proteins for packaging the genome and delivering it to more host cells, and proteins that modify the structure or function of the host cell to enhance the replication of the virions (**Figure 24–16**). In the second section of this chapter, we focus primarily on this third class of viral proteins. Many viral genomes also encode a fourth class of proteins, which modulate or subvert the host's normal immune defense mechanisms. Several of these are described in the final section of this chapter.



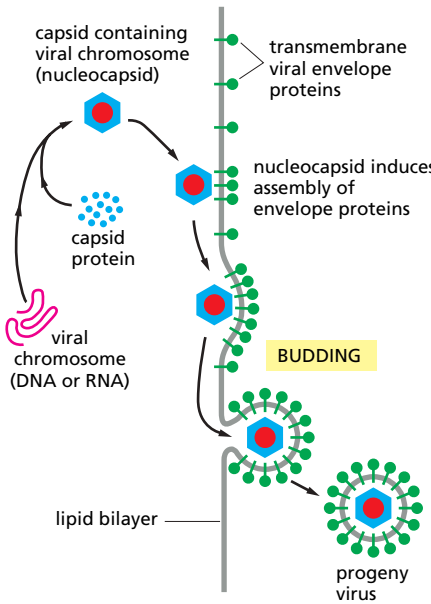


Figure 24–15 Acquisition of a viral envelope. (A) Electron micrograph of an animal cell from which six copies of an enveloped virus *(Semliki forest virus)* are budding. (B) Schematic drawing of the envelope assembly and budding processes. The lipid bilayer that surrounds the viral capsid is derived directly from the plasma membrane of the host cell. In contrast, the proteins in this lipid bilayer (shown in *green*) are encoded by the viral genome. (A, courtesy of M. Olsen and G. Griffith.)

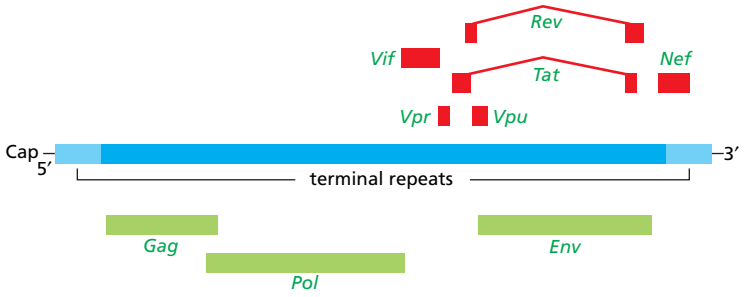


Figure 24–16 A map of the HIV genome. This retroviral RNA genome consists of about 9000 nucleotides and contains nine genes, the locations of which are shown in *green* and *red*. Three of the genes *(green)* are common to all retroviruses: *Gag* encodes capsid proteins, *Env* encodes envelope proteins, and *Pol* encodes both the reverse transcriptase (which copies the RNA into DNA) and the integrase (which inserts the DNA copy into the host cell genome) (discussed in Chapter 5). The HIV genome is unusually complex, because in addition to the three large genes *(green)* normally required for the retrovirus life cycle, it contains six small genes *(red)*. At least some of these small genes encode proteins that regulate viral gene expression *(Tat* and *Rev*—see Figure 7–103); others encode proteins that modify host cell processes, including protein trafficking *(Vpu* and *Nef)* and progression through the cell cycle *(Vpr)*. As indicated by the *red* lines, RNA splicing (using host cell spliceosomes) is required to produce the Rev and Tat proteins.

Since the host cell's machinery performs most of the critical steps in viral replication, the identification of effective antiviral drugs is problematic. Whereas the antibiotic tetracycline specifically poisons bacterial ribosomes, for example, it will not be possible to find a drug that specifically poisons viral ribosomes, as viruses use the host cell's ribosomes to make their proteins. The best strategy for containing viral diseases is to prevent them by vaccinating the potential hosts. Highly successful vaccination programs have effectively eliminated smallpox from the planet, and the eradication of poliomyelitis may be imminent (**Figure 24–17**).

Prions Are Infectious Proteins

All information in biological systems is encoded by structure. We are used to thinking of biological information in the form of nucleic acid sequences (as in our description of viral genomes), but the sequence itself is a shorthand code for describing nucleic acid structure. The replication and expression of the information encoded in DNA and RNA depend strictly on the structure of these nucleic acids and their interactions with other macromolecules. The propagation of genetic information primarily requires that the information be stored in a structure that can be duplicated from unstructured precursors. Nucleic acid sequences are the simplest and most robust solution that organisms have found to the problem of faithful structural replication.

Nucleic acids are not the only solution, however. **Prions** are infectious agents that replicate in the host by copying an aberrant protein structure. They have been found to occur in organisms ranging from yeasts to sea slugs to humans, and they cause various neurodegenerative diseases in mammals. The most well-known infection caused by prions is *bovine spongiform encephalopathy* (BSE, or mad cow disease), which occasionally spreads to humans who eat infected parts of the cow (**Figure 24–18**); it can also be transmitted from human to human via blood transfusions. Isolation of the infectious prions that cause the disease *scrapie* in sheep, followed by years of painstaking laboratory characterization of scrapie-infected mice, eventually established that the protein itself is infectious.

Intriguingly, the host makes the infectious prion protein, and the prion's amino acid sequence is identical to that of a normal host protein. Moreover, the prion and normal forms of the protein are indistinguishable in their post-translational modifications. The only difference between them appears to be in their folded three-dimensional structure. The misfolded prion protein tends to aggregate to form regular helical fibers called *amyloid*. The amyloid fibers grow at the

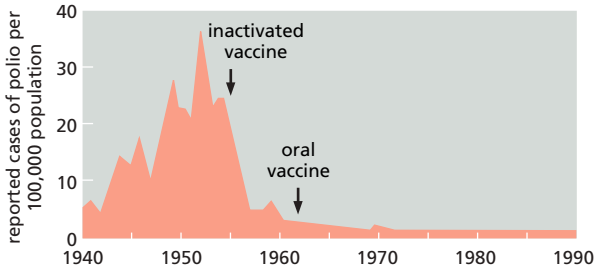


Figure 24–17 Eradication of a viral disease through vaccination. The graph shows the number of cases of poliomyelitis reported per year in the United States. The *arrows* indicate the timing of the introduction of the Salk vaccine (inactivated virus given by injection) and the Sabin vaccine (live attenuated virus given orally).

ends, much like the cytoskeletal protein filaments discussed in Chapter 16, except that the protein subunits undergo a structural conversion from the normal folded form of the protein to the misfolded form as they become part of the amyloid polymer (see Figure 6–95). In other words, the misfolded prion form has the remarkable capacity to cause the normal protein to adopt its misfolded prion conformation and thereby to become infectious, which is equivalent to the prion's having replicated itself in the host. When one of the amyloid fibrils is broken into smaller pieces, each one can seed the conversion process in a new cell; the prion can therefore propagate as well as replicate. If eaten by another susceptible host, these newly misfolded prions can transmit the infection from organism to organism.

It is not known how most normal proteins are able to find the single, correct, folded conformation, among the billions of other possibilities, without becoming stuck in dead-end intermediates (discussed in Chapters 3 and 6). Prions are a good example of how protein folding can go dangerously wrong. But why are the prion diseases so uncommon? What are the constraints that determine whether a misfolded protein will behave like a prion, or simply get refolded or degraded by the cell that made it? We do not yet have answers to these questions, and the study of prions remains an area of intense research.

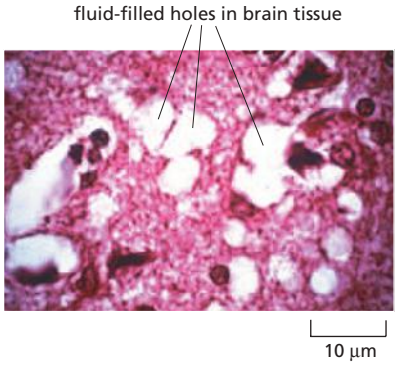


Figure 24–18 Neural degeneration in a prion infection. This micrograph shows a slice from the brain of a person who died of kuru. Kuru is a human prion disease, very similar to BSE, that was spread from one person to another by ritual mortuary practices in New Guinea. The large fluid-filled holes are places where neurons have died. These characteristic holes are why prion-based neurological diseases are called *spongiform encephalopathies*. (Courtesy of Gary Baumbach.)

Infectious Disease Agents Are Linked To Cancer, Heart Disease, and Other Chronic Illnesses

Thus far, we have considered microorganisms primarily in their roles as causative agents of infectious disease. It is clear, however, that in many cases viral and bacterial infections can contribute to the pathogenesis of important life-threatening illnesses that are not normally classified as infectious diseases. One obvious example is cancer. The oncogene concept, which is that certain altered genes can trigger cell transformation and tumor development, came initially from studies of the *Rous sarcoma virus*, which causes a form of cancer (sarcomas) in chickens. One of the genes encoded by the virus was eventually found to encode an overactive homolog of the host tyrosine kinase Src, which has since been implicated in many kinds of cancer.

Although Rous sarcoma virus does not cause cancer in humans, several human cancers are now known to have a viral origin. *Human papillomavirus*, for example, which causes genital warts, is also responsible for more than 90% of cervical cancers. Worldwide, cervical cancer is the second most common cancer in women and has a mortality rate of ~40%. In wealthy countries, widespread screening using the Pap smear test has reduced the incidence and severity of cervical cancer, but it is still very common in developing countries. The recent development of a vaccine against the most abundant cancer-associated strains of human papillomavirus raises the hope that this form of cancer can be largely prevented worldwide by a simple and cost-effective measure.

The *Epstein–Barr virus (EBV) provides a* more complex example of human cancer linked to a viral infection. Infection by this DNA virus is so common that nearly 90% of adults in the United States over the age of 40 have detectable levels of anti-EBV antibodies in their blood. EBV prefers to invade B cells of the adaptive immune system, especially long-lived memory B cells (discussed in Chapter 25). Most people infected as children have few symptoms and are unaware that they have been infected, but teenagers and young adults infected for the first time often develop *mononucleosis* (also called *glandular fever*), a severe flu-like disease that can lead to high fever, painful swelling of lymph nodes, and fatigue that can persist for several months. After symptoms subside, EBV can remain dormant in the B cells for life, with its genome maintained as an extrachromosomal plasmid in the B cell nucleus. Some of the gene products encoded by the EBV genome inhibit apoptosis and thereby presumably help to prevent the virus from being cleared from the body. Thus, when an infected B cell acquires cancer-promoting mutations, the usual mechanism for eliminating precancerous cells by apoptosis is inhibited, and a form of B cell cancer called Burkitt's lymphoma may develop.

In some cases, chronic tissue damage caused by infection can increase the likelihood of cancer developing in the infected tissue. The stomach-dwelling bacterium *Helicobacter pylori* has been implicated as a major cause of stomach cancers as well as gastric ulcers, and the *hepatitis viruses* that cause chronic infections in the liver (chronic hepatitis) are associated with more than 60% of liver cancers.

Along with cancer, the other major cause of death in wealthy industrialized nations is cardiovascular disease, frequently brought on by *atherosclerosis*, the accumulation of fatty deposits in blood vessel walls that can block blood flow. The resulting ischemia has dire consequences in the heart and brain. A hallmark of early atherosclerosis is the appearance in blood vessel walls of clumps of strange-looking macrophages, called foam cells because they are loaded with engulfed fatty globules. The foam cells secrete cytokines that recruit other white blood cells into the forming *atherosclerotic plaque*, which also accumulates extracellular matrix. The continued accumulation of cells and matrix can gradually block blood flow, or, alternatively, the plaque can rupture, causing an overlying thrombus to form, which acutely blocks blood flow; moreover, pieces of the thrombus can break off to form emboli that block smaller blood vessels downstream. Interestingly, foam cells in atherosclerotic plaques often contain the bacterial pathogen *Chlamydia pneumoniae*, which commonly causes pneumonia in humans (**Figure 24–19**). Numerous lines of evidence suggest that *C. pneumoniae* infection is a significant risk factor for atherosclerosis in humans and animal models. DNA from other bacterial species has also been found in atherosclerotic plaques, including DNA from bacteria usually associated with teeth and gums, such as *Porphyromonas gingivalis*. The connection between infectious agents and atherosclerosis is an area of active current research.

In addition to contributing to the life-threatening conditions of cancer and cardiovascular disease, infectious agents are also thought to have a role in many chronic illnesses, although it is often hard to tell whether infection causes these diseases or is a consequence of the diseases. A clear-cut case of an infectious cause for a chronic ailment is *Lyme disease*, a bacterial infection caused by the spirochete *Borrelia burgdorferi*. The infection is acquired by a tick bite and can cause painful chronic arthritis if it is not detected and treated early with antibiotics. Several other bacterial infections, particularly infections by Gram-positive cocci and by small bacteria without a wall, called *Mycoplasma*, can also trigger immune responses leading to arthritis. In some people, *Mycoplasma*, *Chlamydia pneumoniae,* or both, are associated with chronic asthma. As we learn more about the complex interactions between pathogens and the human body, it seems likely that more and more chronic conditions will be found to have a link to an infectious agent. As has been the case with peptic ulcers, curing the infection may cure the disease, or at least alleviate the painful symptoms.

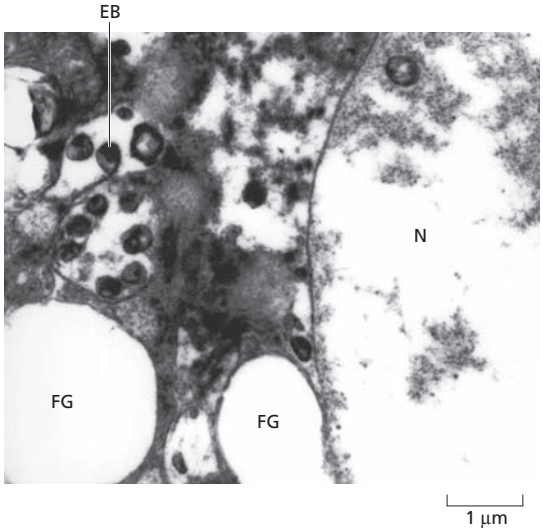


Figure 24–19 *Chlamydia pneumoniae* within a foam cell macrophage in an atherosclerotic plaque. The labels on this electron micrograph indicate: EB, elementary body (bacterium); FG, fat globule; N, macrophage nucleus. (From L.A. Campbell and C.C. Kuo, *Nat. Rev. Microbiol.* 2:23–32, 2004. With permission from Macmillan Publishers Ltd.)

Summary

Infectious diseases are caused by pathogens, which include bacteria, fungi, protozoa, worms, viruses, and even infectious proteins called prions. All pathogens must have mechanisms for entering their host and for evading immediate destruction by the host. Most bacteria are not pathogenic. Those that are contain specific virulence genes that mediate interactions with the host, eliciting responses from the host cells that promote the replication and spread of the pathogen. Pathogenic fungi, protozoa, and other eucaryotic parasites typically pass through several different forms during the course of infection; the ability to switch among these forms is usually required for the parasites to survive in a host and cause disease. In some cases, such as malaria, parasites must pass sequentially through several host species to complete their life cycles. Unlike bacteria and eucaryotic parasites, viruses have no metabolism of their own and no intrinsic ability to produce the proteins encoded by their DNA or RNA genomes; they rely entirely on subverting the machinery of the host cell to produce their proteins and to replicate their genomes. Prions, the smallest and simplest infectious agents, contain no nucleic acid; instead, they are rare, aberrantly folded proteins that replicate by catalyzing the misfolding of normal host proteins with the same amino acid sequence as the prion.

CELL BIOLOGY OF INFECTION

The mechanisms that pathogens use to cause disease are as diverse as the pathogens themselves. Nonetheless, all pathogens must carry out certain common tasks: they must colonize the host, reach an appropriate niche, avoid host defenses, replicate, and exit from the infected host to spread to an uninfected one. In this section, we examine the common strategies that many pathogens use to accomplish these tasks.

Pathogens Cross Protective Barriers to Colonize the Host

The first step in infection is for the pathogen to colonize the host. A thick and fairly tough covering of skin protects most parts of the human body from the environment. The protective boundaries of some other human tissues (eyes, nasal passages, respiratory tract, mouth, digestive tract, urinary tract, and female genital tract) are less robust. In the lungs and small intestine, for example, where oxygen and nutrients, respectively, are absorbed from the environment, the barrier is just a single monolayer of epithelial cells.

Skin and many other epithelial barriers are densely populated by normal flora. Some pathogens also colonize these surfaces and attempt to outcompete the normal flora, but most pathogens avoid such competition by crossing the barriers to gain access to unoccupied niches within the host.

Wounds in barrier epithelia allow pathogens direct access to such niches. This avenue of entry requires little in the way of pathogen specialization, and many members of the normal flora can cause serious illness if they enter through such wounds. Anaerobic bacteria of the genus *Bacteroides*, for example, are carried as harmless flora at very high density in the large intestine, but they can cause life-threatening peritonitis if they enter the peritoneal cavity through a perforation in the intestinal wall caused by trauma, surgery, or infection. *Staphylococcus* from the skin and nose, or *Streptococcus* from the throat and mouth, are also responsible for many serious infections resulting from breaches in epithelial barriers.

Primary pathogens, however, need not wait for a wound to gain access to their host. A particularly efficient way for a pathogen to cross the skin is to catch a ride in the saliva of a biting arthropod. Many insects and ticks nourish themselves by sucking mammalian blood, and a diverse group of bacteria, viruses, and protozoa have developed the ability to survive in arthropods and then use them as *vectors* to spread from one mammalian host to another. As discussed earlier, the *Plasmodium* protozoan that causes malaria develops through several

forms in its life cycle, including some that are specialized for survival in a human and some that are specialized for survival in a mosquito (see Figure 24–10). Viruses that are spread by insect bites include the causative agents for several types of hemorrhagic fever, including yellow fever and Dengue fever, as well as the causative agents for many kinds of viral encephalitis (inflammation of the brain). All these viruses replicate in both insect cells and mammalian cells, as required for their transmission by an insect vector. Bloodborne viruses such as HIV that are not capable of replicating in insect cells are rarely, if ever, spread from insect to human.

The efficient spread of a pathogen via an insect vector requires that individual insects consume blood meals from numerous mammalian hosts. In a few striking cases, the pathogen appears to alter the behavior of the insect so that its transmission is more likely. Like most animals, the tsetse fly (whose bite spreads the protozoan parasite *Trypanosoma brucei*, which causes sleeping sickness in Africa) stops eating when it is satiated. Tsetse flies carrying trypanosomes, however, bite much more frequently and ingest more blood than do uninfected flies. The presence of trypanosomes impairs the function of the insect mechanoreceptors that measure blood flow through the gullet to assess the fullness of the stomach, effectively fooling the tsetse fly into thinking that it is still hungry. The bacterium *Yersinia pestis*, which causes bubonic plague, uses a different mechanism to ensure that a flea carrying it bites repeatedly: it multiplies in the flea's foregut to form aggregated masses that eventually enlarge and physically block the digestive tract. The insect is then unable to feed normally and begins to starve. During repeated attempts to alleviate its hunger by feeding, some of the bacteria in the foregut are flushed into the bite site, thus transmitting plague to a new host (**Figure 24–20**).

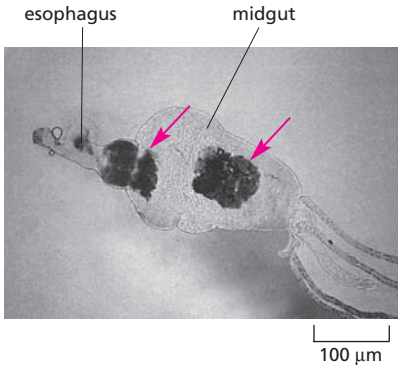


Figure 24–20 Plague bacteria within a flea. This light micrograph shows the digestive tract dissected from a flea that had dined about two weeks previously on the blood of an animal infected with the plague bacterium, *Yersinia pestis*. The bacteria multiplied in the flea gut to produce large cohesive aggregates (*red arrows*); the bacterial mass on the left is occluding the passage between the esophagus and the midgut. This type of blockage prevents a flea from digesting its blood meals, so that hunger causes it to bite repeatedly, disseminating the infection. (From B.J. Hinnebusch, E.R. Fischer and T.G. Schwann, *J. Infect. Dis.* 178:1406–1415, 1998. With permission from University of Chicago Press.)

Pathogens That Colonize Epithelia Must Avoid Clearance by the Host

Hitching a ride through the skin on an insect proboscis is just one strategy that pathogens use to pass through host barriers. Whereas many epithelial barriers such as the skin and the lining of the mouth and large intestine are densely populated by normal flora, others, including the lining of the lower lung, the small intestine, and the bladder, are normally kept nearly sterile, despite the presence of a relatively direct route to the outside world. How do these epithelia avoid bacterial colonization? As discussed in Chapter 23, a layer of protective mucus covers the respiratory epithelium, and the coordinated beating of cilia sweeps the mucus and trapped bacteria and debris up and out of the lung. The epithelia lining the bladder and the upper gastrointestinal tract also have a thick layer of mucus, and these organs are periodically flushed by urination and by peristalsis, respectively, which washes away undesirable microbes. The pathogenic bacteria and parasites that infect these epithelial surfaces have specific mechanisms for overcoming these host-cleaning processes. Those that infect the urinary tract, for example, resist the washing action of urine by adhering tightly to the epithelium lining the tract via specific **adhesins**, which are proteins or protein complexes that recognize and bind to host cell-surface molecules. An important group of adhesins in *E. coli* strains that infect the kidney are components of the *P pili* that help the bacteria adhere to the kidney epithelial cells. These surface projections can be several micrometers long and are thus able to span the thickness of the protective mucus layer (see Figure 24–4D). At the tip of each pilus is an adhesin protein that binds tightly to a particular glycolipid disaccharide that is found on the surface of kidney cells. Strains of *E. coli that* infect the bladder rather than the kidney express a second kind of pilus that enables them to adhere to bladder epithelial cells. It is the adhesion specificity of the adhesin proteins on the tips of two types of pili that is responsible for the bacteria's colonizing the different parts of the urinary tract (**Figure 24–21**). The specificity of the adhesins also restricts the host range for these and many other pathogenic bacteria.

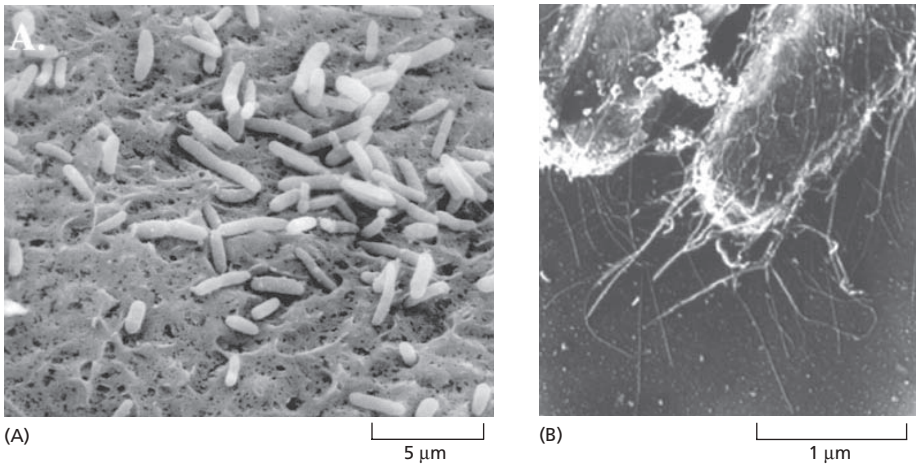


Figure 24–21 Uropathogenic *E. coli* in the infected bladder of a mouse. (A) Scanning electron micrograph of uropathogenic *E. coli*, a common cause of bladder and kidney infections. The bacteria are attached to the surface of epithelial cells lining the infected bladder. (B) A close-up view of one of the bacteria showing the pili on its surface. (A, from G.E. Soto and S.J. Hultgren, *J. Bact.* 181:1059–1071, 1999; B, courtesy of D.G. Thanassi and S.J. Hultgren, *Meth. Comp. Meth. Enzym.* 20:111–126, 2000. With permission from Academic Press.)

One of the hardest organs for a microbe to colonize is the stomach. Besides the thick layer of mucus and peristaltic washing, the stomach is filled with acid (average pH ~2). This extreme environment is lethal to almost all bacteria ingested in food. Nonetheless, the stomach is colonized by the hardy and enterprising bacterium *Helicobacter pylori*, which was recognized only recently as the major causative agent of stomach ulcers and some cases of stomach cancer. Remarkably, it is able to persist for life as a harmless commensal in most of its hosts. Although the older treatments for ulcers (acid-reducing drugs and bland diets) are still used to reduce inflammation, a short and relatively cheap course of antibiotics can now effectively cure a patient of recurrent stomach ulcers. The hypothesis that a persistent bacterial infection of the stomach lining could cause stomach ulcers was initially met with great skepticism. The young Australian doctor who made the initial discovery finally proved the point: he drank a flask of a pure culture of *H. pylori* and developed gastritis. One way in which *H. pylori* survives in the stomach is by producing the enzyme *urease*, which converts urea to ammonia and carbon dioxide; in this way, the bacterium surrounds itself with a layer of ammonia, which neutralizes the acid in its immediate vicinity. The bacteria also express at least five types of adhesins, which enable them to adhere to the stomach epithelium, and they produce several cytotoxins that destroy the stomach epithelial cells, creating painful ulcers. The resulting chronic inflammation promotes cell proliferation and thus predisposes the infected individual to stomach cancer.

Bordetella pertussis, the bacterium that causes whooping cough, provides another remarkable example of active colonization. The first step in a *B. pertussis* infection is colonization of the respiratory epithelium. The bacteria circumvent the normal clearance mechanism that clears the respiratory tract (the *mucociliary escalator* described in Chapter 23) by binding tightly to the surface of the ciliated cells that line the tract and multiplying on them. *B. pertussis* expresses at least four adhesins that bind tightly to particular glycolipids on the ciliated cells. The adherent bacteria produce various toxins that eventually kill the ciliated cells, compromising the host's ability to clear the infection. The most familiar of these is *pertussis toxin*, which—like cholera toxin—is an ADP-ribosylating enzyme. It ADP-ribosylates the α subunit of the G protein G_i, inhibiting the G protein from suppressing the activity of the host cell's adenylyl cyclase, thereby increasing the production of cyclic AMP (discussed in Chapter 15). This toxin also interferes with the chemotactic pathway that neutrophils use to seek out and destroy invading bacteria (see Figure 16–101). Not content with this, *B. pertussis* also produces an adenylyl cyclase of its own, which is active only when bound to the eucaryotic Ca^{2+}-binding protein calmodulin in the cytoplasm of the host cell. Although both *B. pertussis* and *V. cholerae* drastically increase cAMP levels in the host cells to which they adhere, the symptoms of the diseases differ because the two bacteria colonize different sites in the host: *B. pertussis* colonizes the respiratory tract and causes paroxysmal coughing, whereas *V. cholerae* colonizes the gut and causes watery diarrhea.

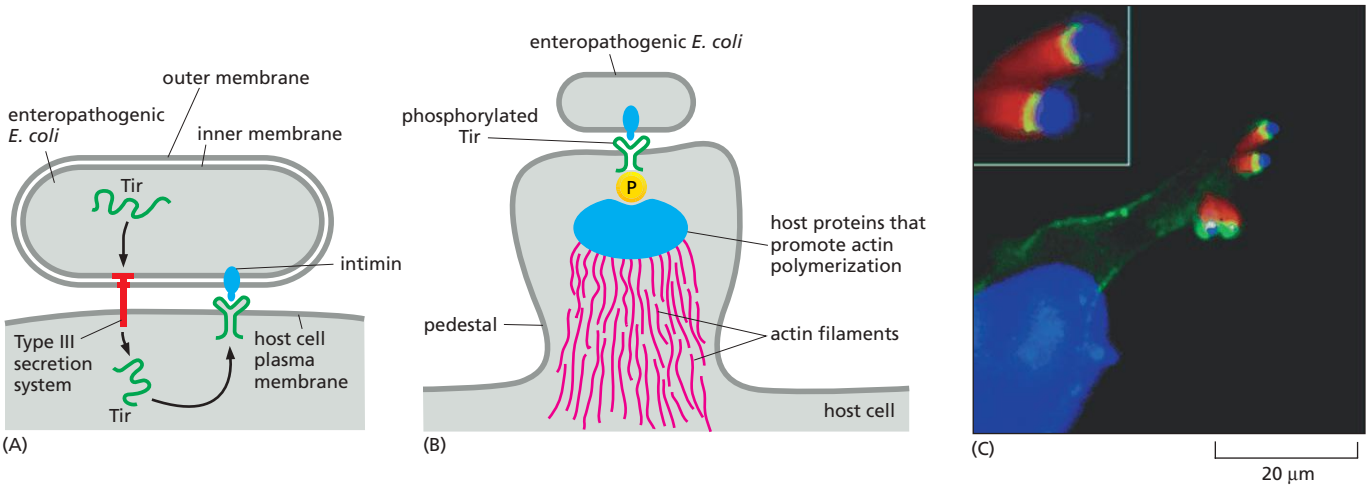


Figure 24–22 Interaction of enteropathogenic *E. coli* (EPEC) with host cells in the gut. (A) When EPEC contacts an epithelial cell in the lining of the human gut, it delivers a bacterial protein called Tir into the host cell through a type III secretion system. Tir then inserts into the plasma membrane of the host cell, where it functions as a receptor for the bacterial adhesin protein intimin. (B) A host cell protein tyrosine kinase phosphorylates the intracellular domain of Tir on tyrosines. Phosphorylated Tir then recruits various host cell proteins in turn, which trigger actin polymerization. Consequently, a dense accumulation of actin filaments assembles underneath the bacterium, forming an actin pedestal. (C) EPEC on a pedestal. In this fluorescence micrograph, the DNA of the EPEC and host cell are labeled in *blue*, Tir protein is labeled in *green*, and host cell actin filaments are labeled in *red*. The inset shows a close up view of the two upper bacteria on pedestals. (C, from D. Goosney et al., *Annu. Rev. Cell Dev. Biol.* 16:173–189, 2000. With permission from Annual Reviews.)

Not all examples of specific colonization require that the bacterium express adhesins that bind to host cell glycolipids or proteins. Enteropathogenic *E. coli*, which causes diarrhea in young children, instead uses a type III secretion system (see Figure 24–8) to deliver its own bacterially produced receptor protein (called Tir) into its host cell (**Figure 24–22**A). After Tir inserts into the host cell's plasma membrane, a bacterial surface protein binds to the extracellular domain of Tir, triggering a remarkable series of events inside the host cell. First, a host protein tyrosine kinase phosphorylates the Tir receptor protein on tyrosines, which is unusual because bacteria generally do not phosphorylate tyrosines on proteins. The phosphorylated Tir is then thought to recruit a member of the Rho family of small GTPases, which promotes actin polymerization through a series of intermediate steps (discussed in Chapter 16). The polymerized actin then forms a unique cell-surface protrusion, called a *pedestal*, that pushes the tightly adherent bacteria up about 10 μm from the host cell surface (Figure 24–22B, C).

These examples of host colonization illustrate the importance of host–pathogen communication in both the infection process and its evolution. Pathogenic organisms have acquired genes that encode proteins that interact specifically with particular molecules of the host cells. In some cases, such as the *B. pertussis* adenylyl cyclase, an ancestor of the pathogen may have acquired the cyclase gene from its host, whereas in others, such as Tir, random mutation may have produced protein motifs that are recognized by a eucaryotic protein tyrosine kinase.

Intracellular Pathogens Have Mechanisms for Both Entering and Leaving Host Cells

Many pathogens, including *V. cholerae* and *B. pertussis*, infect their host without entering host cells; they are referred to as *extracellular pathogens*. Others, however, including all viruses and many bacteria and protozoa, are **intracellular pathogens**. Their preferred niche for replication and survival is within the cytosol or intracellular compartments of particular host cells. This strategy has several advantages. The pathogens are not accessible to *antibodies* (discussed in

Chapter 25), and they are not easy targets for phagocytic cells (discussed later); furthermore, they are bathed in a rich source of many sugars, amino acids, and other nutrients present in host cell cytoplasm. This lifestyle, however, requires that the pathogen develop mechanisms for entering host cells, for finding a suitable subcellular niche where it can replicate, and for exiting from the infected cell to spread the infection. In the remainder of this section, we consider some of the myriad ways that individual intracellular pathogens exploit and modify host cell biology to satisfy these requirements.

Virus Particles Bind to Molecules Displayed on the Host Cell Surface

The first step for any intracellular pathogen is to bind to the surface of the host target cell. Viruses accomplish this binding through the association of a viral surface protein with a specific receptor on the host cell surface. Of course, no host cell receptor evolved for the sole purpose of allowing a pathogen to bind to it; these receptors all have other functions. The first such "virus receptor" identified was the *E. coli* surface protein that allows the bacteriophage lambda to bind to the bacterium. Its normal function is as a transport protein responsible for the uptake of maltose. Receptors need not be proteins, however; herpes simplex virus, for example, binds to heparan sulfate proteoglycans through specific viral membrane proteins.

Virions that infect animal cells generally use cell-surface receptor molecules that are either very abundant (such as sialic-acid-containing oligosaccharides, which are used by the influenza virus) or uniquely found on those cell types in which the virions can replicate (such as the nerve growth factor receptor, the nicotinic acetylcholine receptor, or the cell–cell adhesion protein N-CAM, all of which are used by rabies virus to infect neurons specifically). Often, many types of virus use a single type of receptor, and some viruses can use several different receptors. Moreover, different viruses that infect the same cell type may each use a different receptor. For example, members of at least six virus families, all of which preferentially replicate in liver cells (hepatocytes), cause hepatitis. Receptors for four of these have been identified, and they all differ. Many virions bind to receptors expressed on cells of the immune system. While seemingly paradoxic, as we might expect that triggering an immune response does not enhance viral survival, invading an immune cell may be a useful way to travel around the body and be taken to lymphoid organs, which are filled with other immune system cells.

Virions often require both a primary receptor and a secondary co-receptor for efficient attachment and entry into host cells. An important example is the AIDS virus *HIV*. Its primary receptor is CD4, a cell-surface protein on helper T cells and macrophages that is involved in immune recognition (discussed in Chapter 25). Viral entry also requires a co-receptor, either CCR5 (a receptor for β-chemokines) or CXCR4 (a receptor for α-chemokines), depending on the particular variant of the virus (**Figure 24–23**). Macrophages are susceptible only to HIV

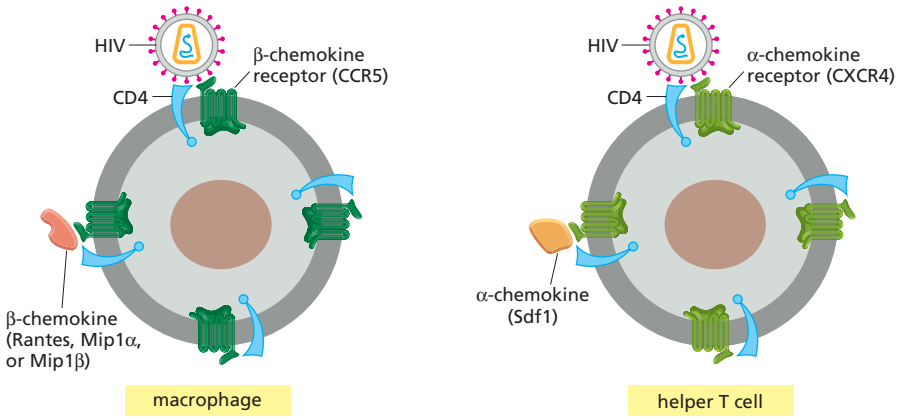
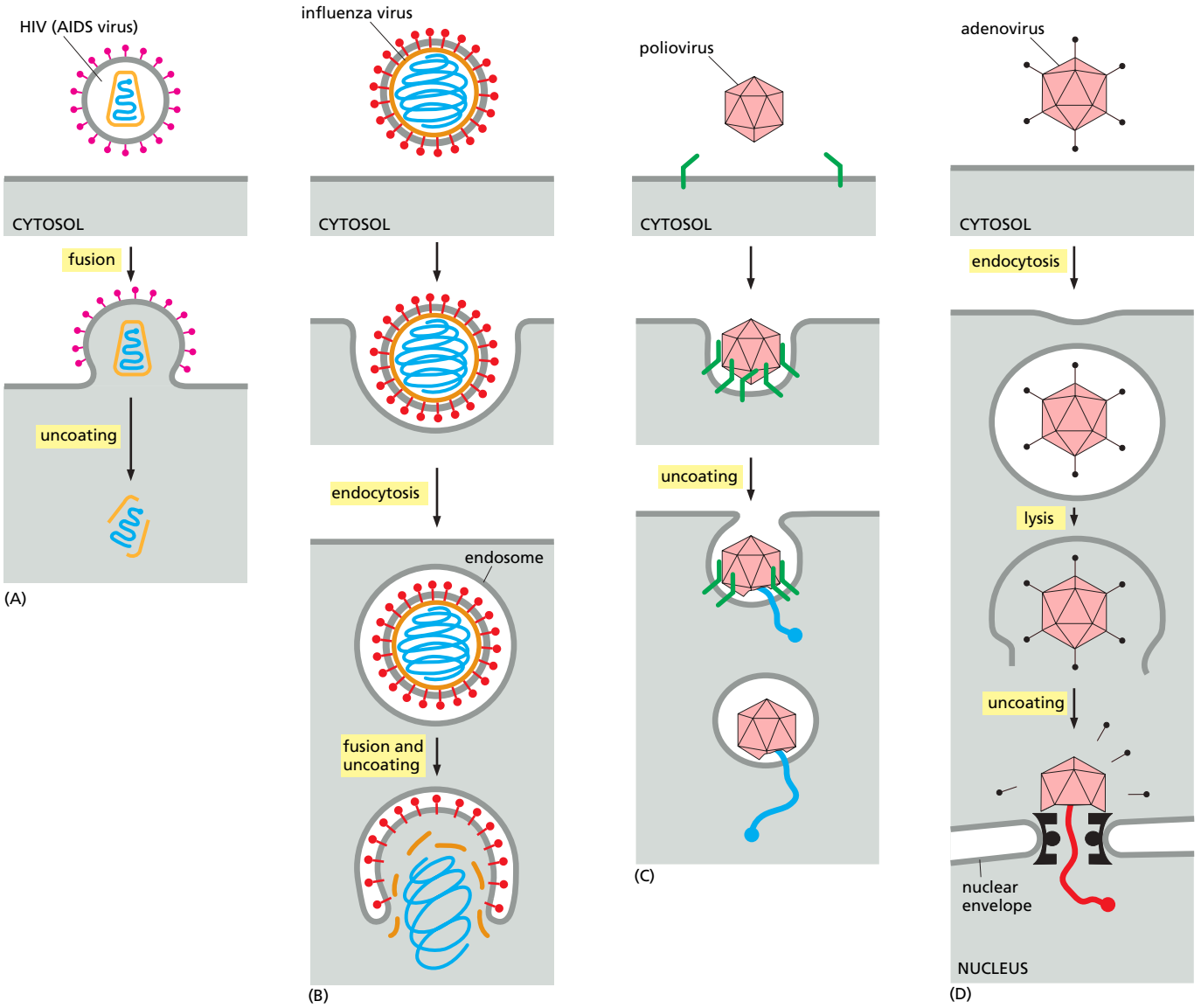


Figure 24–23 Receptor and co-receptors for HIV. All strains of HIV require the CD4 protein as a primary receptor. Early in an infection, most of the viruses use CCR5 as a co-receptor, allowing them to infect macrophages and their precursors, monocytes. As the infection progresses, mutant variants arise that now use CXCR4 as a co-receptor, enabling them to infect helper T cells efficiently. The natural ligand for the chemokine receptors (Sdf1 for CXCR4; Rantes, Mip1α, or Mip1β for CCR5) blocks co-receptor function and prevents viral invasion.



variants that use CCR5 for entry, whereas helper T cells are most efficiently infected by variants that use CXCR4. The viruses that are found within the first few months after HIV infection almost invariably require CCR5, which presumably explains why individuals fortunate enough to carry a defective *Ccr5* gene are not susceptible to HIV infection. In the later stages of infection, viruses may either switch to use the CXCR4 co-receptor or adapt to use both co-receptors; in this way, the virus can change the cell types it infects as the disease progresses.

Virions Enter Host Cells by Membrane Fusion, Pore Formation, or Membrane Disruption

After recognition and attachment to the host cell surface, the typical next steps for a virion are to enter the host cell and release its nucleic acid genome from its protective protein coat or lipid envelope. In most cases, the liberated nucleic acid remains complexed with some viral proteins. **Enveloped viruses** enter the host cell by fusing either with the plasma membrane or with the endosomal membrane following endocytosis (**Figure 24–24**A,B). Fusion is thought to proceed via a mechanism similar to SNARE-mediated fusion of intracellular vesicles during normal vesicular traffic (discussed in Chapter 13). <ATAG>

The virus regulates fusion both to ensure that its particles fuse only with the appropriate host cell membrane and to prevent the particles from fusing with

Figure 24–24 Four virus uncoating strategies. (A) Some enveloped viruses, such as HIV, fuse directly with the host cell plasma membrane to release their genome *(blue)* and capsid proteins *(orange)* into the cytosol. (B) Other enveloped viruses, such as influenza virus, first bind to cell-surface receptors, triggering receptor-mediated endocytosis. When the endosome acidifies, the virus envelope fuses with the endosomal membrane, releasing the viral genome *(blue)* and capsid proteins *(orange)* into the cytosol. (C) Poliovirus, a nonenveloped virus, binds to a receptor *(green)* on the host cell surface and then forms a pore (not shown) in the host cell membrane to extrude its RNA genome *(blue)*. (D) Adenovirus, another nonenveloped virus, uses a more complicated strategy. It induces receptor-mediated endocytosis and then disrupts the endosomal membrane, releasing part of the capsid and its DNA genome into the cytosol. The trimmed-down virus eventually docks onto a nuclear pore and releases its DNA *(red)* directly into the nucleus.

one another. For viruses such as HIV that fuse at neutral pH at the plasma membrane, binding to receptors or co-receptors usually triggers a conformational change in the viral envelope protein to expose a normally buried fusion peptide (see Figure 13–16). Other enveloped viruses, such as influenza, only fuse with a host cell membrane after endocytosis; in this case, it is frequently the acid environment in the early endosome that triggers the conformational change in a viral surface protein that exposes the fusion peptide (see Figure 24–24B). The H^+ pumped into the early endosome enters the influenza particle through an ion channel and triggers the *uncoating* of the viral RNA, which is directly released into the cytosol as the virus fuses with the endosomal membrane. For some viruses, uncoating occurs after release into the cytosol. In the case of Semliki forest virus, for example, the binding of host ribosomes to the capsid causes the capsid proteins to separate from the viral genome.

It is more difficult to envision how **nonenveloped viruses** enter host cells, as it is not obvious how large assemblies of protein and nucleic acid can cross the plasma or endosomal membrane. Where the entry mechanism is understood, nonenveloped viruses generally either form a pore in the cell membrane to deliver the viral genome into the cytoplasm or they disrupt the endosomal membrane after endocytosis.

Poliovirus uses the first strategy. Binding of poliovirus to its receptor triggers both receptor-mediated endocytosis and a conformational change in the viral particle. The conformational change exposes a hydrophobic projection on one of the capsid proteins, which apparently inserts into the endosomal membrane to form a pore. The viral genome then enters the cytoplasm through the pore, leaving the capsid either in the endosome on the cell surface, or in both places (see Figure 24–24C).

Adenovirus uses the second strategy. It is initially taken up by receptor-mediated endocytosis. As the endosome matures and becomes more acidic, the virus undergoes multiple uncoating steps that remove structural proteins sequentially from the capsid. Some of these steps require the action of a viral protease, which is inactive in the extracellular virus particle (probably because of intrachain disulfide bonds) but is activated in the reducing environment of the endosome. One of the proteins released from the capsid lyses the endosomal membrane, releasing the remainder of the virus into the cytosol. This trimmed-down virus then docks onto the nuclear pore complex, and the viral DNA genome is transported through the pore into the nucleus, where it is transcribed (see Figure 24–24D).

In these various entry strategies, viruses exploit a variety of host cell molecules and processes, which can include cell-surface components, receptor-mediated endocytosis, endosomal maturation steps, and nuclear transport. These strategies illustrate again the sophisticated ways that pathogens have evolved to utilize the basic cell biology of their hosts.

Bacteria Enter Host Cells by Phagocytosis

Bacteria are much larger than viruses, and they are too large to be taken up either through pores or by receptor-mediated endocytosis. Instead, they enter host cells by phagocytosis. Phagocytosis of bacteria is a normal function of macrophages. They patrol the tissues of the body and ingest and destroy unwanted microbes. Some pathogens, however, have acquired the ability to survive and replicate within macrophages after they have been phagocytosed.

Mycobacterium tuberculosis is one such pathogen. As discussed earlier, it causes **tuberculosis**, a serious lung infection that is widespread in some urban populations. It is usually acquired by inhalation of the bacterium into the lungs, where it is phagocytosed by alveolar macrophages. Although the microbe can survive and replicate within macrophages, with the help of the adaptive immune system, the macrophages of most healthy individuals contain the infection within a lesion called a *tubercle.* In most cases, the lesion becomes walled off by a fibrous capsule that eventually undergoes calcification and can then easily be seen on an X-ray of the lungs. Remarkably, *M. tuberculosis* in such

lesions can survive for decades, and, later in life, especially if drugs or disease weaken the immune system, the infection may be reactivated and spread in the lung and even to other organs.

Tuberculosis has infected human populations for thousands of years, but another bacterium that lives within alveolar macrophages was first recognized as a human pathogen only in 1976. *Legionella pneumophila* is normally a parasite of freshwater amoebae, which take it up by phagocytosis. When droplets of water containing *L. pneumophila* or infected amoebae are inhaled into the lung, the bacteria can invade and live inside alveolar macrophages (**Figure 24–25**), which, to the bacteria, must seem just like large amoebae. This infection leads to the type of pneumonia known as **Legionnaire's disease**. The pathogen is commonly spread by central air-conditioning systems, as the amoebae that are the bacterium's normal host are particularly adept at colonizing air-conditioning cooling towers; moreover, these cooling systems produce microdroplets of water that are easily inhaled. The incidence of Legionnaire's disease has increased dramatically in recent decades, with outbreaks frequently traced to the air-conditioning systems in office buildings, hospitals, and hotels. Other forms of modern aerosolization are sometimes responsible, including decorative fountains and produce sprayers in supermarkets.

Some bacteria invade cells that are normally nonphagocytic. One way in which bacteria can induce such a cell to phagocytose them is by expressing an adhesin that binds with high affinity to a host cell adhesion protein that the host cell normally uses to adhere to another host cell or to the extracellular matrix (discussed in Chapter 19). For example, a bacterium that causes diarrhea, *Yersinia pseudotuberculosis* (a close relative of the plague bacterium *Yersinia pestis*), expresses a protein called *invasin* that binds to β1 integrins, and a bacterium that causes a rare but serious form of food poisoning, *Listeria monocytogenes,* expresses a protein that binds to E-cadherin. Binding of the bacterial proteins to these transmembrane host adhesion proteins fools the host cell into attempting to form a cell junction, and it begins moving actin and other cytoskeletal components to the site of bacterial attachment. Since the bacterium is small relative to the host cell, the host cell's attempt to spread over the adhesive surface of the bacterium results in the phagocytic uptake of the bacterium—a process known as the *zipper mechanism* of invasion (**Figure 24–26**A). The similarity of this form of invasion to the natural process of cell adhesion was revealed when the three-dimensional structure of invasin was determined. The bacterial protein has an RGD motif with a structure almost identical to the RGD motif of the integrin-binding site in the extracellular matrix protein laminin (discussed in Chapter 19).

A second pathway by which bacteria can invade nonphagocytic cells is known as the *trigger mechanism* (Figure 24–26B). It is used by various pathogens, including *Salmonella enterica*, which causes food poisoning. When the bacterium injects a set of effector molecules into the host cell cytoplasm through a type III secretion system, it initiates this dramatic form of invasion. Some of these effector molecules activate Rho-family GTPases, which stimulate actin polymerization (discussed in Chapter 16). Others interact with cytoskeletal elements more directly, severing actin filaments and causing the rearrangement of cross-linking proteins. The net effect is to cause dramatic localized ruffling on the surface of the host cell (Figure 24–26C), which throws up large actin-rich protrusions that fold over and trap the bacterium within a large endocytic vesicle called a *macropinosome* (Figure 24–26D). The overall appearance of cells being invaded by the trigger mechanism is similar to the dramatic ruffling induced by some extracellular growth factors, suggesting that similar intracellular signaling pathways are activated in both cases.



Figure 24–25 Uptake of *Legionella pneumophila* by a human phagocyte. This electron micrograph shows the unusual coil structure induced on the surface of the phagocyte by the bacterium. Some other pathogens, including the bacterium *Borrelia burgdorferi*, which causes Lyme disease, the eucaryotic pathogen *Leishmania*, and the yeast *Candida albicans*, may also sometimes invade cells using this type of coiling phagocytosis. (From M.A. Horwitz, *Cell* 36:27–33, 1984. With permission from Elsevier.)

Intracellular Eucaryotic Parasites Actively Invade Host Cells

The host cell supplies the energy required for the uptake of viruses by receptor-mediated endocytosis and bacteria by phagocytosis or macropinocytosis. The pathogen is a relatively passive participant, usually providing a trigger to initiate the invasion process. In contrast, intracellular eucaryotic parasites, which are

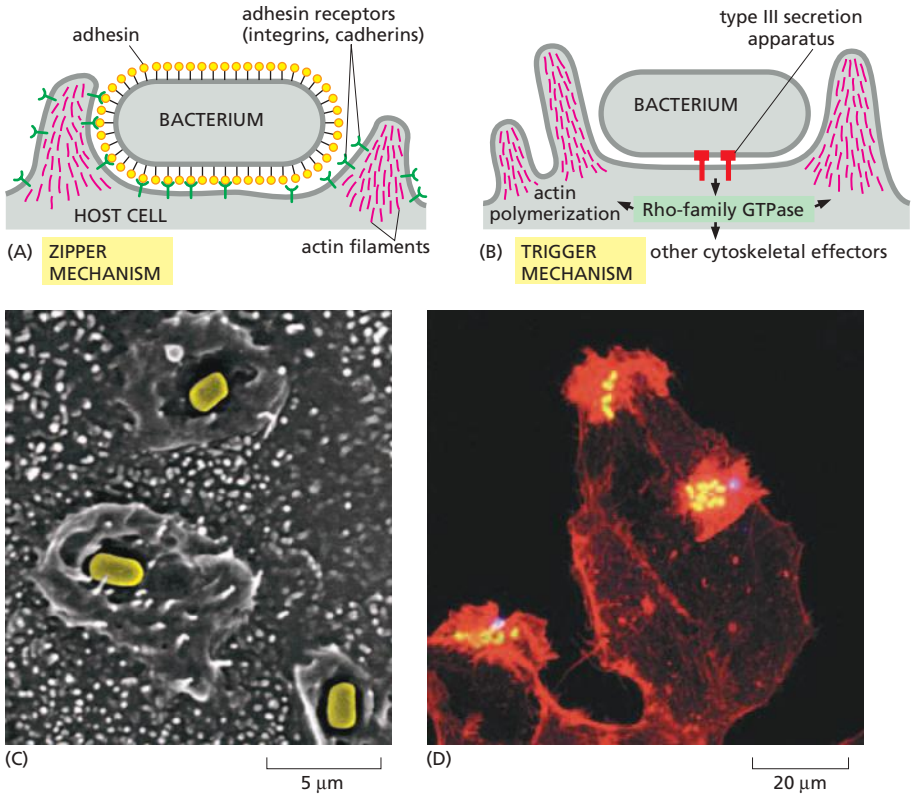


Figure 24–26 Mechanisms used by bacteria to induce phagocytosis by nonphagocytic host cells. (A) The zipper and (B) trigger mechanisms for pathogen-induced phagocytosis both require the polymerization of actin at the site of bacterial entry. (C) A scanning electron micrograph showing a very early stage of *Salmonella enterica* invasion by the trigger mechanism. Each of the three bacteria (pseudocolored *yellow*) is surrounded by a small membrane ruffle. (D) Fluorescence micrograph showing that the large ruffles that engulf the *Salmonella* bacteria are actin-rich. The bacteria are labeled in *green* and actin filaments in *red*; because of the color overlap, the bacteria appear *yellow*. (From J.E. Galan, *Annu. Rev. Cell Dev. Biol.* 17:53–86, 2001. With permission from Annual Reviews.)

typically much larger than bacteria, invade host cells through a variety of complex pathways that usually require significant energy expenditure on the part of the parasite.

Toxoplasma gondii, the cat parasite that also causes occasional serious human infections, is an example. When this protozoan contacts a host cell, it protrudes an unusual microtubule-based structure called a *conoid*, which it uses to push its way slowly into the host cell. The energy for invasion seems to come entirely from the parasite and requires at least one highly unusual myosin (Class XIV; see Figure 16–57); depolymerizing the actin cytoskeleton in the parasite, but not in the host cell, disrupts this process. As the parasite moves into the host cell, a membrane derived from the invaginated host cell plasma membrane surrounds it. Remarkably, the parasite somehow removes host transmembrane proteins from the surrounding membrane as it forms, so that the parasite is protected in a membrane-enclosed compartment, which does not fuse with lysosomes and does not participate in host cell membrane trafficking processes (**Figure 24–27**). The specialized membrane allows the parasite to take up metabolic intermediates and nutrients from the host cell's cytosol but excludes larger molecules. Malaria parasites invade red blood cells by using a very similar mechanism.

The protozoan *Trypanosoma cruzi*, which causes multiorgan Chagas disease, mainly in Mexico and Central and South America, uses an entirely different, but no less peculiar, invasion strategy. After attachment to host cell surface receptors, this parasite induces a local elevation of Ca^{2+} in the host cell's cytosol. The Ca^{2+} signal recruits lysosomes to the site of parasite attachment, which fuse with the host cell's plasma membrane during the internalization process, allowing the parasites rapid access to the lysosomal compartment (**Figure 24–28**). As we discuss below, most intracellar pathogens go to great lengths to avoid exposure to the hostile, proteolytic environment of the lysosome, but *Trypanosoma cruzi* uses the lysosome as its means of entry. In the lysosomal compartment, the parasite secretes an enzyme that removes sialic acid from lysosomal glycoproteins and transfers it to its own surface molecules, thereby coating itself with host cell sugars. Next, the parasite secretes a pore-forming toxin that lyses the lysosome membrane, releasing the parasite into the host cell's cytosol, where it proliferates.

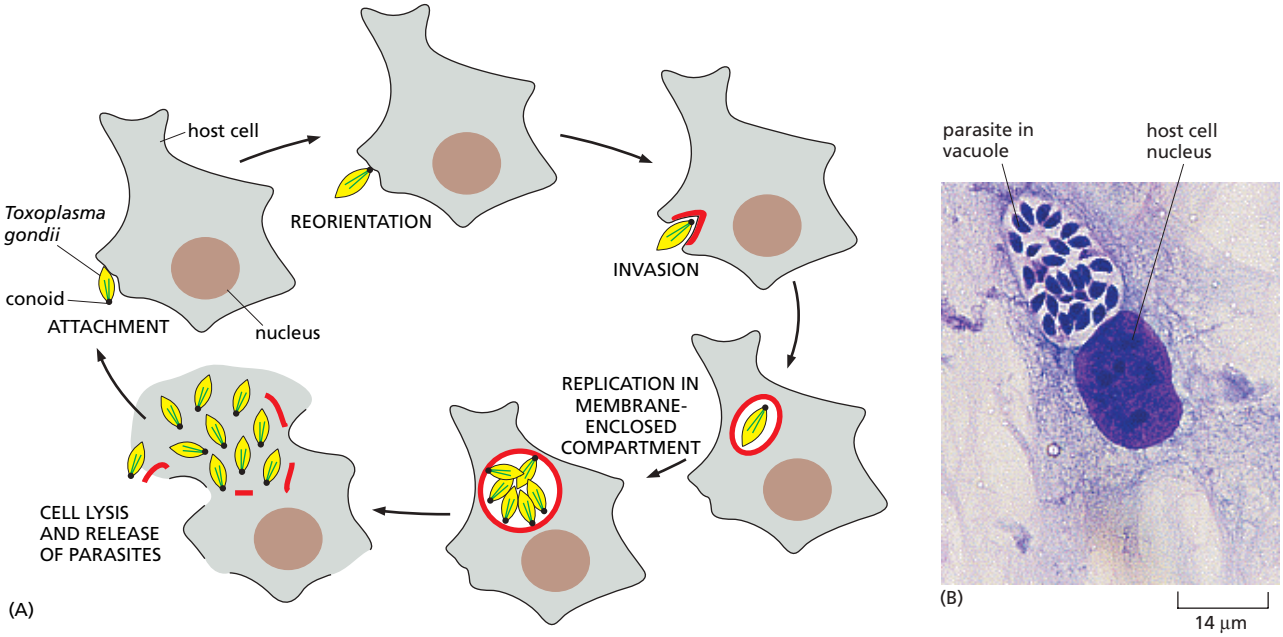
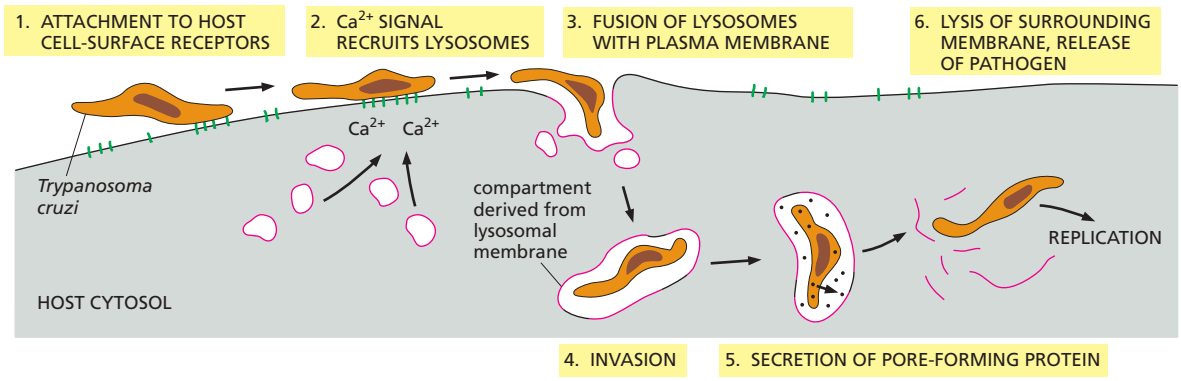


Figure 24–27 The life cycle of the intracellular parasite *Toxoplasma gondii*. (A) After attachment to a host cell, *T. gondii* uses its conoid to push its way into the host cell. As the host cell's plasma membrane invaginates to surround the parasite, the parasite somehow removes the host cell proteins associated with normal endosomes or phagosomes, so that the compartment (shown in *red*) does not fuse with lysosomes. After several rounds of replication, the parasite causes the compartment to break down and the host cell to lyse, releasing the progeny into the extracellular space from which they can infect other host cells. (B) Light micrograph of *T. gondii* replicating within a membrane-enclosed compartment in a cultured cell. (B, courtesy of Manuel Camps and John Boothroyd.)

The *microsporidia* use perhaps the most bizarre active invasion mechanism. These tiny, obligate, intracellular, eucaryotic parasites are only about 5 µm long and have among the smallest known genomes for a eucaryotic cell, only 2,900,000 nucleotide pairs. Normally, microsporidia cause disease primarily in insects, but they can also cause opportunistic infections in people with AIDS. Having adapted over a long period to a parasitic lifestyle, they depend on their host cells for some metabolic functions and have lost many of the genes and cell structures required for a free-living existence; for example, they no longer have mitochondria or peroxisomes. However, they do have a strange extrusion apparatus, the polar tube, that enables them to invade host cells. In the environmentally resistant spore stage of its life cycle, the polar tube is wound in a coil around the nucleus (**Figure 24–29**A). On contact with an appropriate host cell, the polar tube discharges explosively, uncoiling in less than 2 seconds to form a mature structure that can be more than ten times the length of the spore. The tip of the discharging polar tube, traveling at a speed of 100 µm/sec, penetrates the host cell and delivers (apparently by osmotic pressure) the internal contents of the spore, including the microsporidian's nucleus, into the cytoplasm of the host cell, where the parasite replicates to form up to a hundred progeny (Figure 24–29B and C). Eventually, the progeny mature into spores, and the host cell

Figure 24–28 Invasion of host cells by *Trypanosoma cruzi*. This parasite recruits host cell lysosomes to its site of attachment. The lysosomes fuse with the invaginating plasma membrane to create an intracellular compartment constructed almost entirely of lysosomal membrane. After a brief stay in the compartment, the parasite secretes a pore-forming protein that disrupts the surrounding membrane, allowing the parasite to escape into the host cell cytosol and proliferate.



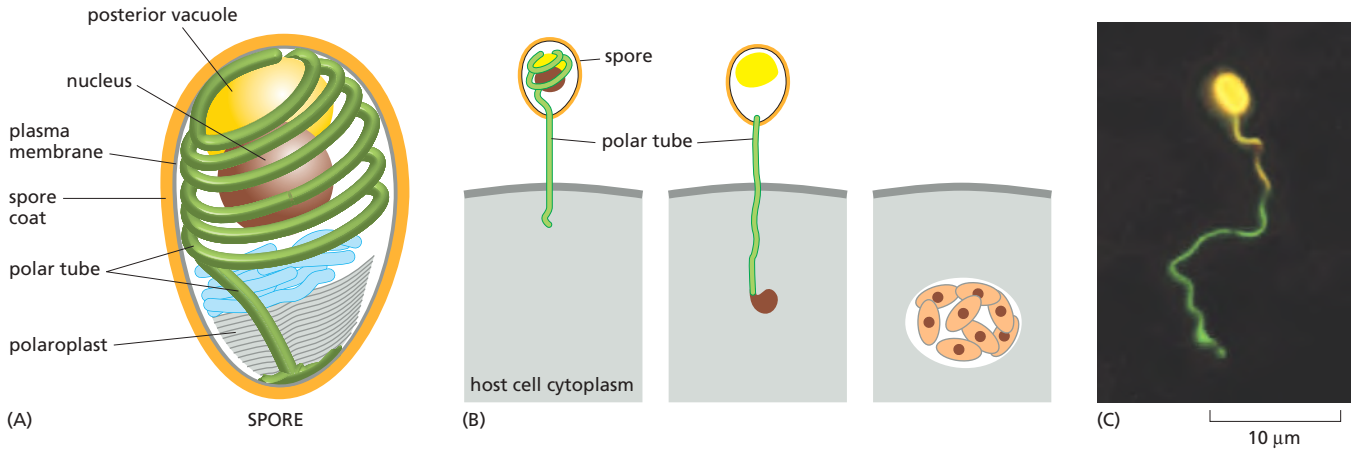


Figure 24–29 Invasion of host cells by microsporidia. (A) The spore form of the parasite is covered with a rigid coat and harbors a coiled polar tube that wraps many times around the nucleus. (B) The polar tube extends explosively when the spore comes into contact with an appropriate host cell, penetrating the host cell and delivering the nucleus and other contents of the spore into the cytoplasm of the host cell. The microsporidian then proliferates in the host cell. (C) Immunofluorescence micrograph of a spore of the microsporidian *Encephalitozoon cuniculi*. The parts of the spore that are outside the host cell are stained *yellow,* and the parts that are inside the host cell are stained *green*. The discharged polar tube changes from yellow to green at the point where it has entered the host cell, which is not visible. Scale bar 10 μm. (C, from C. Franzen, *Trends Parasitol.* 20:275–279, 2004. With permission from Elsevier.)

lyses to release them. Microsporidia spores are sufficiently small for macrophages to phagocytose them. When this happens, the spores discharge their polar tube from within the confines of the phagosome, again delivering their contents into the cytosol of the host cell.

Many Pathogens Alter Membrane Traffic in the Host Cell

The three examples of intracellular parasites just discussed raise a general problem that faces all intracellular pathogens, including viruses, bacteria, and eucaryotic parasites. They must deal in some way with membrane traffic in the host cell. After endocytosis by a host cell, they usually find themselves in an endosomal compartment that normally would fuse with lysosomes to form a *phagolysosome*. They therefore must either modify the compartment to prevent its fusion with lysosomes, escape from the compartment before such fusion, escape after fusion but before getting digested, or find ways to survive in the hostile environment of the phagolysosome (**Figure 24–30**).

Most pathogens use the first or second strategy. As we have seen, *Trypanosoma cruzi* uses the escape route, as do essentially all viruses (see Figure 24–24). The bacterium *Listeria monocytogenes* also uses this strategy. It is taken up into cells via the zipper mechanism discussed earlier and secretes a protein called *listeriolysin O* that forms large pores in the phagosomal membrane, releasing the bacteria into the cytosol before they are digested. Once in the

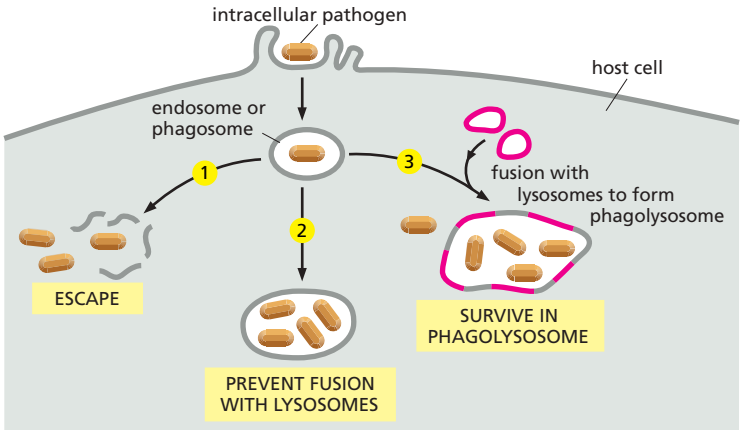


Figure 24–30 Choices that an intracellular pathogen faces. After entry into a host cell, generally through endocytosis or phagocytosis into a membrane-enclosed compartment, intracellular pathogens can use one of three strategies to survive and replicate. Pathogens that follow strategy (1) include all viruses, *Trypansoma cruzi*, *Listeria monocytogenes*, and *Shigella flexneri*. Those that follow strategy (2) include *Mycobacterium tuberculosis*, *Salmonella enterica*, *Legionella pneumophila*, and *Chlamydia trachomatis*. Those that follow strategy (3) include *Coxiella burnetii* and *Leishmania*.

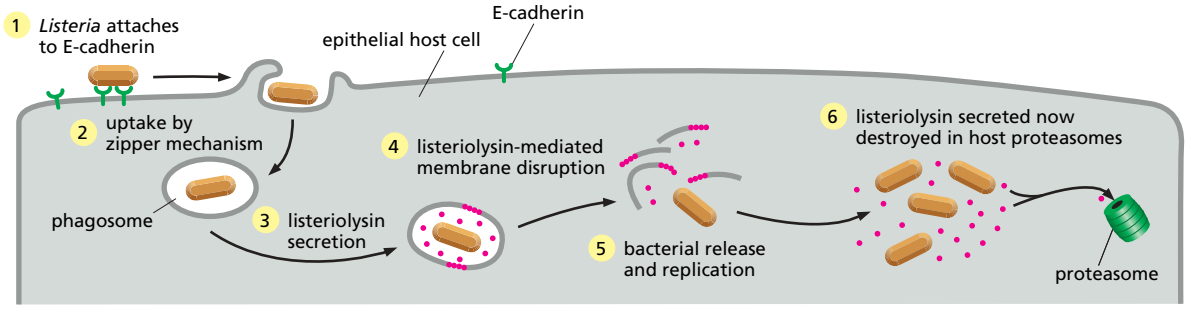


Figure 24–31 Selective destruction of the phagosomal membrane by *Listeria monocytogenes*. *L. monocytogenes* attaches to E-cadherin on the surface of epithelial cells and induces its own uptake by the zipper mechanism (see Figure 24–26A). Within the phagosome, the bacterium secretes the hydrophobic protein listeriolysin O, which forms oligomers in the host cell membrane, thereby creating large pores and eventually disrupting the membrane. Once in the host cell cytosol, the bacteria begin to replicate and continue to secrete listeriolysin O. Because the listeriolysin O in the cytosol is rapidly degraded by proteasomes, the host cell's plasma membrane remains intact.

cytosol, the bacteria continue to secrete listeriolysin O, but it does not destroy other cell plasma membranes for two reasons: first, it is 10 times more active at the acidic pH found in the phagosome than at neutral pH found in the cytosol; second, it is rapidly degraded in the cytosol by host cell proteasomes (see Figure 6–80), which do not have access to the listeriolysin O in the phagosome (**Figure 24–31**).

If a pathogen is to survive and replicate in a host cell membrane-enclosed compartment, it must modify membrane trafficking in the host cell, and it can use various methods to do this. It must modify the compartment in at least two ways: first, it must prevent lysosomal fusion, and second, it must provide a pathway for importing nutrients from the host cytosol. In addition, many pathogens (particularly viruses) alter membrane trafficking pathways to prevent presentation of their tell-tale foreign antigens on the host cell surface; otherwise, T cells could detect their presence and kill the host cell (discussed in Chapter 25).

Different pathogens have distinct strategies for altering host cell membrane traffic (**Figure 24–32**). As we have seen, *Toxoplasma gondii* creates a membrane-enclosed compartment that does not participate in normal host cell membrane traffic and yet specifically allows nutrient import. *Mycobacterium tuberculosis* somehow prevents the very early endosome that contains it from maturing, so that the endosome never acidifies or acquires the characteristics of a late endosome or lysosome. Endosomes containing *Salmonella enterica*, in contrast, do acidify and acquire markers of late endosomes, but they arrest their maturation at a stage prior to lysosomal fusion. Other bacteria seem to find shelter in intracellular compartments that are completely distinct from the usual endocytic pathway. *Legionella pneumophila*, for example, replicates in compartments that are enclosed by layers of rough endoplasmic reticulum

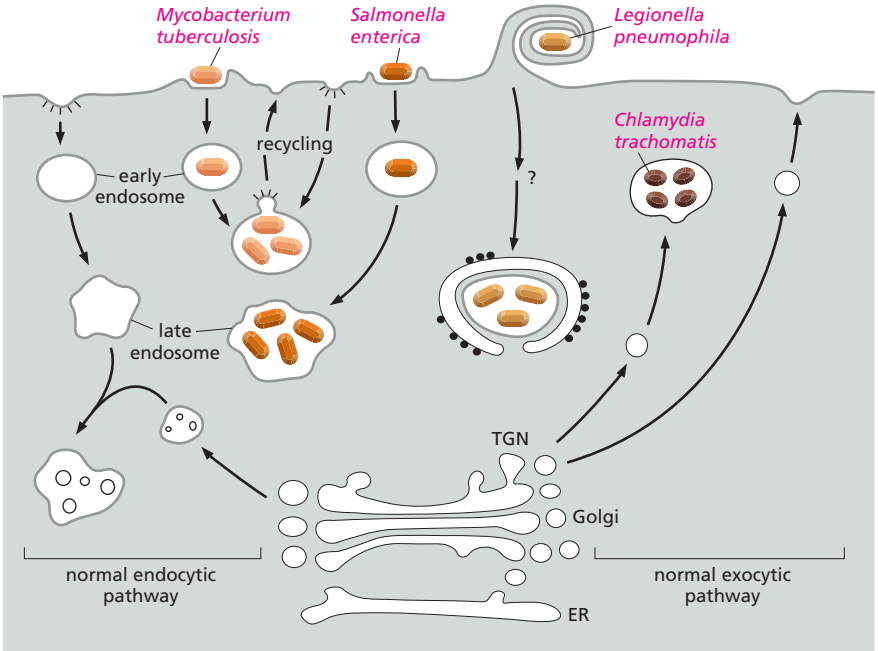


Figure 24–32 Modifications of host cell membrane trafficking by bacterial pathogens. Four intracellular bacterial pathogens, *Mycobacterium tuberculosis*, *Salmonella enterica*, *Legionella pneumophila*, and *Chlamydia trachomatis*, all replicate in membrane-enclosed compartments, but the four compartments differ. *M. tuberculosis* remains in a compartment that has early endosomal markers and continues to communicate with the plasma membrane via transport vesicles. *S. enterica* replicates in a compartment that has late endosomal markers and does not communicate with the plasma membrane. *L. pneumophila* replicates in an unusual compartment that is wrapped in several layers of rough endoplasmic reticulum (ER) membrane; only one layer is shown for simplicity. *C. trachomatis* replicates in an exocytic compartment that fuses with vesicles coming from the *trans* Golgi network (TGN).

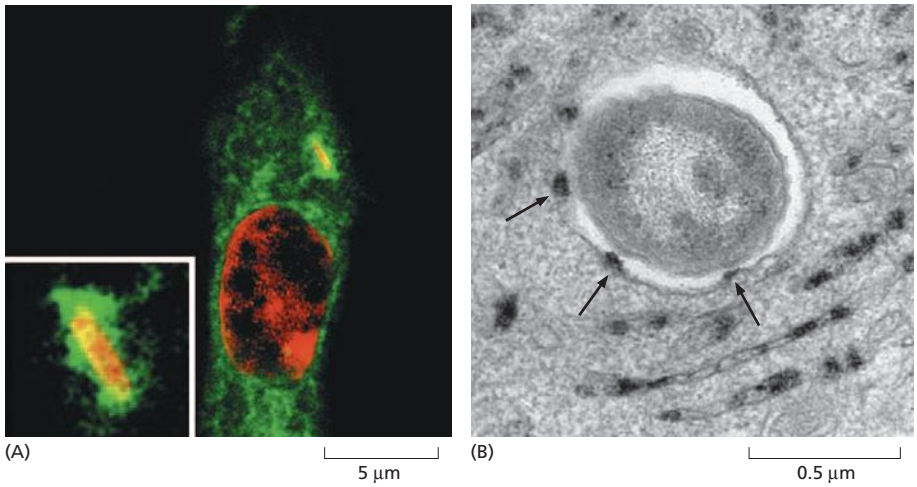
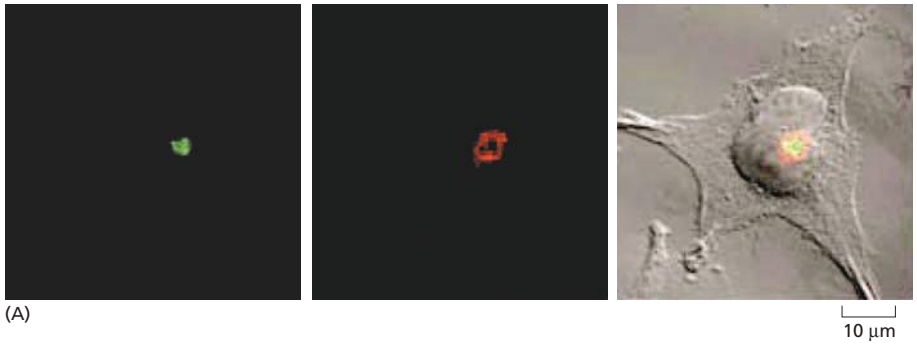


Figure 24–33 Association of host cell endoplasmic reticulum (ER) membrane with intracellular bacterial pathogens. (A) Shortly after infection with *Legionella pneumophila*, the ER *(green)* of the host cell is recruited to surround the intracellular bacterium. The DNA of the bacterium and of the host cell are labeled *red*. The inset shows the close association between the ER and the bacterium. (B) *Brucella abortus*, a Gram-negative bacterium that can cause abortions in cattle, also replicates within an ER-associated compartment. In this electron micrograph, the black smudges within the ER indicate the presence of the ER-specific enzyme glucose-6-phosphatase. The *black arrows* show concentrations of the enzyme on the membrane enclosing the bacterium, indicating that it has fused directly with the ER. (A, from J.C. Kagan and C.R. Roy, *Nat. Cell Biol*. 4:945–954, 2002. With permission from Macmillan Publishers Ltd; B, from J. Celli and J.P. Gorvel, *Curr. Opin. Microbiol*. 7:93–97, 2004. With permission from Elsevier.)

(**Figure 24–33**). *Chlamydia trachomatis*, a sexually transmitted bacterial pathogen that can cause sterility and blindness, replicates in a compartment that seems similar to part of the exocytic pathway. Some intracellular bacterial pathogens seem to be able to manipulate the location of other membrane-enclosed organelles not in direct physical contact with their own compartment. For example, *Salmonella*-containing late endosomes are normally found in very close apposition to the Golgi apparatus (**Figure 24–34**). The mechanisms used by these organisms to alter their membrane compartments and other aspects of membrane traffic in the host cell are still poorly understood.

Viruses also often alter membrane traffic in the host cell. Enveloped viruses must acquire their membrane from host cell phospholipids. In the simplest cases, virally encoded proteins are inserted into the ER membrane and follow the usual path through the Golgi apparatus to the plasma membrane, undergoing various post-translational modifications *en route*. The viral capsid and genome then assemble at the plasma membrane and bud off from the cell surface. This is one mechanism used by HIV. Other enveloped viruses interact in complex ways with membrane trafficking pathways in the host cell (**Figure 24–35**). Even some nonenveloped viruses alter membrane traffic in the host cell to suit their own purposes. For example, a membrane-associated, virus-encoded RNA polymerase carries out poliovirus replication. The replication proceeds more quickly if the surface area of host cell membrane is increased. To accomplish this, the virus induces increased membrane lipid synthesis in the host cell and blocks membrane transport from the ER. ER membrane thereby accumulates, expanding the surface area on which viral RNA replication can occur (**Figure 24–36**). Many viral pathogens and some bacterial pathogens are frequently found in association with autophagosomes, which form by autophagy (discussed



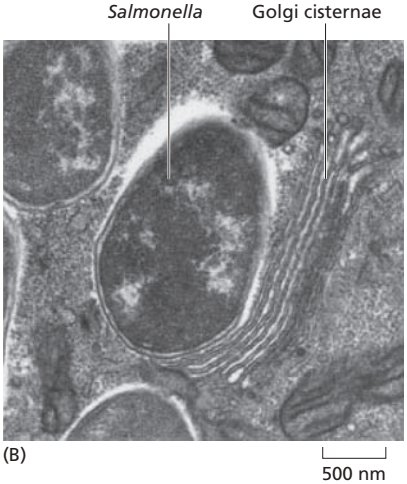


Figure 24–34 Proximity of Golgi stacks to endosomes containing *Salmonella enterica*. (A) In this infected cell, the bacteria are *green* and the Golgi has been labeled (*red*) with antibodies against a Golgi-specific structural protein. (B) Electron micrograph showing the very close apposition of a Golgi stack to a late endosome containing the bacterium. (From S.P. Salcedo and D.W. Holden, *EMBO J.* 22:5003–5014, 2003. With permission from Macmillan Publishers Ltd.)

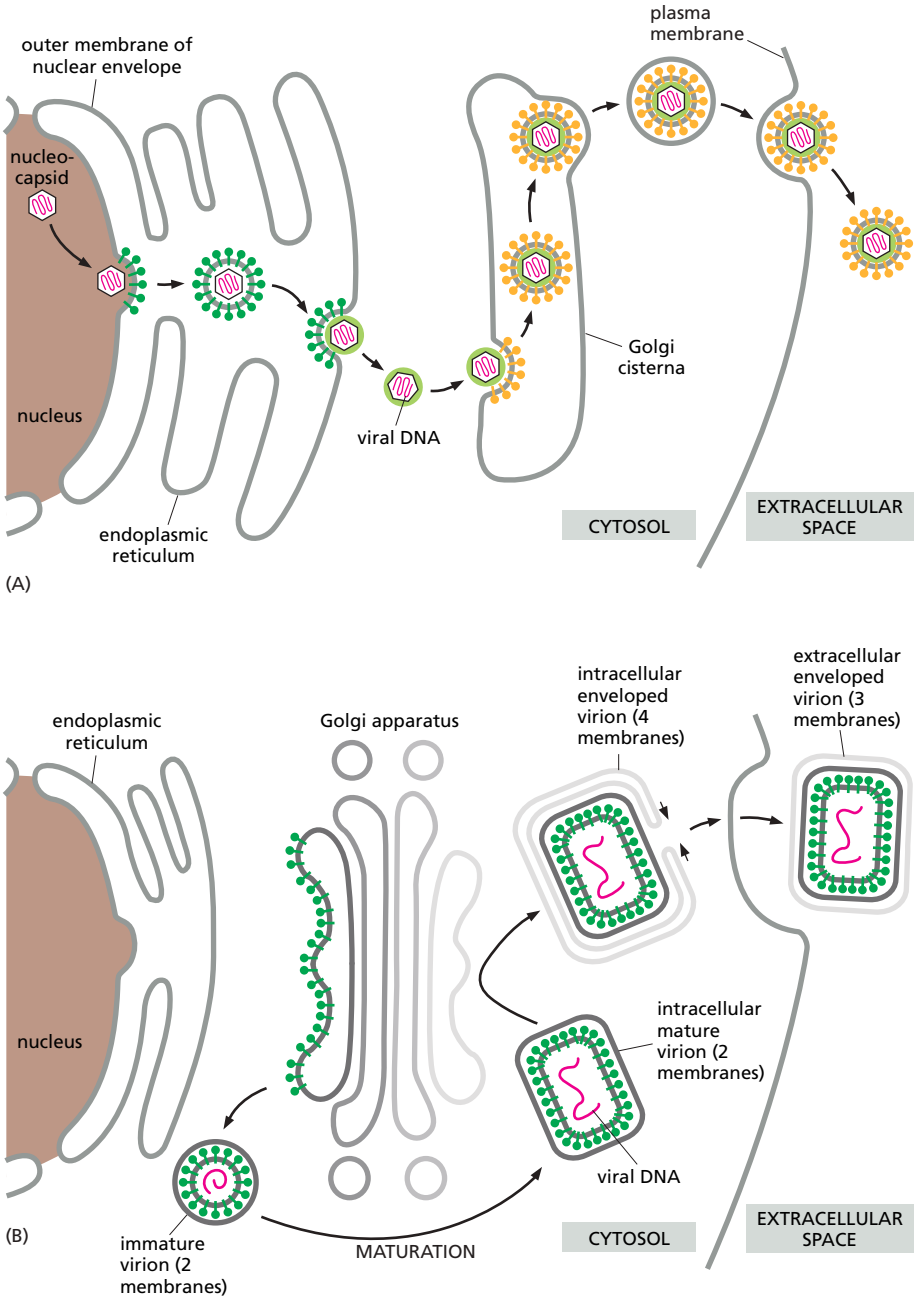


Figure 24–35 Complicated strategies for viral envelope acquisition. (A) Herpes virus nucleocapsids assemble in the nucleus and then bud through the inner nuclear membrane into the space between the inner and outer nuclear membranes, acquiring a membrane coat. The virus particles then apparently lose this coat when they fuse with the outer nuclear membrane to escape into the cytosol. Subsequently, the nucleocapsids bud into the Golgi apparatus and bud out again on the other side, acquiring two new membrane coats. The virus then buds from the cell with a single membrane when its outer membrane fuses with the plasma membrane. (B) Vaccinia virus (which is closely related to the virus that causes smallpox and is used to vaccinate against smallpox) assembles in "replication factories" within the cytosol, far away from the plasma membrane. The first structure that assembles contains two membranes, both acquired from the Golgi apparatus by a poorly defined wrapping mechanism. Some of these viral particles are then engulfed by the membranes of a second intracellular membrane-enclosed compartment; these viral particles have a total of four layers of membrane. After fusion at the plasma membrane, the virus escapes from the cell with three membrane layers.

in Chapter 13). In most cases, it is not clear whether the host cell initiates induction of autophagy as a protective response or the invading pathogen triggers it to assist in the pathogen's replication.

Viruses and Bacteria Use the Host Cell Cytoskeleton for Intracellular Movement

The cytoplasm of mammalian cells is extremely viscous. It is crowded with organelles and supported by networks of cytoskeletal filaments, all of which inhibit the diffusion of particles the size of a bacterium or a viral capsid. To reach a particular part of the cell to carry out part of its replication cycle, a pathogen must actively move there. As with transport of intracellular organelles, generally, pathogens use the host cell's cytoskeleton for active movement.

Several bacteria that replicate in the host cell's cytosol (rather than in membrane-enclosed compartments) have adopted a remarkable mechanism for moving, which depends on actin polymerization. These bacteria, including *Listeria monocytogenes*, *Shigella flexneri*, *Rickettsia rickettsii* (which causes Rocky

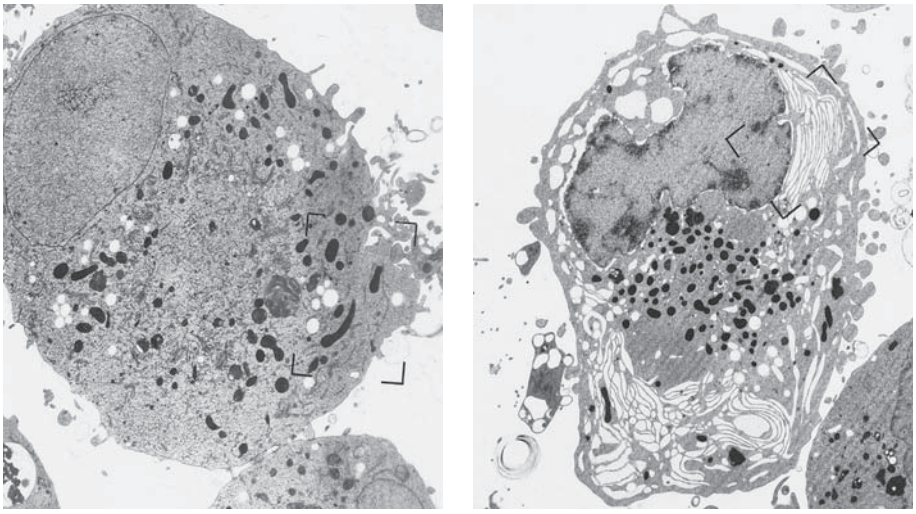
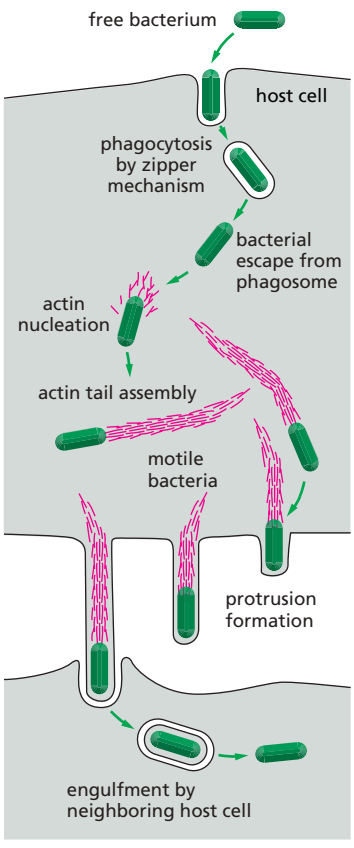
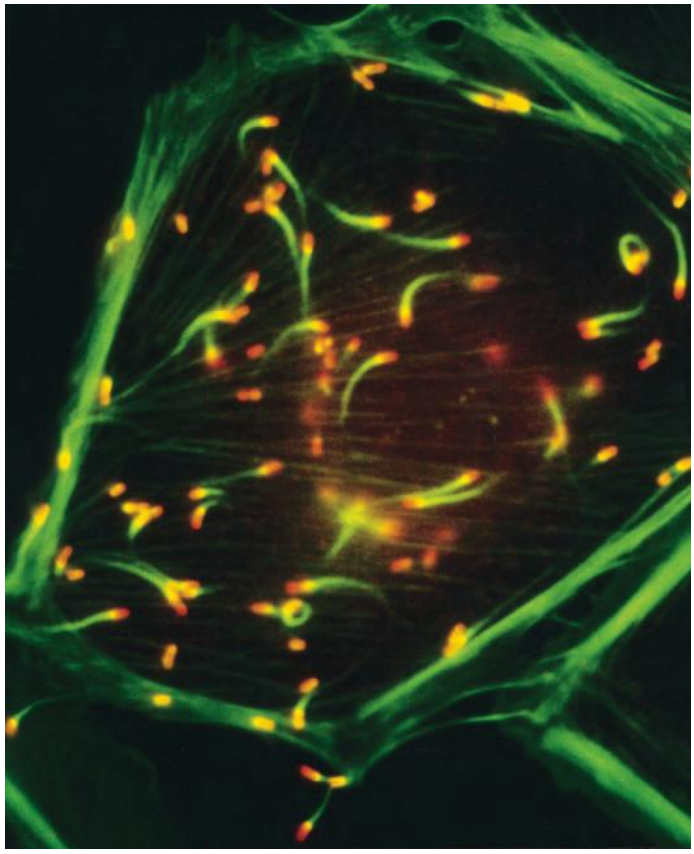


Figure 24–36 Intracellular membrane alterations induced by a poliovirus protein. Poliovirus, like other single-stranded RNA viruses, replicates its RNA genome by using a polymerase that associates with intracellular membranes. Several of the proteins encoded by its genome alter the structure or dynamic behavior of the membrane-enclosed organelles in the host cell. These electron micrographs show an untransfected monkey cell *(left)* and the same type of cell expressing a transgene encoding the 3A protein from poliovirus *(right)*. In the transfected cell, the ER is swollen because the transgene-encoded protein inhibits traffic from the ER to the Golgi apparatus. (From J.J.R. Doedens, T.H. Giddings Jr. and K. Kirkegaard, *J. Virol.* 71:9054–9064, 1997.)

Mountain spotted fever), *Burkholderia pseudomallei* (which causes melioidosis), and *Mycobacterium marinum* (a relative of the bacterium that causes tuberculosis), induce the nucleation and assembly of host cell actin filaments at one pole of the bacterium. The growing filaments generate substantial force and push the bacterium through the cytoplasm at rates of up to 1 μm/sec. New filaments form at the rear of each bacterium and are left behind like a rocket trail as the bacterium advances, depolymerizing again within a minute or so as they encounter depolymerizing factors in the cytosol. When a moving bacterium reaches the plasma membrane, it continues to move outward, inducing the formation of a long, thin host cell protrusion with the bacterium at its tip. A neighboring cell often engulfs this projection, allowing the bacterium to enter the neighbor's cytoplasm without exposure to the extracellular environment, thereby avoiding recognition by antibodies produced by the host's adaptive immune system (**Figure 24–37**).



(A)



(B)

Figure 24–37 The actin-based movement of *Listeria monocytogenes* within and between host cells. <GTAT> (A) These bacteria induce the assembly of actin-rich tails in the host cell cytoplasm, enabling the bacteria to move rapidly. Motile bacteria spread from cell to cell by forming membrane-enclosed protrusions that are engulfed by neighboring cells. (B) Fluorescence micrograph of the bacteria moving in a cell that has been stained to reveal bacteria in *red* and actin filaments in *green*. Note the comet-like tail of actin filaments behind each moving bacterium. Regions of overlap of red and green fluorescence appear *yellow*. (B, courtesy of Julie Theriot and Tim Mitchison.)

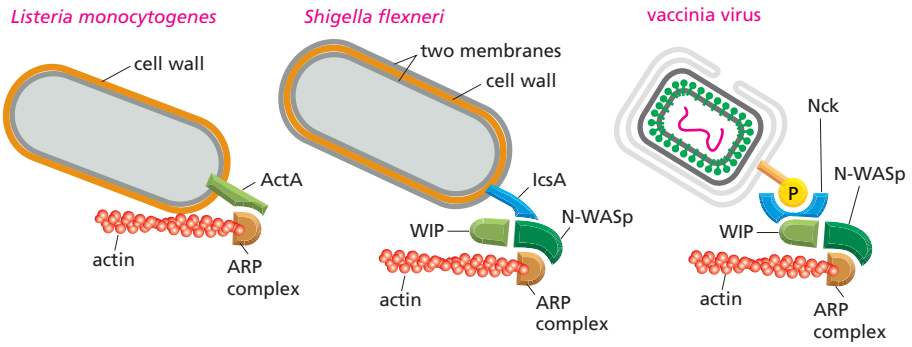


Figure 24–38 Molecular mechanisms for actin nucleation by various pathogens. The bacteria *Listeria monocytogenes* and *Shigella flexneri* and the virus vaccinia all move intracellularly using actin polymerization. To induce actin nucleation, all of these pathogens recruit and activate the ARP complex (see Figure 16–34), although each pathogen uses a different recruitment strategy. *L. monocytogenes* expresses a surface protein, ActA, that directly binds to and activates the ARP complex. *S. flexneri* expresses a surface protein, IcsA (unrelated to ActA), that recruits the host cell signaling protein N-WASp. N-WASp in turn recruits the ARP complex, along with other host proteins, including WIP (WASp-interacting protein). Vaccinia virus expresses an envelope protein that is phosphorylated on tyrosine by a host cell protein tyrosine kinase. The phosphorylated protein then recruits Nck, which binds WIP. WIP binds N-WASp, which recruits and activates the ARP complex. The more complicated mechanism that vaccinia uses is thought to closely resemble how chemotactic factors activate the ARP complex in motile eucaryotic cells. Despite the different molecular strategies used to assemble them, the actin comet tails formed by all three pathogens look very similar, and the pathogens move at similar speeds inside infected cells.

The molecular mechanism of pathogen-induced actin assembly has been determined for several microorganisms. The mechanisms are different for the different pathogens, suggesting that they evolved independently. Although they all use the same host cell regulatory pathway that normally controls the nucleation of actin filaments, they exploit different points in the pathway. As discussed in Chapter 16, activation of the small GTPase Cdc42 by certain extracellular signals leads to the activation of a protein called N-WASp, which in turn activates an ARP complex, which nucleates the growth of a new actin filament. An *L. monocytogenes* surface protein directly binds to and activates the ARP complex to initiate the formation of an actin tail. *B. pseudomallei* and *R. rickettsii* use a similar strategy, although the sequence of the activating proteins differs in all three cases. In contrast, another unrelated surface protein on *S. flexneri* binds to and activates N-WASp, which then activates the ARP complex. Remarkably, vaccinia virus uses yet another mechanism to move intracellularly by inducing actin polymerization, although it exploits the same regulatory pathway (**Figure 24–38**).

Other pathogens rely primarily on microtubule-based transport to move within the host cell. Viruses that infect neurons illustrate this movement. The *neurotropic alpha herpes viruses*, a group that includes the virus that causes chicken pox, provides an important example. The virus enters sensory neurons at the tips of their axons and microtubule-based transport carries the nucleocapsids down the axon to the neuronal nucleus, apparently mediated by attachment of the capsid proteins to the molecular motor protein dynein. After replication and assembly in the nucleus, the enveloped virion is transported along axonal microtubules away from the neuronal cell body, mediated by attachment of a different viral coat component to a kinesin motor protein (**Figure 24–39**). A

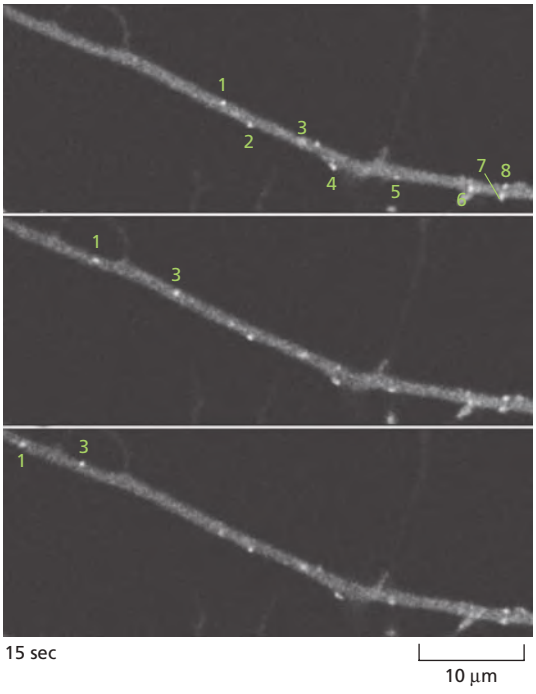


Figure 24–39 Fluorescence micrographs of herpes virus moving in an axon. This nerve cell has been infected with an alpha herpes virus that has been genetically engineered to express green fluorescent protein (GFP) fused to one of its capsid proteins. In this segment of the axon, several viral particles are visible, and two of them (numbered 1 and 3) are moving away from the cell body, which is to the right and out of the picture. (From G.A. Smith, S.P. Gross and L.W. Enquist, *Proc. Natl Acad. Sci. U.S.A.* 98:3466–3470, 2001. With permission National Academy of Sciences.)

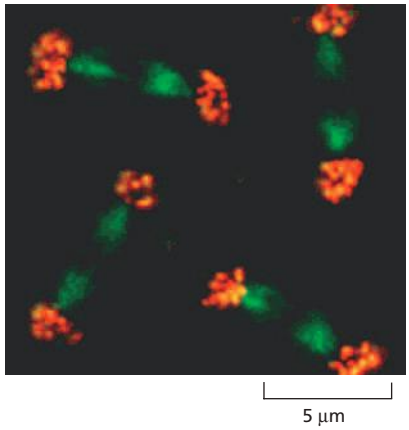


Figure 24–40 The association of *Wolbachia* with microtubules. This fluorescence micrograph shows *Wolbachia (red)* associated with the microtubules *(green)* of four mitotic spindles in a syncytial *Drosophila* embryo. The clumps of bacteria at the spindle poles will segregate along with the chromosomes (not visible at the equator of the spindles) when the chromosomes segregate. (From H. Kose and T.L. Karr, *Mech. Cell Dev.* 51:275–288, 1995. With permission from Elsevier.)

large number of virions, including those of HIV, rabies virus, influenza virus, adenovirus, canine parvovirus, and vaccinia virus (the smallpox relative used for vaccination), have now been shown to associate with dynein or kinesin motor proteins and to undergo directed movement along microtubules at some stage in their replication. A primary function of the microtubule highways laid down in eucaryotic cells is to serve as oriented tracks for membrane traffic; it is not surprising that many viruses have independently evolved the ability to engage these transport systems to enhance their own replication.

One bacterium that is known to associate with microtubules is *Wolbachia*. This fascinating genus includes many species that are parasites or symbionts of insects and other invertebrates, living in the cytosol of each cell in the animal. The infection is spread vertically from mother to offspring, as *Wolbachia* are also present in eggs. The bacteria ensure their transmission into every cell by binding to microtubules, so that the mitotic spindle segregates them simultaneously with chromosome segregation when an infected cell divides (**Figure 24–40**). As we discuss later, *Wolbachia* infection can significantly alter the reproductive behavior of its insect hosts.

Viral Infections Take Over the Metabolism of the Host Cell

Most intracellular bacteria and parasites carry the basic genetic information required for their own metabolism and replication, and they rely on their host cells only for nutrients. Viruses, in contrast, use the basic host cell machinery for most aspects of their reproduction: they all depend on host cell ribosomes to produce their proteins, and some also use host cell DNA and RNA polymerases for replication and transcription, respectively.

Many viruses encode proteins that modify the host transcription or translation apparatus to favor the synthesis of viral proteins over those of the host cell. As a result, the synthetic capability of the host cell is devoted principally to the production of new virus particles. Poliovirus, for example, encodes a protease that specifically cleaves the TATA-binding factor component of TFIID (see Figure 6–18), effectively shutting off all host cell transcription via RNA polymerase II. Influenza virus produces a protein that blocks both the splicing and the polyadenylation of RNA transcripts, which therefore fail to be exported from the nucleus (see Figure 6–40).

Translation initiation of most host cell mRNAs depends on recognition of their 5′ cap by a group of translation initiation factors (see Figure 6–72). Translation initiation of host mRNAs is often inhibited during viral infection, so that the host cell ribosomes can be used more efficiently for the synthesis of viral proteins. Some viral genomes such as that of influenza virus encode endonucleases that cleave the 5′ cap from host cell mRNAs. Some go even further and then use the liberated 5′ caps as primers to synthesize viral mRNAs, a process called *cap snatching*. Several other viral RNA genomes encode proteases that cleave certain translation initiation factors. These viruses rely on 5′ cap-independent translation of the viral RNA, using internal ribosome entry sites (IRESs) (see Figure 7–108).

A few DNA viruses use host cell DNA polymerase to replicate their genome. Unfortunately for the virions, DNA polymerase is expressed at high levels only during S phase of the cell cycle, and most cells that these viral particles infect spend most of their time in G_1 phase. Adenovirus has evolved a mechanism to drive the host cell to enter S phase, producing large amounts of active DNA polymerase that then replicates the viral genome. The adenovirus genome encodes proteins that inactivate both Rb and p53, two key suppressors of cell-

cycle progression (discussed in Chapter 17). As might be expected for any mechanism that induces unregulated DNA replication, these viruses can promote the development of cancer in some circumstances.

Pathogens Can Alter the Behavior of the Host Organism to Facilitate the Spread of the Pathogen

As we have seen, pathogens often alter the behavior of the host cell in ways that benefit the survival and replication of the pathogen. Similarly, pathogens often alter the behavior of the whole host organism to facilitate pathogen spread, as we saw earlier for *Trypanosoma brucei* and *Yersinia pestis*. In some cases, it is difficult to tell whether a particular host response is more for the benefit of the host or for the pathogen. Pathogens such as *Salmonella enterica* that cause diarrhea, for example, usually produce self-limiting infections because the diarrhea efficiently washes out the pathogen. The bacteria-laden diarrhea, however, can spread the infection to a new host. Similarly, coughing and sneezing help to clear pathogens from the respiratory tract, but they also spread the infection to new individuals. A person with a common cold may produce 20,000 droplets in a single sneeze, all carrying rhinovirus or corona virus.

A frightening example of a pathogen modifying host behavior is seen in rabies, as first described in Egyptian writings over 3000 years ago. Rabies virus replicates in neurons and causes infected people or animals to become "rabid": they are unusually aggressive and develop a strong desire to bite. The virus is shed in the saliva and transmitted through the bite wound into the bloodstream of the victim, spreading the infection to a new host.

But *Wolbachia* exhibit the most dramatic example of pathogens modifying host behavior. These bacteria manipulate the sexual behavior of their host to maximize their dissemination. As described earlier, *Wolbachia* are passed vertically into offspring through eggs. If they live in a male, however, they hit a dead end, as they are excluded from sperm. In some species of *Drosophila*, *Wolbachia* modify the sperm of their host so that they can fertilize the eggs only of infected females. This modification creates a reproductive advantage for infected females over uninfected females, so that the overall proportion of *Wolbachia* carriers increases. In other host species, a *Wolbachia* infection kills males but spares females, increasing the number of females in the population and thus the number of individuals that can produce eggs to pass on the infection. In a few types of wasp, *Wolbachia* infections enable the females to produce eggs that develop parthenogenetically, without the need for fertilization by sperm; in this species, males have been completely eliminated. For some of its hosts, *Wolbachia* has become an indispensable symbiont, and curing the infection causes death of the host. In one case, humans are making use of this dependence: the filarial nematode that causes African river blindness is difficult to kill with antiparasite medications, but when people with river blindness are treated with antibiotics that cure the nematode's *Wolbachia* infection, the nematode infection is also arrested.

Pathogens Evolve Rapidly

The complexity and specificity of the molecular interactions between pathogens and their host cells might suggest that virulence would be difficult to acquire by random mutation. Yet, new pathogens are constantly emerging, and old pathogens are constantly changing in ways that make familiar infections difficult to treat. Pathogens have two great advantages that enable them to evolve rapidly. First, they replicate very quickly, providing a great deal of material for the engine of natural selection. Whereas humans and chimpanzees have acquired a 2% difference in genome sequences over about 8 million years of divergent evolution, poliovirus manages a 2% change in its genome in 5 days, about the time it takes the virus to pass from the human mouth to the gut. Second, selective pressures encourage this rapid genetic variation. The host's adaptive immune system and

modern antipathogen drugs, both of which destroy pathogens that fail to change, are the source of these selective pressures.

In many cases, changes in human behavior exacerbate the emergence and evolution of new infectious diseases. Crowded and filthy living conditions in medieval cities, for example, contributed to the spread of the bacterium *Yersinia pestis* to humans from its natural rodent host to cause plague. The tendency of modern humans to live at high population densities in large cities has also created the opportunity for infectious organisms to initiate epidemics, such as influenza, tuberculosis, and AIDS, which could not have spread so rapidly or so far among sparser human populations. Air travel can, in principle, allow an asymptomatic, newly infected host to carry an epidemic to any previously unexposed population within a few hours or days.

Antigenic Variation in Pathogens Occurs by Multiple Mechanisms

A small-scale example of the constant battle between infection and immunity is the phenomenon of **antigenic variation**. An important adaptive immune response against many pathogens is the host's production of antibodies that recognize specific molecules *(antigens)* on the pathogen surface (discussed in Chapter 25). Many pathogens change these antigens during the course of an infection, enabling them to evade elimination by antibodies. Some eucaryotic parasites, for example, undergo programmed rearrangements of the genes encoding their surface antigens. The most striking example occurs in African trypanosomes such as *Trypanosoma brucei*, a protozoan parasite that causes sleeping sickness and is spread by an insect vector. (*T. brucei* is a close relative of *T. cruzi*—see Figure 24–28—but it replicates extracellularly rather than inside cells.) *T. brucei* is covered with a single type of glycoprotein, called *variant-specific glycoprotein* (*VSG*), which elicits a protective antibody response in the host that rapidly clears most of the parasites. The trypanosome genome, however, contains about 1000 *Vsg* genes, each encoding a VSG with distinct antigenic properties. Only one of these genes is expressed at any one time, by being copied into an active expression site in the genome. Gene rearrangements that copy new alleles into the expression site repeatedly change the *Vsg* gene expressed. In this way, a few trypanosomes with an altered VSG escape the antibody-mediated clearance, replicate, and cause the disease to recur, leading to a chronic cyclic infection (**Figure 24–41**). Many other eucaryotic parasites, including the protozoan *Plasmodium falciparum*, which causes malaria, and the fungus *Pneumocystis carinii*, which causes pneumonia in people with AIDS, use very similar strategies to evade the host adaptive immune responses.

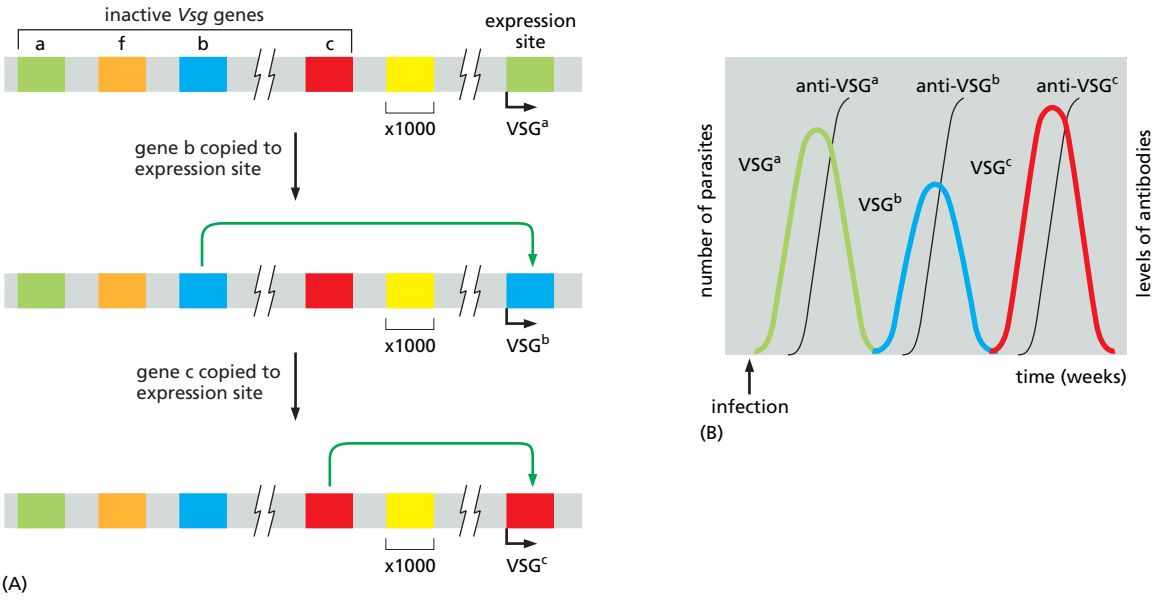


Figure 24–41 Antigenic variation in trypanosomes. (A) There are about 1000 distinct *Vsg* genes in *Trypanosoma brucei*, but there is only one site for *Vsg* gene expression. An inactive gene is copied into the expression site by gene conversion, where it is now expressed. Each *Vsg gene* encodes a different surface protein (antigen). Rare switching events allow the trypanosome to repeatedly change the surface antigen it expresses. (B) A person infected with trypanosomes expressing VSG^a mounts a protective antibody response, which clears most of the parasites expressing this antigen. However, a few of the trypanosomes may have switched to expression of VSG^b, which can now proliferate until anti-VSG^b antibodies clear them. By that time, however, some parasites will have switched to VSG^c, and so the cycle repeats, seemingly indefinitely.

Bacterial pathogens can also rapidly change their surface antigens. Species of the genus *Neisseria* are champions. These Gram-positive cocci can cause meningitis and sexually transmitted diseases. They employ an astonishing variety of mechanisms to promote antigenic variation. First, they undergo genetic recombination very similar to that just described for eucaryotic pathogens, which enables them to vary (over time) the pilin protein they use to make cell-surface pili: by recombination of multiple silent copies of variant *pilin* genes into a single expression locus they can express dozens of slightly different versions of the protein. Second, many cell-surface proteins, as well as many of the biosynthetic enzymes involved in synthesizing cell-surface carbohydrates, have their expression levels continually altered by random slippage and repair of tandem nucleotide repeats in the promoter region or coding sequence of their genes, which modulates transcription or translation. *Neisseria*, for example, have about 10 different genes encoding variants of the Opa family of outer membrane proteins, each of which undergoes random variation of protein expression levels in this way, resulting in a plethora of different surface protein compositions to bewilder the host adaptive immune system. Analysis of the genome sequence for several *Neisseria* species has led to the suggestion that over 100 genes may vary their expression levels using some variation of this mechanism. Third, *Neisseria* are extremely adept at taking up DNA from their environment and incorporating it into their genomes, further contributing to their extraordinary variability. Finally, *Neisseria* lack several of the DNA repair mechanisms present in other bacteria such as *E. coli*, so that the likelihood of their acquiring new mutations through replication error is higher than average. With all these mechanisms working together, it is not surprising that we have not yet developed an effective vaccine against *Neisseria* infections.

Although *Neisseria* is an extreme example, many other bacterial pathogens employ one or more of these techniques to enhance their antigenic variation. Moreover, several studies have shown that pathogenic bacteria isolated from patients with disease symptoms are much more likely to have defects in DNA repair pathways than isolates of the same bacterial species from environmental reservoirs. This intriguing finding suggests that the human immune system may act to accelerate bacterial evolution.

Horizontal gene transfer, rather than point mutations, often causes rapid evolution in bacteria. The acquisition of plasmids and bacteriophages mediates most of this horizontal gene transfer. Bacteria readily pick up pathogenicity islands and virulence plasmids (see Figure 24–5) from other bacteria. Once a bacterium acquires a new set of virulence-related genes, it may quickly establish itself as a new cause of human epidemics. *Yersinia pestis*, for example, is a bacterium endemic to rats and other rodents; it first appeared in human history in 542 A.D., when the city of Constantinople was devastated by plague. Sequence comparisons of *Y. pestis* with those of its close relative *Y. pseudotuberculosis*, which causes a severe diarrheal disease, suggest that *Y. pestis* may have emerged as a distinct strain only a few thousand years ago, not long before its devastating debut as plague.

Error-Prone Replication Dominates Viral Evolution

Error-prone replication mechanisms rather than genomic rearrangements are mainly responsible for antigenic variation in viruses. Retroviral genomes, for example, acquire on average one point mutation every replication cycle, because the viral reverse transcriptase that produces DNA from the viral RNA genome cannot correct nucleotide misincorporation errors. A typical, untreated HIV infection may eventually produce HIV genomes with every possible point mutation. In some ways, the high mutation rate is beneficial for the pathogen. By a microevolutionary process of mutation and selection within each host, most virions change over time from a form that is most efficient at infecting macrophages to one more efficient at infecting T-cells, as described earlier (see Figure 24–23). Similarly, once a patient is treated with an antiviral drug, the viral genome can quickly mutate and be selected for its resistance to the drug. If the

reverse transcriptase error rate were too high, however, deleterious mutations might accumulate too rapidly for the virus to survive. Furthermore, a variant that is successful in one host does not necessarily spread to others, as a mutated virion may not be able to infect a new host. For HIV-1, we can estimate the extent of this constraint by examining the sequence diversity among different infected individuals. Remarkably, only about one-third of the nucleotide positions in the coding sequence of the viral genome are invariant, and nucleotide sequences in some parts of the genome, such as the *env* gene, can differ by as much as 30%. This extraordinary genomic plasticity greatly complicates attempts to develop vaccines against HIV, and it can also lead to rapid drug resistance (discussed below). It has also led to the rapid emergence of new HIV strains. Sequence comparisons between various strains of HIV and the very similar simian immunodeficiency virus (SIV) from a variety of different monkey species suggest that the most virulent type of HIV, HIV-1, may have jumped from chimpanzees to humans as recently as 1930 (**Figure 24–42**).

Influenza viruses are an important exception to the rule that error-prone replication dominates viral evolution. They are unusual in that their genome consists of several (usually eight) strands of RNA. When two strains of influenza infect the same host, the strands of the two strains can recombine to form a novel type of influenza virus. Prior to 1900, the influenza strain that infected humans caused a very mild disease; a different influenza strain infected fowl such as ducks and chickens, but it only rarely infected humans. In 1918, a particularly virulent variant of avian (bird) influenza crossed the species barrier to infect humans, triggering the catastrophic epidemic of 1918 called the Spanish flu, which killed 20–50 million people worldwide, more than were killed in World War I. Subsequent influenza pandemics have been triggered by recombination, in which a new DNA segment from an avian form of the virus replaced one or more of the viral DNA segments governing human immune response to the virus (**Figure 24–43**). Such recombination events allow the new virus to replicate rapidly and spread through an immunologically naive human population. Generally, within two or three years, the human population develops immunity to the new recombinant strain of virus, and, as a result, the infection rate drops to a steady-state level. In normal years, influenza is a mild disease in healthy adults but can be life-threatening in the very young and very old. In pandemic years, however, especially in the 1918 pandemic, healthy adults seem unusually susceptible to lethal influenza infection, perhaps because of the tissue damage caused by an overzealous immune response. Because the recombination events are unpredictable, it is not possible to know when the next influenza pandemic will occur or how severe it might be.

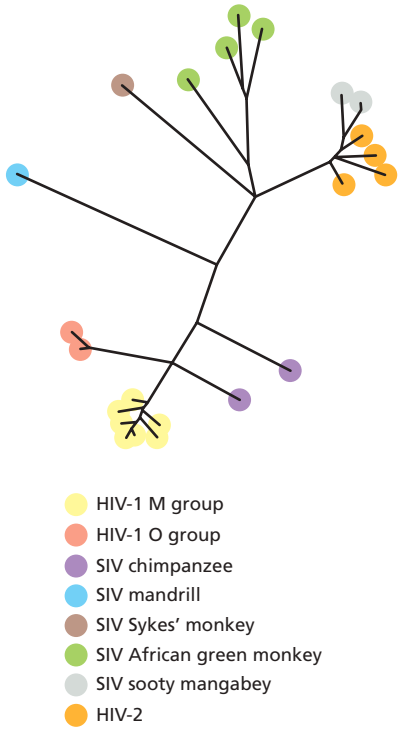


Figure 24–42 Diversification of HIV-1, HIV-2, and related strains of SIV. The genetic distance between any two viral isolates is found by following the shortest path joining them in the tree. HIV-1 is divided into two groups, major (M) and outlier (O). The HIV-1 M group is responsible for the global AIDS epidemic. HIV-1 M is further subdivided into various subtypes (A through G—not named in the figure). Subtype B is dominant in America and Europe; B, C, and E are dominant in Asia; and all subtypes are found in Africa. At least two monkey viruses, chimpanzee and mandrill, are more closely related to HIV-1 than is HIV-2, suggesting that HIV-1 and HIV-2 arose independently of each other. HIV-1 and SIV chimpanzee are estimated to have diverged in about 1930. This tree was constructed from the nucleotide sequences of the *Gag* gene, using a database containing about 16,000 sequences from virus samples isolated around the world.

Drug-Resistant Pathogens Are a Growing Problem

While human activities such as air travel have promoted the spread of certain infectious diseases, advances in public sanitation and in medicine have prevented or ameliorated the suffering caused by many others. Effective vaccines and worldwide vaccination programs have eliminated smallpox and severely reduced poliomyelitis, and many deadly childhood infections such as mumps and measles are now rarities in wealthy industrialized nations. Nonetheless, there are still many widespread and devastating infectious diseases, such as malaria, for which no effective vaccines are available.

The development of drugs that cure rather than prevent infections has also had a major impact on human health. **Antibiotics**, which kill bacteria, comprise the most successful class. Penicillin was one of the first antibiotics used to treat infections in humans, introduced into clinical use just in time to prevent tens of thousands of deaths from infected battlefield wounds in World War II. Because bacteria form a kingdom distinct from the eucaryotes they infect, much of their basic machinery for DNA replication, transcription, translation, and fundamental metabolism differs from that of their host. These differences enable us to find antibacterial drugs that specifically inhibit these processes in bacteria, without disrupting them in the host. Most of the antibiotics that we use to treat bacterial

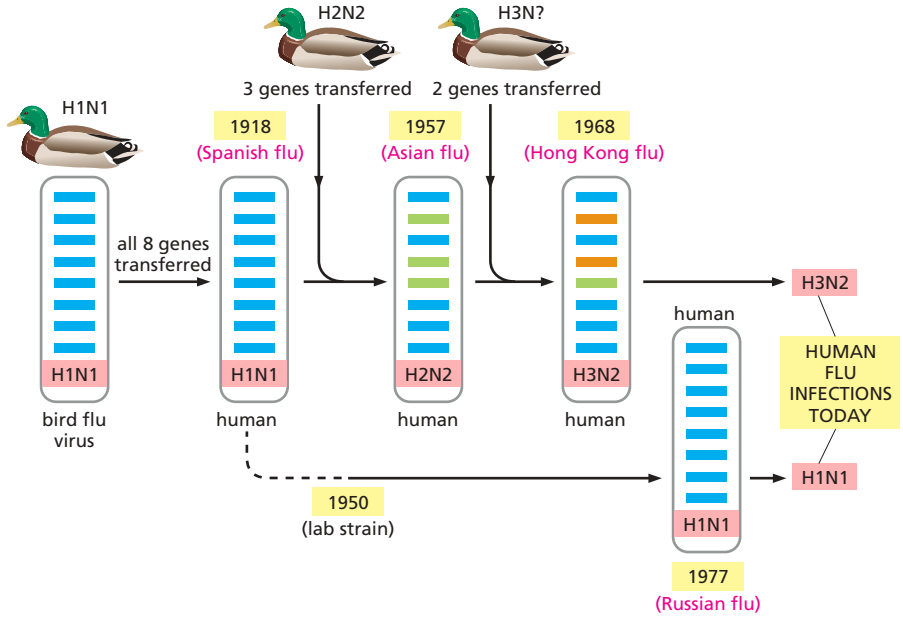


Figure 24–43 Model for the evolution of pandemic strains of influenza virus by recombination. Influenza A virus is a natural pathogen of birds, particularly waterfowl, and is always present in wild bird populations. Before 1900, influenza virus caused sporadic mild disease in humans. In 1918, a particularly virulent form of the virus crossed the species barrier from birds to humans and caused a devastating worldwide epidemic. This strain was designated H1N1, referring to the specific forms of its main antigens, hemagglutinin (HA, or H) and neuraminidase (NA, or N). Changes in the virus, rendering it less virulent, and the rise of adaptive immunity in the human population, prevented the pandemic from continuing in subsequent seasons, although H1N1 influenza strains continued to cause serious disease every year in very young and very old people. In 1957, a new pandemic arose when the genes encoding HA and NA were replaced by equivalent genes from an avian virus (*green bars*); the new strain (designated H2N2) was not effectively cleared by antibodies in people who had previously contracted only H1N1 forms of influenza. In 1968, another pandemic was triggered when the HA gene of H2N2 (but not the NA gene) was replaced from another avian virus; the new virus was designated H3N2. In 1977, there was a resurgence of H1N1 influenza, which had previously been almost completely replaced by the N2 strains. Molecular sequence information suggests that this minor pandemic may have been caused by an accidental release of an influenza strain that had been held in a laboratory since about 1950. As indicated, most human influenza today is caused by H1N1 and H3N2 strains.

Poultry farming brings humans into very close contact with large numbers of birds, which may carry different versions of the influenza virus. These avian viruses occasionally cause fatal disease in humans who have been in direct contact with infected birds (so-called bird flu), but thus far these viruses have not acquired the ability to spread effectively from human to human. Monitoring these cross-species jumps and attempting to predict the origin of the next serious influenza epidemic remain significant public health challenges.

infections are small molecules that inhibit macromolecular synthesis in bacteria by targeting bacterial enzymes that either are distinct from their eucaryotic counterparts or are involved in pathways, such as cell wall biosynthesis, that are absent in humans (**Figure 24–44** and Table 6–3).

The rapid evolution of pathogens, however, enables bacteria to develop resistance to antibiotics very quickly. The typical lag between the introduction of an antibiotic into clinical use and the appearance of the first resistant strains is only a few years. Similar drug resistance also arises rapidly when treating viral infections with antiviral drugs. The virus population in an HIV-infected person treated with the reverse transcriptase inhibitor AZT, for example, will acquire complete resistance to the drug within a few months. The current protocol for treatment of HIV infections involves the simultaneous use of three drugs, which helps to minimize the acquisition of resistance.

There are three general strategies by which a pathogen can develop drug resistance: (1) it can alter the molecular target of the drug so that it is no longer sensitive to the drug; (2) it can produce an enzyme that destroys the drug; or (3) it can prevent access to the target by, for example, actively pumping the drug out of the pathogen (**Figure 24–45**).

Once a pathogen has chanced upon an effective drug-resistance strategy, the newly acquired or mutated genes that confer the resistance are frequently spread throughout the pathogen population and may even transfer to pathogens of different species that are treated with the same drug. The highly

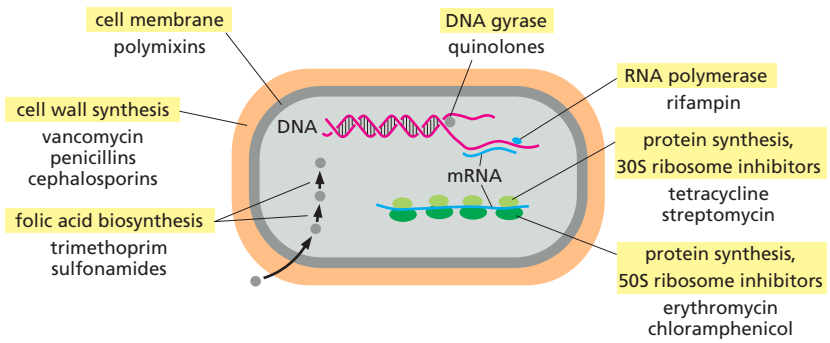


Figure 24–44 Antibiotic targets. Although there are many antibiotics in clinical use, they have a narrow range of targets, which are highlighted in *yellow*. A few representative antibiotics in each class are listed. Nearly all antibiotics used to treat human infections fall into one of these categories. The vast majority inhibit either bacterial protein synthesis or bacterial cell wall synthesis.

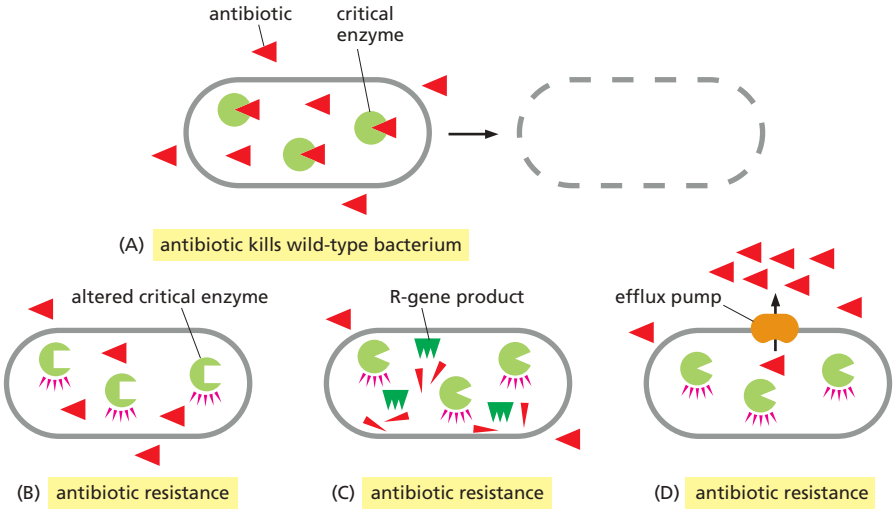


Figure 24–45 Three general mechanisms of antibiotic resistance. (A) Non-resistant bacterial cells bathed in a drug (*red triangles*) that binds to and inhibits a critical enzyme (*light green*) will be killed due to enzyme inhibition. (B) Bacteria that have altered the target enzyme so that it no longer binds to the drug will survive and proliferate. In many cases, a single point mutation in the gene encoding the target protein can generate resistance. (C) Bacteria that express a resistance gene (R-factor) that encodes an enzyme (*dark green*) that either degrades or covalently modifies the drug will survive and proliferate. Some resistant bacteria, for example, make Beta-lactamase enzymes, which cleave penicillin and similar compounds. (D) Yet other resistant bacteria express or up-regulate an efflux pump that ejects the drug from the bacterial cytoplasm using energy derived from either ATP hydrolysis or the electrochemical gradient across the bacterial plasma membrane. Some efflux pumps, such as the TetR efflux pump that confers resistance to tetracycline, are specific for a single drug; others, called multi-drug resistance (MDR) efflux pumps, are capable of exporting a wide variety of structurally dissimilar drugs and so can render a bacterium resistant to a very large number of different antibiotics in a single step.

effective and very expensive antibiotic *vancomycin*, for example, has been used as a treatment of last resort for many severe hospital-acquired bacterial infections that are already resistant to most other known antibiotics. Vancomycin prevents one step in bacterial cell wall biosynthesis, by binding to part of the growing peptidoglycan chain and preventing it from being cross-linked to other chains (see Figure 24–4). Resistance can arise if the bacterium synthesizes a different type of cell wall, using different subunits that do not bind vancomycin. The most effective form of vancomycin resistance depends on a transposon containing seven genes, the products of which work together to sense the presence of vancomycin, shut down the normal pathway for bacterial cell wall synthesis, and generate a different type of cell wall. Although the joining of these genes into a single transposon must have been a difficult evolutionary step (it took 15 years for vancomycin resistance to develop, rather than the typical year or two), the transposon can now be readily transmitted to many other pathogenic bacterial species.

Where do drug-resistance genes come from? Sometimes, when the bacteria are under selective pressure due to drug exposure, resistance genes arise by spontaneous mutation and expand within a population. In many cases, however, they appear in a pathogen's genome as new DNA segments acquired by horizontal transfer, frequently carried on transposons or replicative plasmids. Unlike eucaryotic cells, bacteria commonly exchange genetic material across species boundaries.

Drug-resistance genes acquired by horizontal transfer frequently seem to come from environmental reservoirs, where they play an important part in the competition between microorganisms. Nearly all antibiotics used to treat bacterial infections today are not synthetic creations of chemists; instead, most are natural products produced by fungi or bacteria: penicillin, for example, is made by the mold *Penicillium*, and more than 50% of the antibiotics currently used in the clinic are made by the Gram-positive genus *Streptomyces*. It is believed that these microorganisms produce antimicrobial compounds as weapons in their competition with other microorganisms in the environment. Many of these compounds have probably existed on Earth for at least hundreds of millions of years, which is ample time for other microorganisms, as well as those that produce the antibiotics themselves, to have evolved resistance mechanisms. Unbiased surveys of bacteria taken from soil samples that have never been deliberately exposed to antibiotic drugs reveal that the bacteria are typically already resistant to about seven or eight of the antibiotics widely used in clinical practice. When pathogenic microorganisms are faced with the selective pressure provided by antibiotics treatments, they can apparently draw upon this world-wide and essentially inexhaustible source of genetic material to acquire resistance.

Like most other aspects of infectious disease, human behavior has exacerbated the problem of drug resistance. Many patients choose to take antibiotics

for viral conditions such as influenza, colds, sore throats, and earaches that the drugs do not help. Persistent and chronic misuse of antibiotics in this way can eventually result in antibiotic-resistant normal flora, which can then transfer the resistance to pathogens. Several antibiotic-resistant outbreaks of infectious diarrhea caused by *Shigella flexneri*, for example, originated in this way. The problem is particularly severe in countries where antibiotics are available without a physician's prescription, as in Brazil, where more than 80% of the strains of *S. flexneri* found in infected patients are resistant to four or more antibiotics. Antibiotics are also misused in agriculture, where they are commonly employed as food additives to promote the health of farm animals. An antibiotic closely related to vancomycin was commonly added to cattle feed in Europe; the resulting resistance in the normal flora of these animals is widely believed to be one of the original sources for vancomycin-resistant bacteria that now threaten the lives of hospitalized patients.

Because the acquisition of drug resistance is almost inevitable, it is crucial that we continue to develop innovative novel treatments for infectious diseases. We must also try to do better at delaying the onset of drug resistance.

Summary

All pathogens share the ability to interact with host cells in various ways that promote replication and spread of the pathogen, but these host–pathogen interactions are diverse. Pathogens often colonize the host by adhering to or invading the epithelial surfaces that line the respiratory, gastrointestinal, and urinary tracts, as well as the other surfaces in direct contact with the environment. Intracellular pathogens, including all viruses and many bacteria and protozoa, invade host cells and replicate inside them. They invade by one of several mechanisms. Viruses rely largely on receptor-mediated endocytosis, whereas bacteria exploit cell adhesion and phagocytic pathways; in both cases, the host cell provides the machinery and energy. Protozoa, by contrast, employ unique invasion strategies that usually require significant metabolic expense on the part of the invader. Once inside, intracellular pathogens seek out a cell compartment that is favorable for their replication, frequently altering host membrane traffic and exploiting the host cell cytoskeleton for intracellular movement. Besides altering the behavior of individual host cells, pathogens frequently alter the behavior of the host organism in ways that favor spread to a new host. Pathogens evolve rapidly, so that new infectious diseases frequently emerge, and old pathogens acquire new ways to evade our attempts at treatment, prevention, and eradication.

BARRIERS TO INFECTION AND THE INNATE IMMUNE SYSTEM

Humans are exposed to millions of potential pathogens daily, through contact, ingestion, and inhalation. Our ability to avoid infection depends in large part on our adaptive immune system (discussed in Chapter 25), which remembers previous encounters with specific pathogens and specifically destroys or eliminates them when they attack again. Adaptive immune responses, however, are slow to develop on first exposure to a new pathogen, as specific clones of B and T cells that can respond to it have to become activated and proliferate; it can therefore take a week or so before the responses are effective. By contrast, a single bacterium with a doubling time of 1 hour can produce almost 20 million progeny, a full-blown infection, in a single day. Therefore, during the first critical hours and days of exposure to a new pathogen, we rely on our **innate immune system** to protect us from infection. As we discuss in Chapter 25, we also rely on the innate immune system to help activate adaptive immune responses.

Innate immune responses are not specific to a particular pathogen in the way that the adaptive immune responses are. Generally, there are three lines of innate immune defenses that can prevent an infection or stop it in its tracks before the adaptive immune system needs be called into play. The first of these

are the physical and chemical barriers that prevent easy access of microorganisms to the interior of the human body. These include the thick layer of dead keratinized cells that forms the surface of our skin, the tight junctions between epithelial cells, the acidic pH of the stomach, and components of the mucus layers that inhibit colonization or even kill pathogenic bacteria. The normal flora also have a role in protecting body surfaces against invaders by competing for the same ecological niche and thereby limiting colonization.

The second line of innate defenses comprise cell-intrinsic responses, by which an individual cell recognizes that it has been infected and takes measures to kill or cripple the invader. Most cells that have taken up a bacterium by pathogen-induced phagocytosis (see Figure 24–26), for example, will immediately direct the phagosome to fuse with a lysosome, exposing the invading microorganism to a barrage of digestive enzymes. Another ancient intrinsic defense mechanism is the ability of host cells to degrade double-stranded RNA, which is a common intermediate in viral replication; the infected cells will even degrade any single-stranded RNA that shares sequence identity with the double-stranded trigger. This mechanism not only serves as an effective intrinsic defense against many viral infections, it also enables cell biologists to manipulate gene expression by using the technique of RNA interference (RNAi).

The third line of innate immune defenses depends on a specialized set of proteins and phagocytic cells that recognize conserved features of pathogens and become quickly activated to help destroy invaders. These include professional phagocytic cells such as neutrophils and macrophages, natural killer cells, and the complement system. Whereas the adaptive immune system arose in evolution less than 500 million years ago and is confined to vertebrates, innate immune responses operate in both vertebrates and invertebrates, as well as in plants, and the basic mechanisms that regulate them are similar in these organisms. As discussed in Chapter 25, the innate immune responses in vertebrates are also required to activate adaptive immune responses by producing extracellular signal molecules that help call the adaptive immune system into action.

Epithelial Surfaces and Defensins Help Prevent Infection

In vertebrates, the skin and other epithelial surfaces, including those lining the respiratory, intestinal, and urinary tracts (**Figure 24–46**), provide a physical barrier between the inside of the body and the outside world. A mucus layer provides additional protection against microbial, mechanical, and chemical insults of the interior epithelial surfaces; many amphibians and fish also have a mucus layer covering their skin. The slimy mucus coating is made primarily of secreted mucin and other glycoproteins, and it physically helps prevent pathogens from adhering to the epithelium. It also facilitates the clearance of pathogens by beating cilia on the epithelial cells (discussed in Chapter 23).

The mucus layer also contains substances that either kill pathogens or inhibit their proliferation. Among the most abundant of these are antimicrobial peptides, called **defensins**, which are found in all animals and plants. They are generally short (12–50 amino acids) and positively charged, and have hydrophobic or amphipathic domains. They constitute a diverse family with a broad spectrum of antimicrobial activities, including the ability to kill or inactivate Gram-negative and Gram-positive bacteria, fungi (including yeasts), parasites (including protozoa and nematodes), and even enveloped viruses such as HIV. Defensins are also the most abundant proteins in neutrophils (see below), which use them to kill phagocytosed pathogens.

It is still uncertain how defensins kill pathogens. One possibility is that they use their hydrophobic or amphipathic domains to insert into the surface membrane of their victims, thereby disrupting the integrity of the membrane. Some of their selectivity for pathogens over host cells may come from their preference for membranes that do not contain cholesterol. After disrupting the membrane of the pathogen, the positively charged peptides may also interact with various negatively charged targets within the microbe, including DNA. Because of the relatively nonspecific nature of the interaction between antimicrobial peptides

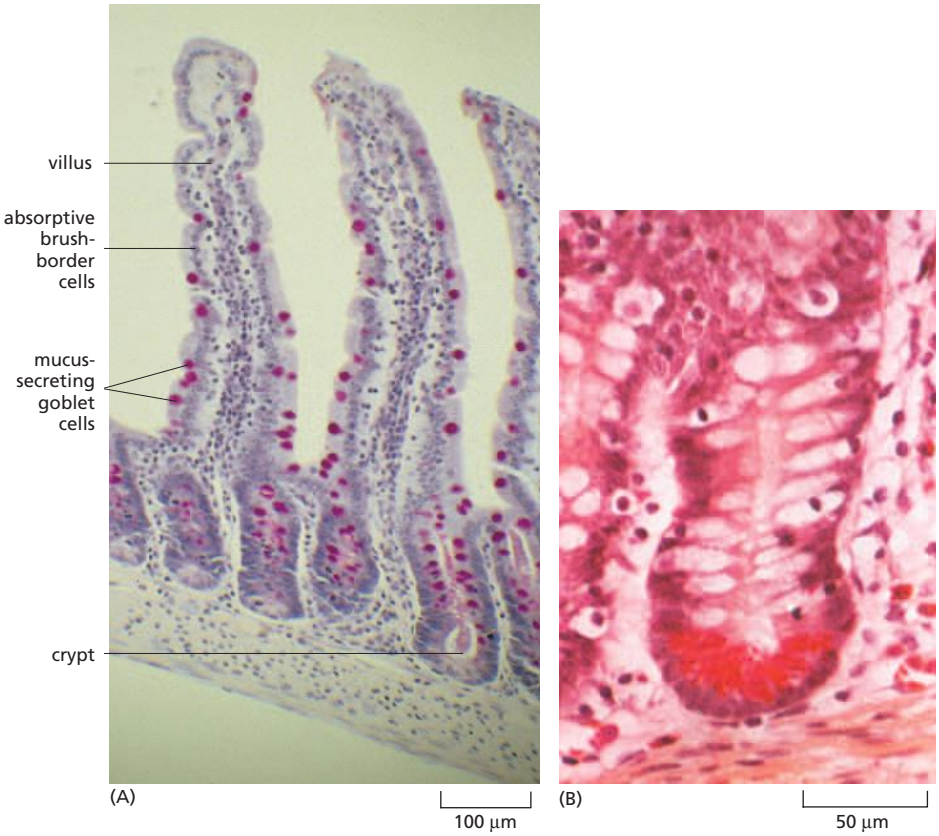


Figure 24–46 Epithelial defenses against microbial invasion. (A) Light micrograph of a cross section through the wall of the human small intestine, showing three villi. Goblet cells secreting mucus are stained *magenta*. The protective mucus layer covers the exposed surfaces of the villi. At the base of the villi lie the *crypts,* where the epithelial cells proliferate. (B) Close-up view of a crypt, stained to show the granules in the Paneth cells *(scarlet*). These cells secrete large quantities of defensins and other antimicrobial peptides into the intestinal lumen. (B, courtesy of H.G. Burkitt, from P.R. Wheater, Functional Histology, 2nd ed. London: Churchill–Livingstone, 1987.)

and the microbes they kill, it is difficult for pathogens to acquire resistance to them. Thus, in principle, defensins and other antimicrobial peptides might be useful therapeutic agents to combat infection, either alone or in combination with more traditional drugs.

Human Cells Recognize Conserved Features of Pathogens

Microorganisms do occasionally breach the epithelial barricades. It is then up to the innate and adaptive immune systems to recognize and destroy them, without harming the host. Consequently, the immune systems must be able to distinguish self from nonself. We discuss in Chapter 25 how the adaptive immune system does this. The innate immune system relies on the recognition of particular types of molecules that are common to many pathogens but are absent in the host. These pathogen-associated molecules (called *pathogen-associated* or *microbe-associated immunostimulants*) trigger two types of innate immune responses—*inflammatory responses* (discussed below) and phagocytosis by professional phagocytes (neutrophils and macrophages), and by *dendritic cells*, which activate T cells of the adaptive immune system (discussed in Chapter 25). Both the inflammatory and phagocytic responses can occur quickly, even if the host has never been previously exposed to a particular pathogen.

The **microbe-associated immunostimulants** are of various types. Most are not exclusive to pathogens, but are found in many bacteria, benign as well as harmful. Bacterial translation initiation differs from eucaryotic translation initiation in that *formylated methionine*, rather than regular methionine, is generally used as the first amino acid. Therefore, any peptide containing formylmethionine at the N-terminus must be of bacterial origin. Formylmethionine-containing peptides act as very potent chemoattractants for neutrophils, which migrate quickly to the source of such peptides and engulf the bacteria producing them (see Figure 16–101).

In addition, molecules that do not occur in multicellular hosts compose the outer surface of many microorganisms, and these molecules also act as

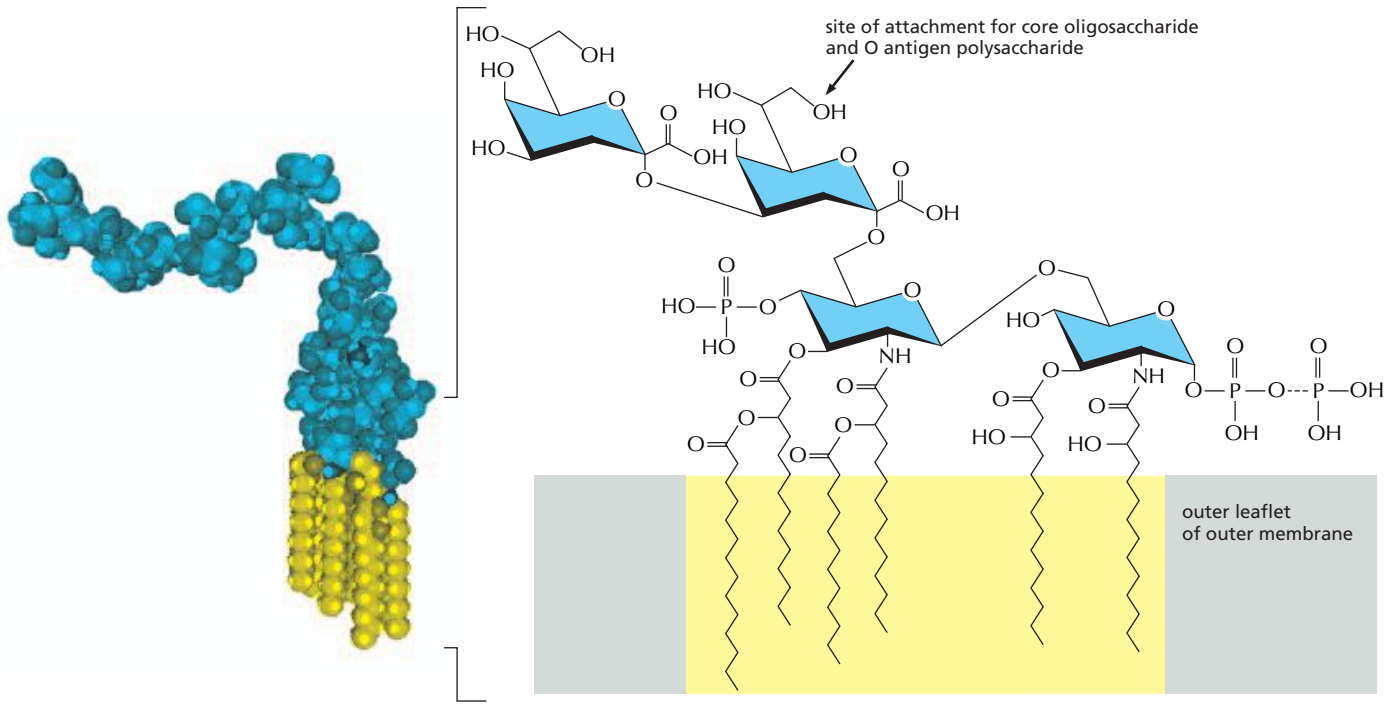


Figure 24–47 Structure of lipopolysaccharide (LPS). On the *left* is the three-dimensional structure of a molecule of LPS, with the fatty acid tails shown in *yellow* and the sugars in *blue*. The molecular structure of the bottom part of LPS is shown on the *right*. The membrane anchor is made up of two linked glucosamine sugars attached to three phosphates and six fatty acid tails. Attached to this basic structure is a long chain of sugars, which is usually highly branched. This drawing shows the simplest type of LPS that will allow *E. coli* to live; it has just two sugar molecules in the chain, and both are the same (3-deoxy-D-manno-octulosonic acid). At the position marked by the *arrow*, wild-type Gram-negative bacteria also attach a *core oligosaccharide* made up of 8–12 linked sugars and a long *O antigen* polysaccharide made up of an oligosaccharide unit that is repeated up to 40 times. The sugars making up the core oligosaccharide and O antigen polysaccharide vary from one bacterial species to another, and even among different strains of the same species. All forms of LPS stimulate strong innate immune responses.

immunostimulants. They include the peptidoglycan cell wall and flagella of bacteria, as well as lipopolysaccharide (LPS) on Gram-negative bacteria (**Figure 24–47**) and teichoic acids on Gram-positive bacteria (see Figure 24–4B). They also include molecules in the cell walls of fungi, including mannan, glucan, and chitin. Many eucaryotic parasites also contain unique membrane components that act as immunostimulants, including glycosylphosphatidylinositol in *Plasmodium*. In order to avoid making unnecessary immune responses, the host must be able to distinguish between microbe-associated immunostimulants produced by pathogens and very similar or identical molecules released from normal flora. In many cases, differences in concentration of the immunostimulant may be sufficient; chronic low concentrations of the molecule may be monitored by the immune system but do not provoke a reaction, while sudden increases in concentration or the appearance of immunostimulants in normally sterile areas of the body will trigger an innate immune response.

Short sequences in bacterial or viral DNA can also act as immunostimulants. The culprit is a "CpG motif," which consists of the unmethylated dinucleotide CpG flanked by two 5′ purine residues and two 3′ pyrimidines. This short sequence is at least 20 times less common in vertebrate DNA than in bacterial or viral DNA, and it can activate innate immune responses.

The various classes of microbe-associated immunostimulants often occur in repeating patterns and are therefore often called *pathogen-associated molecular patterns (PAMPs)*. Several types of dedicated receptors in the host, collectively called **pattern recognition receptors**, recognize these patterns. These receptors include soluble receptors in the blood (components of the *complement system*, which we discuss below) and membrane-bound receptors on or in host cells (including members of the *Toll-like receptor family*, which we consider

later). The cell-associated receptors have two functions: they initiate the phagocytosis of the pathogen, and they activate a program of gene expression in the host cell responsible for innate immune responses. Some of the complement components also aid in phagocytosis and, in some cases, the direct killing of the pathogen, as we now discuss.

Complement Activation Targets Pathogens for Phagocytosis or Lysis

The **complement system** consists of about 20 interacting soluble proteins that are made mainly by the liver and circulate in the blood and extracellular fluid. Most are inactive until an infection activates them. They were originally identified by their ability to amplify and "complement" the action of antibodies, but some complement components are also pattern recognition receptors that microbe-associated immunostimulants activate directly.

The *early complement components* are activated first. There are three sets of these, belonging to three distinct pathways of complement activation—the *classical pathway*, the *lectin pathway*, and the *alternative pathway*. The early components of all three pathways act locally to activate C3, which is the pivotal component of complement (**Figure 24–48**). Individuals with a C3 deficiency are subject to repeated bacterial infections. The early components are proenzymes, which are activated sequentially by proteolytic cleavage. The cleavage of each proenzyme in the series activates the next component to generate a serine protease, which cleaves the next proenzyme in the series, and so on. Since each activated enzyme cleaves many molecules of the next proenzyme in the chain, the activation of the early components consists of an amplifying, *proteolytic cascade*.

Many of these cleavages liberate a biologically active small peptide fragment that can attract phagocytic cells such as neutrophils, and a membrane-binding larger fragment. The binding of the large fragment to a cell membrane, usually the surface of a pathogen, helps to carry out the next reaction in the sequence. In this way, complement activation is confined largely to the particular cell surface where it began. The larger fragment of C3, called C3b, binds covalently to the surface of the pathogen, where it recruits fragments of cleaved C2 and C3 to form a proteolytic complex (C4b, C2b, C3b) that catalyzes the subsequent steps in the complement cascade. Specific receptors on phagocytic cells that enhance the ability of these cells to phagocytose the pathogen also recognize C3b. In addition, receptors on B cells recognize C3b, which is the reason that C3b-coated pathogens are especially efficient at stimulating B cells to make

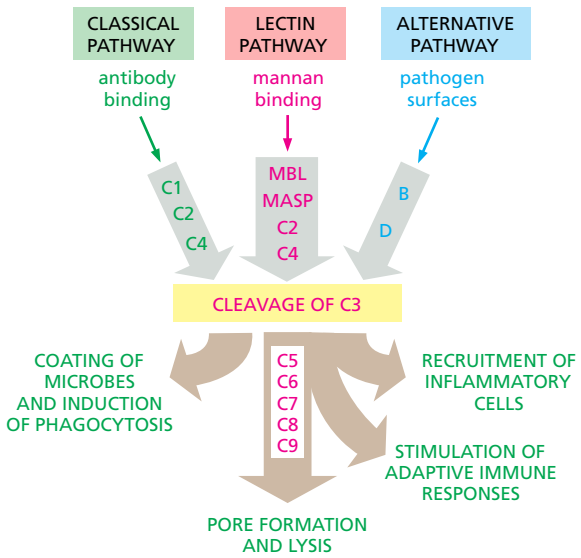


Figure 24–48 The principal stages in complement activation by the classical, lectin, and alternative pathways. In all three pathways, the reactions of complement activation usually take place on the surface of an invading microbe, such as a bacterium, and lead to the cleavage of C3. C1–C9, mannose-binding lectin (MBL), MBL-associated serine protease (MASP), and factors B and D are the central components of the complement system; various other components regulate the system. The early components are shown within *gray arrows*, while the late components are shown within a *brown arrow*.

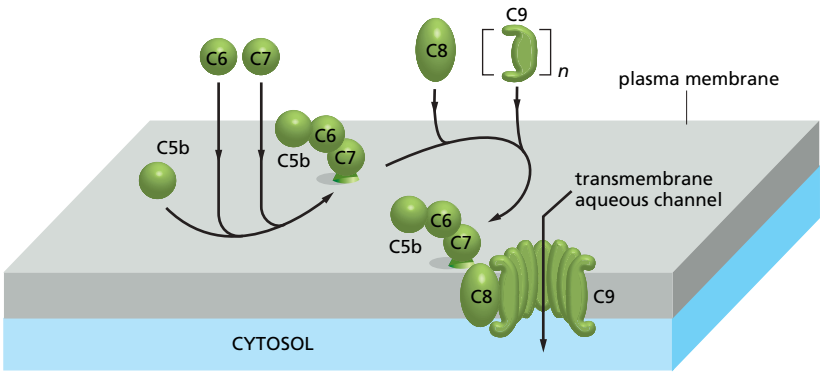


Figure 24–49 Assembly of the late complement components to form a membrane attack complex. When C3b is produced by any of the three complement activation pathways, it is immobilized on a membrane, where it recruits C4b and C2b to form a proteolytic complex. This complex then cleaves the first of the late components, C5, to produce C5a (not shown) and C5b. C5b remains loosely bound to C3b (not shown) and rapidly assembles with C6 and C7 to form C567, which then binds firmly via C7 to the membrane, as illustrated. To this complex is added one molecule of C8 to form C5678. The binding of a molecule of C9 to C5678 induces a conformational change in C9 that exposes a hydrophobic region and causes C9 to insert into the lipid bilayer of the target cell. This starts a chain reaction in which the altered C9 binds a second molecule of C9, which can then bind another molecule of C9, and so on. In this way, a chain of C9 molecules forms a large transmembrane channel in the membrane.

antibodies (discussed in Chapter 25). The smaller fragment of C3 (called C3a), as well as fragments of C4 and C5 (see Figure 24–48), act independently as diffusible signals to promote an inflammatory response by recruiting phagocytes and lymphocytes to the site of infection.

IgG or IgM antibody molecules (discussed in Chapter 25) bound to the surface of a microbe activate the **classical pathway**. **Mannan-binding lectin**, the protein that initiates the second pathway of complement activation, is a serum protein that forms clusters of six carbohydrate-binding heads around a central collagen-like stalk. This assembly binds specifically to mannose and fucose residues in bacterial cell walls that have the correct spacing and orientation to match up perfectly with the six carbohydrate-binding sites, providing a good example of a pattern recognition receptor. These initial binding events in the classical and lectin pathways cause the recruitment and activation of the early complement components. Molecules on the surface of pathogens often activate the **alternative pathway**; activation of the classical or lectin pathways also activates the alternative pathway, forming a positive feedback loop that amplifies the effects of the classical or lectin pathway.

Host cells produce various proteins and surface modifications that prevent the complement reaction from proceeding on their cell surface. The most important of these is the carbohydrate moiety sialic acid, a common constituent of cell surface glycoproteins and glycolipids. Because pathogens generally lack these surface components, they are singled out for destruction, while host cells are spared. At least one pathogen, *Neisseria gonorrhoeae*, the bacterium that causes the sexually transmitted disease gonorrhea, has developed the ability to exploit this feature of host self-protection. Coating itself with a layer of sialic acid, it effectively hides from the complement activation system.

Membrane-immobilized C3b, produced by any of the three pathways, triggers a further cascade of reactions that leads to the assembly of the late complement components to form *membrane attack complexes* (**Figure 24–49**). These complexes assemble in the pathogen membrane near the site of C3 activation and have a characteristic appearance in negatively stained electron micrographs, where they are seen to form aqueous pores through the membrane (**Figure 24–50**). For this reason, and because they perturb the structure of the bilayer in their vicinity, they make the membrane leaky and can, in some cases, cause the microbial cell to lyse, much like the defensins mentioned earlier.

The self-amplifying, inflammatory, and destructive properties of the complement cascade make it essential that key activated components be rapidly inactivated after they are generated to ensure that the attack does not spread to nearby host cells. Inactivation is achieved in at least two ways. First, specific inhibitor proteins in the blood or on the surface of host cells terminate the cascade, by either binding or cleaving certain components once they have been activated by proteolytic cleavage. Second, many of the activated components in the cascade are unstable; unless they bind immediately to either another appropriate complement component in the cascade or to a nearby membrane, they rapidly become inactive.

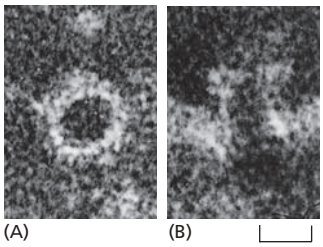


Figure 24–50 Electron micrographs of negatively stained complement lesions in the plasma membrane of a red blood cell. The lesion in (A) is seen *en face*, while that in (B) is seen from the side as an apparent transmembrane channel. The negative stain fills the channels, which therefore look black. This red blood cell has been deliberately sensitized to be susceptible to lysis by complement. (From R. Dourmashkin, *Immunology* 35:205–212, 1978. With permission from Blackwell Publishing.)

Toll-like Proteins and NOD Proteins Are an Ancient Family of Pattern Recognition Receptors

Many of the mammalian pattern recognition receptors responsible for triggering innate immune responses to pathogens are members of the **Toll-like receptor** (**TLR**) family. *Drosophila* Toll is a transmembrane protein with a large extracellular domain consisting of a series of leucine-rich repeats (see Figure 15–82). It was originally identified as a protein involved in the establishment of dorsoventral polarity in developing fly embryos (discussed in Chapter 22). It is also involved, however, in the adult fly's resistance to fungal infections. The intracellular signal transduction pathway activated downstream of Toll when a fly is exposed to a pathogenic fungus leads to the translocation of the NFκB protein (discussed in Chapter 15) into the nucleus, where it activates the transcription of various genes, including those encoding antifungal defensins. The leucine-rich repeats found in Toll and TLRs are versatile structural motifs that are useful for binding a wide variety of ligands. Besides their role in pathogen recognition in both animals and plants, proteins with leucine-rich repeats have roles in signal transduction, DNA repair, and cell–cell and cell–matrix adhesion.

Humans have at least 10 TLRs, several of which play important parts in recognizing microbe-associated immunostimulants made by bacteria, viruses, fungi, and parasites. Different ligands activate different TLRs: TLR4, for example, recognizes lipopolysaccharide (LPS) from the outer membrane of Gram-negative bacteria, TLR9 recognizes CpG DNA, and TLR5 recognizes the protein that makes up the bacterial flagellum. Most TLRs are on cell surfaces; they are abundant, for instance, on the surface of macrophages, dendritic cells, and neutrophils, as well as on the surface of epithelial cells lining the respiratory and intestinal tracts. Others, however, are associated with intracellular membranes, where they can detect intracellular pathogens. TLRs act as an alarm system to alert both the innate and adaptive immune systems that an infection is brewing. In mammals, they activate a variety of intracellular signaling pathways, which in turn stimulate the transcription of hundreds of genes, especially those that promote inflammatory responses (discussed later) and help induce adaptive immune responses (**Figure 25–51**).

A second family of pattern recognition receptors is exclusively intracellular. They are called **NOD proteins** and also have leucine-rich repeat motifs. They are also functionally similar to TLRs but recognize a distinct set of ligands, including

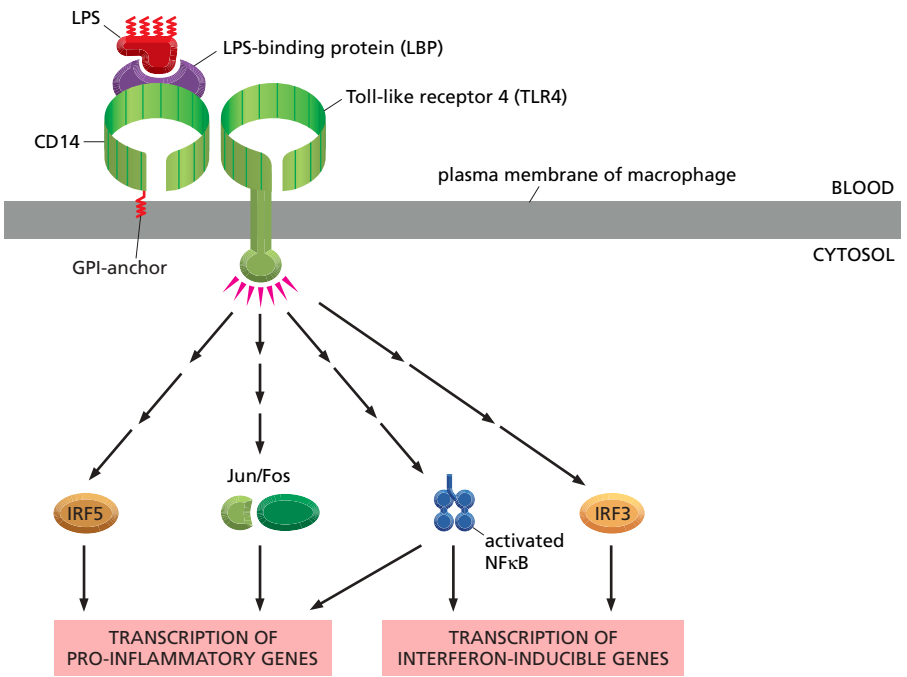


Figure 24–51 The activation of a macrophage by lipopolysaccharide (LPS). LPS binds to LPS-binding protein (LBP) in the blood, and the complex binds to the GPI-anchored protein CD14 on the macrophage surface. The ternary complex then activates Toll-like receptor 4 (TLR4), which activates multiple, downstream, intracellular signaling pathways. As a result, at least four gene regulatory proteins are activated, including NFκB, an AP1 complex of Jun and Fos, and two interferon regulatory factors, IRF3 and IRF5.This strong, multifaceted transcriptional response results in the production of interferons and pro-inflammatory cytokines, including chemokines that recruit various white blood cells to the site of macrophage activation, reflecting the significant danger that a macrophage perceives when it encounters a high concentration of LPS.



Figure 24–52 Microbial disease in a plant. The tomato leaves shown here are infected with the leaf mold fungus *Cladosporium fulvum*. Resistance to this type of infection depends on recognition of a fungal protein by host cell receptors that are structurally related to TLRs. (Courtesy of Jonathan Jones.)

bacterial cell wall components. The different allelic forms of NODs and TLRs that an individual expresses play an important part in influencing their susceptibility to certain infectious diseases; particular polymorphisms in TLR4 and TLR5, for example, correlate with susceptibility to *Legionella pneumophila*, and members of families that express a particular allele of NOD2 have a greatly increased chance of suffering from Crohn's disease, a chronic inflammatory disease of the small intestine, which is thought to be triggered by bacterial infection.

Proteins related to Toll, TLRs, and NODs are apparently involved in innate immunity in all multicellular organisms. In plants, proteins with leucine-rich repeats and with domains homologous to the cytosolic portion of TLRs are required for resistance to fungal, bacterial, and viral pathogens (**Figure 24–52**). Thus, at least two families of proteins that function in innate immunity—the defensins and the TLR/NOD families—seem to be evolutionarily very ancient, perhaps predating the split between animals and plants over a billion years ago. Their conservation during evolution underlines the importance of innate immune responses in the defense against microbial pathogens.

Phagocytic Cells Seek, Engulf, and Destroy Pathogens

In all animals, invertebrate as well as vertebrate, the recognition of a microbial invader is usually quickly followed by its engulfment by a phagocytic cell. Plants, however, lack this type of innate immune response. In vertebrates, *macrophages* are professional phagocytes that reside in tissues throughout the body and are especially abundant in areas where infections are likely to arise, including the respiratory and intestinal tracts, for example. They are also present in large numbers in connective tissues, in the liver and spleen. These long-lived cells patrol the tissues and are among the first cells to encounter invading microbes. *Neutrophils* are the second major type of professional phagocytic cells in vertebrates. In contrast to macrophages, they are short-lived cells, abundant in blood but not present in normal, healthy tissues. They are rapidly recruited to sites of infection by activated macrophages, by molecules such as formylmethionine-containing peptides that the microbes themselves release, and by peptide fragments of cleaved complement components. Neutrophils can detect complement-derived chemoattractants at concentrations as low as 10^{-11} M.

Macrophages and neutrophils display a variety of cell-surface receptors that enable them to recognize and engulf pathogens. <TGTA> These include pattern recognition receptors such as TLRs, receptors for antibodies produced by the adaptive immune system, and receptors for the C3b component of complement. Binding to any of these receptors induces actin polymerization at the site of pathogen attachment, causing the phagocyte's plasma membrane to surround the pathogen and engulf it in a large membrane-enclosed phagosome (**Figure 24–53**). Although some bacteria can actually actively induce a host cell such as an epithelial cell to phagocytose them as a mechanism for invading the cell (see

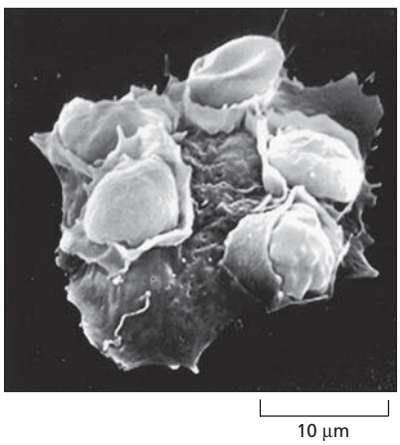


Figure 24–53 Phagocytosis. This scanning electron micrograph shows a macrophage in the midst of ingesting five red blood cells that have been coated with an antibody against a surface glycoprotein. (From E.S. Gold et al., *J. Exp. Med.* 190:1849–1856, 1999. With permission from The Rockefeller University Press.)

Figure 24–25), phagocytosis by a macrophage or neutrophil generally leads to the ingested pathogen's death. Unsurprisingly, some pathogens use specific mechanisms to avoid phagocytosis by macrophages or neutrophils. One strategy is to secrete a thick, slimy layer of polysaccharides, called a *capsule*, which blocks access of complement components to the bacterial surface and also makes it physically difficult for the phagocytic cell to bind to and engulf the bacterium. Another strategy, used by *Yersinia pestis* (the causative bacterium of plague), for example, is to deliver a toxin into the macrophage via a type III secretory system (see Figure 24–8) that disrupts the assembly of the actin cytoskeleton and thereby prevents phagocytosis.

When a macrophage or neutrophil engulfs a pathogen, it unleashes an impressive armory of weapons to kill it. Exposure to both microbe-associated immunostimulants and chemical signals produced by the immune response to the pathogen enhances the phagocytic and killing power of the phagocytes. This exposure is said to "activate" the phagocyte, putting it in a state of high alert, in which not only is it more effective at phagocytosing and killing pathogens, but it also releases cytokines to attract more white blood cells to the site of infection. The location of the phagocyte's weaponry is readily visible in the light or electron microscope as dense membrane-enclosed organelles called *granules*. These specialized lysosomal derivatives fuse with the phagosomes, delivering enzymes such as lysozyme and acid hydrolases that can degrade the pathogen's cell wall and proteins. The granules also contain defensins, the antimicrobial peptides that make up about 15% of the total protein in neutrophils. In addition, the phagocyte assembles *NADPH oxidase complexes* on the phagolysosomal membrane, which catalyze the production of highly toxic oxygen-derived compounds, including superoxide (O_2^-), hypochlorite (HOCl, the active ingredient in bleach), hydrogen peroxide, and hydroxyl radicals. A transient increase in oxygen consumption by the phagocytic cells, called the *respiratory burst*, accompanies the production of these toxic compounds. It is not only these highly reactive oxygen-derived compounds that damage the pathogen trapped within the phagolysosome. The action of NADPH oxidase transports electrons into the phagolysosome and induces a compensatory movement of K^+ along with the electron, which has the net effect of raising the pH. The high pH in the phagolysosome activates a group of potent neutral proteases, which the low pH of the lysosomal granule kept inactive prior to fusion with the phagosome. The neutral proteases quickly destroy the hapless pathogens trapped in the phagolysosome. Whereas macrophages generally survive this killing frenzy and live to kill again, neutrophils usually do not. Dead and dying neutrophils are a major component of the pus that forms in acutely infected wounds. The distinctive greenish tint of pus is due to the abundance in neutrophils of the copper-containing enzyme myeloperoxidase, which is one of the components active in the respiratory burst.

If a pathogen is too large to be successfully phagocytosed (if it is a large parasite such as a nematode, for example), a group of macrophages, neutrophils, or eosinophils (discussed in Chapter 23) will gather around the invader. They secrete their defensins and other bactericidal products contained in their granules by exocytosis, and they also release the toxic products of the respiratory burst (**Figure 24–54**). This barrage is generally sufficient to destroy the pathogen. In some cases, neutrophils have been observed to eject large parts of their chromatin along with the contents of their granules. The ejected DNA, with its attached histones, form a sticky web that entraps nearby bacteria, preventing their escape (**Figure 24–55**). Because its sole function is to sacrifice itself to kill invading pathogens, a neutrophil has no hesitation in using every tool available, including its own DNA, to accomplish this task.

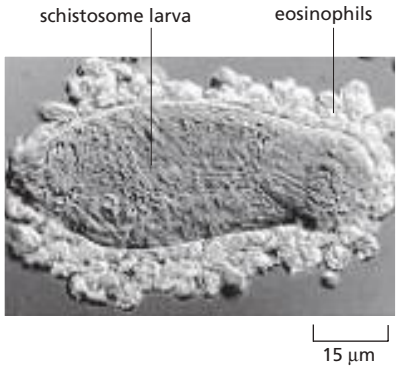


Figure 24–54 Eosinophils attacking a schistosome larva. Phagocytes cannot ingest large parasites such as the schistosome larva shown here. When the larva is coated with antibody or complement, however, eosinophils (and other white blood cells) can recognize it and collectively kill it by secreting various toxic molecules. (Courtesy of Anthony Butterworth.)

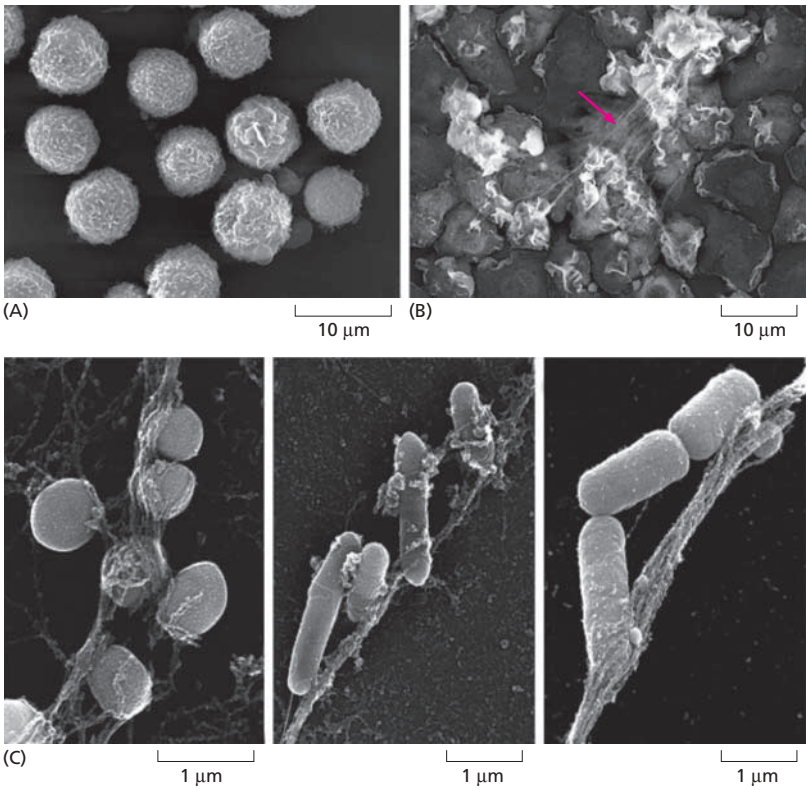


Figure 24–55 Neutrophils eject their chromatin to trap bacteria in a sticky web. (A) Scanning electron micrograph showing resting neutrophils. (B) Activated neutrophils have multiple surface ruffles and are associated with fibrous strands (*arrow*). The strands contain DNA and histones and appear to be chromatin from the neutrophil nucleus that has been ejected during the process of activation. (C) The sticky chromatin can entrap many different kinds of bacteria, including (left to right) *Staphylococcus aureus*, *Salmonella enterica*, and *Shigella flexneri*. (From V. Brinkmann et al., *Science* 303:1532–1535, 2004. With permission from AAAS.)

Activated Macrophages Contribute to the Inflammatory Response at Sites of Infection

When a pathogen invades a tissue, it almost always elicits an **inflammatory response**. Changes in local blood vessels cause a response characterized by pain, redness, heat, and swelling at the site of infection (physicians have recognized these four signs of inflammation, in Latin *dolor*, *rubor*, *calor*, and *turgor*, for thousands of years). The blood vessels dilate and become permeable to fluid and proteins, leading to local swelling and an accumulation of blood proteins that aid in defense, including components of the complement cascade. At the same time, the endothelial cells lining the local blood vessels are stimulated to express cell adhesion proteins (discussed in Chapter 19) that facilitate the attachment and extravasion of white blood cells, initially neutrophils, followed later by lymphocytes and monocytes (the blood-borne precursors of macrophages). <CCCC>

Whereas neutrophils usually die at the site of inflammation, macrophages frequently survive their initial encounter with invading pathogens and can migrate to other parts of the body. Pathogens that can survive within the macrophage, such as the bacterium *Salmonella enterica* serovar *Typhi*, for example, can use the macrophages as a transport system to spread a localized infection to distant sites in the body, converting a minor local invasion event in the gut into a severe systemic disease, typhoid fever.

Various signaling molecules mediate the inflammatory response at the site of an infection. Activation of TLRs results in the production of both lipid signaling molecules, such as prostaglandins, and protein (or peptide) signaling molecules, such as cytokines (discussed in Chapter 15), all of which contribute to the inflammatory response, as do the complement fragments released during complement activation. Some of the cytokines produced by activated macrophages are chemoattractants (called *chemokines*). Some of these attract neutrophils, which are the first cells recruited in large numbers to the site of a new infection. Other cytokines trigger *fever*, a rise in body temperature. On balance, fever helps fight

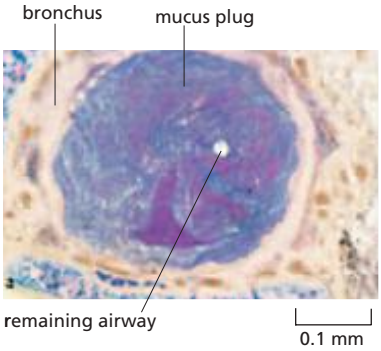


Figure 24–56 Inflammation of the airways in severe asthma. Light micrograph of a section through a bronchus of a patient who died during a prolonged and severe asthmatic attack. There is almost total occlusion of the airway by a mucus plug. The mucus plug is a dense inflammatory infiltrate that includes eosinophils, neutrophils, and lymphocytes. (Courtesy of Thomas Krausz.)

infection, since most bacterial and viral pathogens proliferate better at lower temperatures, whereas adaptive immune responses are more potent at higher temperatures. Still other proinflammatory signaling molecules stimulate endothelial cells to express proteins that trigger blood clotting in local small vessels. By occluding the vessels and cutting off blood flow, this response can help prevent the pathogen from entering the bloodstream and spreading the infection to other parts of the body.

The same inflammatory responses that help control local infections, however, can have disastrous consequences when they occur in response to a disseminated infection in the bloodstream, a condition called *sepsis*. The systemic release of proinflammatory signaling molecules into the blood causes dilation of blood vessels, and loss of plasma volume, which, together, cause a large fall in blood pressure, or *shock*; in addition, there is widespread blood clotting. The end result, known as *septic shock,* is often fatal. Inappropriate or overzealous local inflammatory responses can also contribute to chronic diseases, such as *asthma* (**Figure 24–56**) and arthritis.

Just as some pathogens have developed mechanisms to avoid the lethal consequences of phagocytosis, so some have acquired mechanisms to either prevent the inflammatory response or, in some cases, take advantage of it to spread the infection. Many viruses, for example, encode potent cytokine antagonists that block aspects of the inflammatory response. Some of these antagonists are simply modified forms of cytokine receptors, encoded by genes that the virus originally acquired from the host genome. They bind the cytokines with high affinity and block their activity. Some bacteria, such as *Salmonella*, induce an inflammatory response in the gut at the initial site of infection, thereby recruiting macrophages and neutrophils that they then invade. In this way, the bacteria hitch a ride to other tissues in the body.

Virus-Infected Cells Take Drastic Measures to Prevent Viral Replication

The microbe-associated immunostimulants on the surface of bacteria and parasites that are so important in eliciting innate immune responses against these pathogens are generally not present on the surface of viruses. Host cell ribosomes construct viral proteins, and host cell lipids form the membranes of enveloped viruses. The only general way that a host cell can recognize the presence of a virus is to detect unusual elements of the viral genome, such as the double-stranded RNA (dsRNA) that is an intermediate in the life cycle of many viruses. DNA virus genomes frequently contain significant amounts of CpG dinuclotide, which can be recognized by the Toll-like receptor TLR9, as discussed earlier.

Mammalian cells are particularly adept at recognizing the presence of dsRNA, and they can mobilize a program of intracellular responses to eliminate it. The program occurs in two steps. First, the cell degrades the dsRNA into small fragments (about 21–25 nucleotide pairs in length), using the enzyme *Dicer*. These double-stranded fragments bind to any single-stranded RNA (ssRNA) in the host cell that has the same sequence as either strand of the dsRNA fragment, leading to the destruction of the ssRNA. This dsRNA-directed ssRNA destruction is the basis of the technique of *RNA interference* (RNAi) that researchers use to destroy specific mRNAs and thereby block specific gene expression (discussed in Chapter 8). Second, the dsRNA induces the host cell to produce and secrete two cytokines—**interferon-α** (**IFNα**) and **interferon-β** (**IFNβ**), which act in both an

autocrine fashion on the infected cell and a paracrine fashion on uninfected neighbors. The binding of the interferons to their cell-surface receptors stimulates specific gene transcription by the Jak–STAT intracellular signaling pathway (see Figure 15–68), leading to the production of more than 300 gene products, including a large number of cytokines, reflecting the complexity of the cell's acute response to a viral infection.

The interferon response appears to be a general reaction of a mammalian cell to a viral infection, and viral components other than dsRNA can trigger it. In addition to their effects on host cell gene transcription, the interferons activate a latent ribonuclease, which nonspecifically degrades ssRNA. They also indirectly activate a protein kinase that phosphorylates and inactivates the protein synthesis initiation factor eIF-2, thereby shutting down most protein synthesis in the embattled host cell. Apparently, by destroying most of its RNA and transiently halting most of its protein synthesis, the host cell inhibits viral replication without killing itself. If these measures fail, the cell takes the even more extreme step of killing itself by apoptosis to prevent the virus from replicating, often with the help of a killer lymphocyte, as we discuss below and in Chapter 25.

Mammalian cells have a special defense mechanism to help them deal with retroviruses. These viruses activate a family of proteins called APOBEC (named because they are also members of the editing complex that modifies the mRNA for the protein ApoB, which is the major protein component of the low-density lipoprotein, LDL). These enzymes deaminate cytosines in nascent retroviral cDNAs, converting them into uridine and thereby generating large numbers of mutations in the viral genome, leading eventually to the termination of viral replication.

Not surprisingly, many viruses have acquired mechanisms to defeat or avoid these intracellular defense processes. Influenza virus encodes a protein that blocks the recognition of dsRNA by Dicer. HIV encodes a small protein that mediates the ubiquitylation and proteasome-mediated degradation of the APOBEC proteins. Many viruses, including most of those that are able to cause disease in healthy hosts, use various mechanisms to block the activation of the interferon pathway. Some viruses also inhibit host cell apoptosis, which can have the side-effect of promoting the development of cancer; this is one way in which the Epstein–Barr virus occasionally causes Burkitt's lymphoma.

Natural Killer Cells Induce Virus-Infected Cells to Kill Themselves

Interferons have other, less direct ways of blocking viral replication. One of these is to enhance the activity of **natural killer cells** (**NK cells**), which are part of the innate immune system. Like cytotoxic T cells of the adaptive immune system (discussed in Chapter 25), NK cells destroy virus-infected cells by inducing the infected cells to kill themselves by undergoing apoptosis. The ways in which cytotoxic T cells and NK cells distinguish virus-infected cells from uninfected cells, however, is different.

Both cytotoxic T cells and NK cells recognize the same special class of cell-surface proteins to detect virus-infected host cells, The proteins are called **class I MHC proteins**, <AAGT> because they are encoded by genes in the major histocompatibility complex; almost all vertebrate cells express these genes, and we discuss them in detail in Chapter 25. Cytotoxic T cells recognize peptide fragments of viral proteins bound to these MHC proteins on the surface of virus-infected cells. By contrast, NK cells monitor the level of class I MHC proteins on the surface of all host cells: high levels inhibit the killing activity of NK cells, so that NK cells selectively kill host cells expressing low levels, which are mainly virus-infected cells and some cancer cells (**Figure 24–57**).

The reason that class I MHC protein levels are often low on virus-infected cells is that many viruses have developed mechanisms to inhibit the expression of these proteins on the surface of the cells they infect, to avoid detection by cytotoxic T lymphocytes. Adenovirus and HIV, for example, encode proteins that block class I MHC gene transcription. Herpes simplex virus and cytomegalovirus

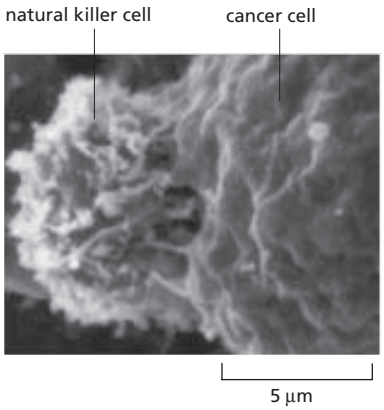


Figure 24–57 A natural killer (NK) cell attacking a cancer cell. The NK cell is the smaller cell on the *left*. This scanning electron micrograph was taken shortly after the NK cell attached, but before it induced the cancer cell to kill itself. (Courtesy of J.C. Hiserodt, in Mechanisms of Cytotoxicity by Natural Killer Cells [R.B. Herberman and D. Callewaert, eds.]. New York: Academic Press, 1995.)

block the peptide translocators in the ER membrane that transport proteasome-derived peptide fragments of viral proteins and host cell proteins from the cytosol into the lumen of the ER; such peptides are required for newly made class I MHC proteins to assemble in the ER membrane and be transported through the Golgi apparatus to the cell surface as peptide–MHC complexes (see Figure 25–59). Cytomegalovirus causes the retrotranslocation of class I MHC proteins from the ER membrane into the cytosol, where they are rapidly degraded in proteasomes. Proteins encoded by still other viruses prevent the delivery of assembled class I MHC proteins from the ER to the Golgi apparatus, or from the Golgi apparatus to the plasma membrane.

By evading recognition by cytotoxic T cells in these ways, however, a virus incurs the wrath of NK cells. The local production of IFNβ activates the killing activity of NK cells and also increases the expression of class I MHC proteins in uninfected cells. The cells infected with a virus that blocks class I MHC expression are thereby exposed as being different and become the victims of the activated NK cells. Thus, it is difficult for viruses to hide from both cytotoxic T cells and NK cells simultaneously. Remarkably, however, some large DNA viruses, including cytomegalovirus, encode MHC-like proteins that are expressed on the surface of the host cells they infect. Like *bona fide* class I MHC proteins, these MHC mimics activate inhibitory receptors on NK cells, and block the killing activity of the NK cells.

Both NK cells and cytotoxic T lymphocytes kill infected target cells by inducing them to undergo apoptosis before the virus has had a chance to replicate. It is not surprising, then, that many viruses have evolved mechanisms to inhibit apoptosis, particularly early in infection. As discussed in Chapter 18, apoptosis depends on an intracellular proteolytic cascade, which the cytotoxic cells can trigger, either through the activation of cell-surface death receptors or by injecting a proteolytic enzyme into the target cell (see Figure 24–47). Viral proteins can interfere with nearly every step in these pathways.

Dendritic Cells Provide the Link Between the Innate and Adaptive Immune Systems

Dendritic cells are crucially important cells of the innate immune system that are widely distributed in the tissues and organs of vertebrates. They display a large variety of pattern recognition receptors, including TLRs and NOD proteins, that enable the cells to recognize and phagocytose invading pathogens and to become activated in the process. The dendritic cells cleave the proteins of the pathogens into peptide fragments, which then bind to MHC proteins that carry the fragments to the cell surface. The activated dendritic cells now carry the pathogen-derived peptides, as complexes with MHC proteins, to a nearby lymphoid organ such as a lymph node, where they activate T cells of the adaptive immune system to join in the battle against the specific invader. In addition to the complexes of MHC proteins and microbial peptides displayed on their cell surface, activated dendritic cells also display special, cell-surface co-stimulatory proteins that help activate the T cells. The activated dendritic cells also secrete a variety of cytokines that influence the type of response that the T cells make, ensuring that it is appropriate to fight the particular pathogen. In these ways, dendritic cells serve as crucial links between the innate immune system, which provides a rapid first line of defense against invading pathogens, and the adaptive immune system, which provides slower but more powerful and highly specific ways of attacking an invader.

The battle between pathogens and host defenses is remarkably balanced. At present, humans seem to be gaining a slight advantage, using public sanitation measures, vaccines, and drugs to aid the efforts of our innate and adaptive immune systems. However, infectious diseases are still a leading cause of death worldwide, and new epidemics such as AIDS continue to emerge. The rapid evolution of pathogens and the almost infinite variety of ways that they can invade the human body and elude immune responses will prevent us from ever winning the battle completely.

In the next Chapter, we consider the unique and remarkable strategies that our adaptive immune system has evolved to defend us against such wily opponents. Amazingly, this immune system can mount pathogen-specific responses against pathogens that have never existed before.

Summary

Physical barriers preventing infection, cell-intrinsic responses to infection, and innate immune responses provide early lines of defense against invading pathogens. All multicellular organisms possess these defenses. In vertebrates, innate immune responses can also recruit specific and more powerful adaptive immune responses. Innate immune responses rely on the body's ability to recognize conserved features of microbial molecules that are not made by the host. These microbe-associated immunostimulants include many types of molecules on microbial surfaces, as well as the double-stranded RNA of some viruses. Many of these microbial molecules are recognized by pattern recognition receptors, including the toll-like receptors (TLRs) found in both plants and animals. In vertebrates, microbial surface molecules also activate complement, a group of blood proteins that are activated in sequence to target the microbe for phagocytosis by macrophages and neutrophils, to disrupt the membrane of the microbe, and to produce an inflammatory response. The phagocytes use a combination of degradative enzymes, antimicrobial peptides, and reactive oxygen species to kill the invading microorganism; in addition, they secrete signal molecules that trigger an inflammatory response. Cells infected with viruses produce interferons, which induce a series of cell responses, inhibit viral replication, and activate the killing activities of natural killer cells. Dendritic cells of the innate immune system ingest microbes at sites of infection and carry them and their products to local lymph nodes, where they activate T cells of the adaptive immune system to make specific responses against the microbes.

REFERENCES

General

Cossart P, Boquet P & Normark S (eds) (2005) Cellular Microbiology, 2nd ed. Washington: ASM Press.

Flint SJ, Enquist LW et al (2003) Principles of Virology: Molecular Biology, Pathogenesis, and Control of Animal Viruses, 2nd ed. Washington: ASM Press.

Murphy KM et al (2008) Janeway's Immunobiology, 7th ed. New York: Garland Science.

Salyers A & Whitt DD (2001) Bacterial Pathogenesis: A Molecular Approach, 2nd ed. Washington: ASM Press.

Schaechter M (ed) (1998) Mechanisms of Microbial Disease, 3rd ed. Philadelphia: Lippincott, Williams & Wilkins.

Introduction to Pathogens

Baltimore D (1971) Expression of animal virus genomes. *Bacteriol Rev* 35:235–241.

Bozdech Z, Llinas M, Pulliam BL et al (2003) The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol* 1:E5.

Brown NF, Wickham ME, Coombes BK & Finlay BB (2006) Crossing the line: selection and evolution of virulence traits. *PLoS Pathog* 2:e42.

Crick FHC & Watson JD (1956) Structure of small viruses. *Nature* 177374–475.

Gal-Mor O & Finlay BB (2006) Pathogenicity islands: a molecular toolbox for bacterial virulence. *Cell Microbiol* 8:1707–1719.

Galan JE & Wolf-Watz H (2006) Protein delivery into eukaryotic cells by type III secretion machines. *Nature* 444:567–573.

Hacker J & Kaper JB (2000) Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* 54:641–679.

Heidelberg JF, Eisen JA, Nelson WC et al (2000) DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* 406:477–483.

Lorber B (1996) Are all diseases infectious? *Annu Intern Med* 125:844–851.

Poulin R & Morand S (2000) The diversity of parasites. *Q Rev Biol* 75:277–293.

Prusiner SB (1996) Molecular biology and genetics of prion diseases. *Cold Spring Harb. Symp Quant Biol* 61:473–493.

Rappleye CA & Goldman WE (2006) Defining virulence genes in the dimorphic fungi. *Annu Rev Microbiol* 60:281–303

Rixon FJ (1990) Structure and assembly of herpesviruses. *Semin Virol* 1:477–487.

Stebbins CE & Galan JE (2001) Structural mimicry in bacterial virulence. *Nature* 412:701–705.

Cell Biology of Infection

Baranowski E, Ruiz-Jarabo CM & Domingo E (2001) Evolution of cell recognition by viruses. *Science* 292:1102–1105.

Berger EA, Murphy PM & Farber JM (1999) Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. *Annu Rev Immunol* 17:657–700.

Cossart P & Sansonetti PJ (2004) Bacterial invasion: the paradigms of enteroinvasive pathogens. *Science* 304:242–248.

Deitsch KW, Moxon ER & Wellems TE (1997) Shared themes of antigenic variation and virulence in bacterial, protozoal, and fungal infections. *Microbiol Mol Biol Rev* 61:281–293.

Garoff H, Hewson R & Opstelten DJ (1998) Virus maturation by budding. *Microbiol Mol Biol Rev* 62:1171–1190.

Ghedin E, Sengamalay NA, Shumway M et al (2005) Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature* 437:1162–1166.

Goldberg MB (2001) Actin-Based Motility of Intracellular Microbial Pathogens. *Microbiol Mol Biol Rev* 65:595–626.

Hackstadt T (2000) Redirection of host vesicle trafficking pathways by intracellular parasites. *Traffic* 1:93–99.

Jones NC (1990) Transformation by the human adenoviruses. *Semin Cancer Biol* 1:425–435.

Kaariainen L & Ranki M (1984) Inhibition of cell functions by RNA-virus infections. *Annu Rev Microbiol* 38:91–109.
Keeling PJ & Fast NM (2002) Microsporidia: biology and evolution of highly reduced intracellular parasites. *Annu Rev Microbiol* 56:93–116.
Kenny B, DeVinney R, Stein M et al (1997) Enteropathogenic *E. coli* (EPEC) transfers its receptor for intimate adherence into mammalian cells. *Cell* 91:511–520.
Knodler LA, Celli J & Finlay BB (2001) Pathogenic trickery: deception of host cell processes. *Nature Rev Mol Cell Biol* 2:578–88.
Lyles DS (2000) Cytopathogenesis and inhibition of host gene expression by RNA viruses. *Microbiol Mol Biol Rev* 64:709–724.
Mullins JI & Jensen MA (2006) Evolutionary dynamics of HIV-1 and the control of AIDS. *Curr Top Microbiol Immunol* 299:171–192.
Overbaugh J & Bangham CR (2001) Selection forces and constraints on retroviral sequence variation. *Science* 292:1106–1109.
Parrish CR & Kawaoka Y (2005) The origins of new pandemic viruses: the acquisition of new host ranges by canine parvovirus and influenza A viruses. *Annu Rev Microbiol* 59:553–586.
Pizarro-Cerda J & Cossart P (2006) Bacterial adhesion and entry into host cells. *Cell* 124:715–727.
Poranen MM, Daugelavicius R & Bamford DH (2002) Common principles in viral entry. *Annu Rev Microbiol* 56:521–538.
*Sibley LD & Andrews NW (2000) Cell invasion by un-palatable parasites. *Traffic* 1:100–106.
Skehel JJ & Wiley DC (2000) Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu Rev Biochem* 69:531–569.
Sodeik B (2000) Mechanisms of viral transport in the cytoplasm. *Trends Microbiol* 8:465–472.
Stephens EB & Compans RW (1988) Assembly of animal viruses at cellular membranes. *Annu Rev Microbiol* 42:489–516.
Tilney LG & Portnoy DA (1989) Actin filaments and the growth, movement, and spread of the intracellular bacterial parasite, *Listeria monocytogenes*. *J Cell Biol* 109:1597–1608.
Walsh C (2000) Molecular mechanisms that confer antibacterial drug resistance. *Nature* 406:775–781.
Wright GD (2007) The antibiotic resistome: the nexus of chemical and genetic diversity. *Nature Rev Microbiol* 5:175–186

Barriers to Infection and the Innate Immune System

Aderem A & Underhill DM (1999) Mechanisms of phagocytosis in macrophages. *Annu Rev Immunol* 17:593–623.
Banchereau J, Briere F, Caux C et al (2000) Immunobiology of dendritic cells. *Annu Rev Immunol* 18:767–811.
Brinkmann V, Reichard U, Goosmann C et al (2004) Neutrophil extracellular traps kill bacteria. *Science* 303:1532–1535.
Ganz T & Lehrer RI (1998) Antimicrobial peptides of vertebrates. *Curr Opin Immunol* 10:41–44.
Girardin SE, Sansonetti PJ & Philpott DJ (2002) Intracellular vs extracellular recognition of pathogens—common concepts in mammals and flies. *Trends Microbiol* 10:193–199.
Guidotti LG & Chisari FV (2001) Noncytolytic control of viral infections by the innate and adaptive immune response. *Annu Rev Immunol* 19:65–91.
Hedrick SM (2004) The acquired immune system: a vantage from beneath. *Immunity* 21:607–615.
Janeway CA Jr & Medzhitov R (2002) Innate immune recognition. *Annu Rev Immunol* 20:197–216.
Kimbrell DA & Beutler B (2001) The evolution and genetics of innate immunity. *Nature Rev Genet* 2:256–267.
Kopp E & Medzhitov R (2003) Recognition of microbial infection by Toll-like receptors. *Curr Opin Immunol* 15:396–401
Muller-Eberhard HJ (1988) Molecular organization and function of the complement system. *Annu Rev Biochem* 57:321–347.
Murphy PM (2001) Viral exploitation and subversion of the immune system through chemokine mimicry. *Nature Immunol* 2:116–122.
Ploegh HL (1998) Viral strategies of immune evasion. *Science* 280:248–253.
Segal AW (2005) How neutrophils kill microbes. *Annu Rev Immunol* 23:197–223.
Super M & Ezekowitz RA (1992) The role of mannose-binding proteins in host defense. *Infect Agents Dis* 1:194–199.
Tomlinson S (1993) Complement defense mechanisms. *Curr Opin Immunol* 5:83–89.
Yang RB, Mark MR, Gray A et al (1998) Toll-like receptor-2 mediates lipopolysaccharide-induced cellular signalling. *Nature* 395:284–288.
Yang D, Biragyn A, Kwak LW & Oppenheim JJ (2002) Mammalian defensins in immunity: more than just microbicidal. *Trends Immunol* 23:291–296.
Yokoyama WM, Kim S & French AR (2004) The dynamic life of natural killer cells. *Annu Rev Immunol* 22:405–429

25 The Adaptive Immune System

Our **adaptive immune system** saves us from certain death by infection. An infant born with a severely defective adaptive immune system will soon die unless extraordinary measures are taken to isolate it from a host of infectious agents, including bacteria, viruses, fungi, and parasites. All multicellular organisms need to defend themselves against infection by such potentially harmful invaders, collectively called **pathogens**. Invertebrates use relatively simple defense strategies that rely chiefly on protective barriers, toxic molecules, and phagocytic cells that ingest and destroy invading microorganisms *(microbes)* and larger parasites (such as worms). Vertebrates, too, depend on such **innate immune responses** as a first line of defense (discussed in Chapter 24), but they can also mount much more sophisticated defenses, called **adaptive immune responses**. In vertebrates, the innate responses call the adaptive immune responses into play, and both work together to eliminate the pathogens (**Figure 25–1**).

Whereas the innate immune responses are general defense reactions, the adaptive responses are highly specific to the particular pathogen that induced them, and they provide long-lasting protection. A person who recovers from measles, for example, is protected for life against measles by the adaptive immune system, although not against other common viruses, such as those that cause mumps or chickenpox. In this chapter, we focus on adaptive immune responses, and, unless we indicate otherwise, we use the term "immune responses" to refer to them.

Adaptive immune responses eliminate or destroy invading pathogens and any toxic molecules they produce. Because these responses are destructive, it is important that they are directed only against foreign molecules and not against molecules of the host itself. The adaptive immune system uses multiple mechanisms to avoid damaging responses against self molecules. Occasionally, however, these mechanisms fail, and the system turns against the host, causing *autoimmune diseases*, which can be fatal.

Many harmless foreign molecules enter the body, and it would be pointless and potentially dangerous to mount adaptive immune responses against them. Allergic conditions such as hayfever and allergic asthma are examples of deleterious adaptive immune responses against apparently harmless foreign molecules. An individual normally avoids such inappropriate responses because the innate immune system only calls adaptive immune responses into play when it recognizes conserved patterns of molecules that are specifically expressed by invading pathogens. The innate immune system can even distinguish between different classes of pathogens and recruit the most effective form of adaptive immune response to eliminate them.

Any substance capable of eliciting an adaptive immune response is referred to as an **antigen** (*anti*body *gen*erator). Most of what we know about such responses has come from studies in which an experimenter tricks the adaptive immune system of a laboratory animal (usually a mouse) into responding to a harmless foreign molecule, such as a foreign protein. The trick involves injecting the harmless molecule together with immunostimulants (usually microbial in origin) called *adjuvants,* which activate the innate immune system. This trick is called **immunization**. If administered in this way, almost any macromolecule, as long as it is foreign to the recipient, can induce an adaptive immune response

In This Chapter

LYMPHOCYTES AND THE CELLULAR BASIS OF ADAPTIVE IMMUNITY 1540

B CELLS AND ANTIBODIES 1551

THE GENERATION OF ANTIBODY DIVERSITY 1562

T CELLS AND MHC PROTEINS 1569

HELPER T CELLS AND LYMPHOCYTE ACTIVATION 1589

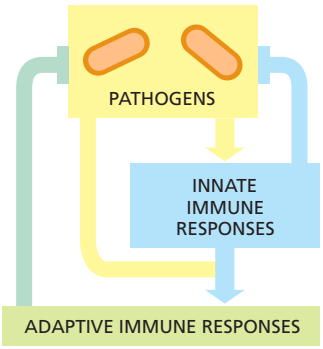


Figure 25–1 Innate and adaptive immune responses. Innate immune responses are activated directly by pathogens and defend all multicellular organisms against infection. In vertebrates, pathogens, together with the innate immune responses they activate, stimulate adaptive immune responses, which then work together with innate immune responses to help fight the infection.

that is specific to that macromolecule. Remarkably, the adaptive immune system can distinguish between antigens that are very similar—such as between two proteins that differ in only a single amino acid, or between two optical isomers of the same molecule. Thus, the adaptive immune system recognizes the fine molecular details of macromolecules.

Adaptive immune responses are carried out by white blood cells called **lymphocytes**. There are two broad classes of such responses—*antibody responses* and *T-cell-mediated immune responses*—and different classes of lymphocytes, called B cells and T cells, respectively, carry them out. In **antibody responses**, B cells are activated to secrete antibodies, which are proteins called *immunoglobulins*. The antibodies circulate in the bloodstream and permeate the other body fluids, where they bind specifically to the foreign antigen that stimulated their production (**Figure 25–2**). Binding of antibody inactivates viruses and microbial toxins (such as tetanus toxin or diphtheria toxin) by blocking their ability to bind to receptors on host cells. Antibody binding also marks invading pathogens for destruction, mainly by making it easier for phagocytic cells of the innate immune system to ingest them.

In **T-cell-mediated immune responses**, the second class of adaptive immune responses, activated T cells react directly against a foreign antigen that is presented to them on the surface of a host cell, which is therefore referred to as an *antigen-presenting cell*. Remarkably, T cells can detect microbes hiding inside host cells and either kill the infected cells or help the infected cells or other cells to eliminate the microbes. The T cell, for example, might kill a virus-infected host cell that has viral antigens on its surface, thereby eliminating the infected cell before the virus has had a chance to replicate (see Figure 25–2). In other cases, the T cell produces signal molecules that either activate macrophages to destroy the microbes that they have phagocytosed or help activate B cells to make antibodies against the microbes.

We begin this chapter by discussing the general properties of lymphocytes. We then consider the functional and structural features of antibodies that enable them to recognize and neutralize extracellular microbes and the toxins they make. Next, we discuss how B cells can produce a virtually unlimited number of different antibody molecules. Finally, we consider the special features of T cells and the immune responses they mediate.

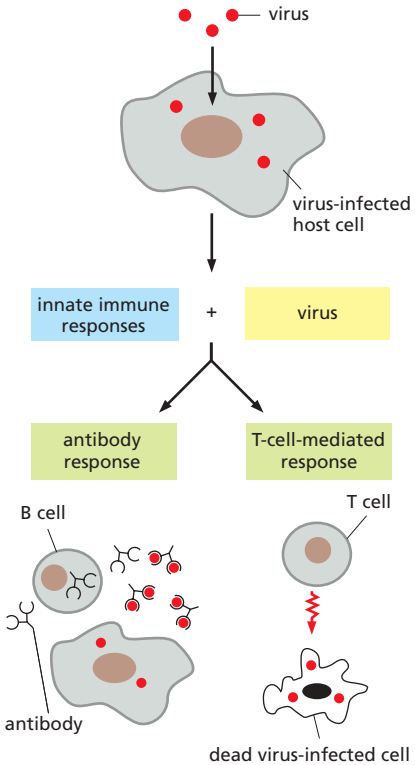


Figure 25–2 The two main classes of adaptive immune responses. Lymphocytes carry out both classes of responses. Here, the lymphocytes are responding to a viral infection. In one class of adaptive response, B cells secrete antibodies that neutralize the virus. In the other, a T-cell-mediated response, T cells kill the virus-infected cells. In both cases, innate immune responses help activate the adaptive immune responses through pathways that are not shown.

LYMPHOCYTES AND THE CELLULAR BASIS OF ADAPTIVE IMMUNITY

Lymphocytes are responsible for the astonishing specificity of adaptive immune responses. They occur in large numbers in the blood and lymph (the colorless fluid in the lymphatic vessels that connect the lymph nodes in the body to each other and to the bloodstream). They are also concentrated in **lymphoid organs**, such as the thymus, lymph nodes (also called lymph glands), spleen, and appendix (**Figure 25–3**). In this section, we discuss the general properties of lymphocytes that apply to both B cells and T cells.

Lymphocytes Are Required for Adaptive Immunity

There are about 2×10^{12} lymphocytes in the human body, making the immune system comparable in cell mass to the liver or the brain. Despite their abundance, their central role in adaptive immunity was not definitively demonstrated until

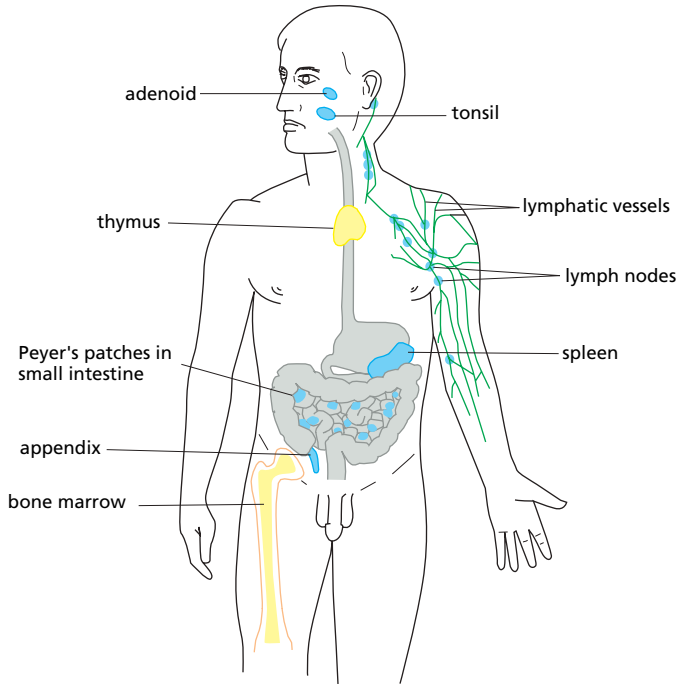


Figure 25–3 Human lymphoid organs. Lymphocytes develop in the thymus and bone marrow *(yellow),* which are therefore called *central* (or *primary*) *lymphoid organs*. The newly formed lymphocytes migrate from these primary organs to *peripheral* (or *secondary*) *lymphoid organs,* where they can react with foreign antigen. Only some of the peripheral lymphoid organs *(blue)* and lymphatic vessels *(green)* are shown; many lymphocytes, for example, are found in the skin and respiratory tract. As we discuss later, the lymphatic vessels ultimately empty into the bloodstream (not shown).

the late 1950s. The crucial experiments were performed in mice and rats that were heavily irradiated to kill most of their white blood cells, including lymphocytes. This treatment makes the animals unable to mount adaptive immune responses. Then, by transferring various types of cells into the animals it was possible to determine which cells reversed the deficiency. Lymphocytes were the only cell type able to restore the adaptive immune responses of irradiated animals, indicating that they are required for these responses (**Figure 25–4**).

The Innate and Adaptive Immune Systems Work Together

As mentioned earlier, lymphocytes usually respond to foreign antigens only if the **innate immune system** is first activated. As discussed in Chapter 24, the rapid innate immune responses to an infection depend largely on **pattern recognition receptors** made by cells of the innate immune system. These receptors recognize

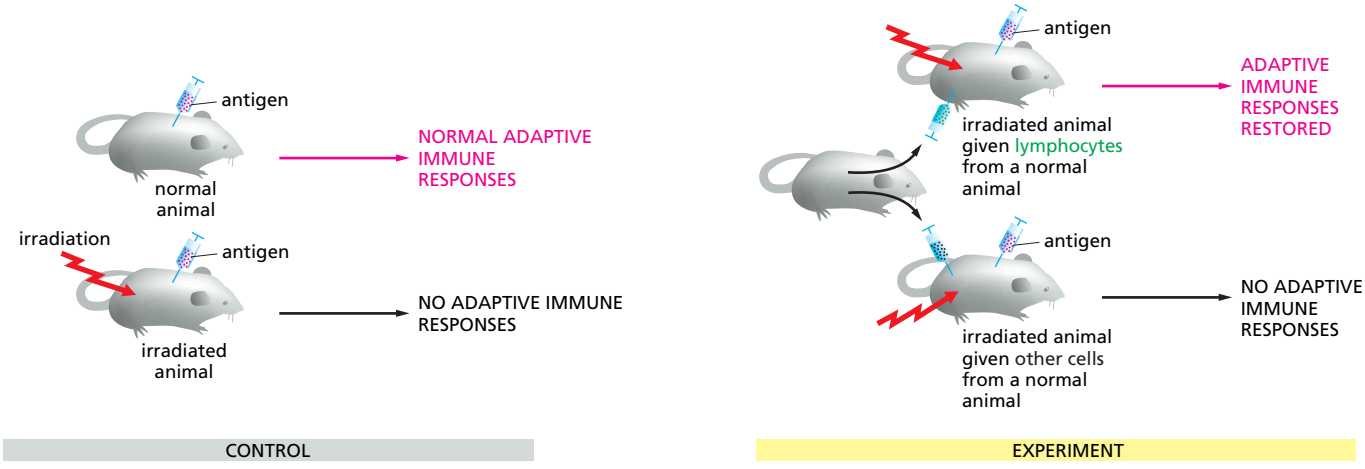


Figure 25–4 A classic experiment showing that lymphocytes are required for adaptive immune responses to foreign antigens. An important requirement of all such cell-transfer experiments is that cells are transferred between animals of the same *inbred strain.* Members of an inbred strain are genetically identical. If lymphocytes are transferred to a genetically different animal that has been irradiated, they react against the "foreign" antigens of the host and can kill the animal. In the experiment shown, the injection of lymphocytes restores both antibody and T-cell-mediated adaptive immune responses, indicating that lymphocytes are required for both types of responses.

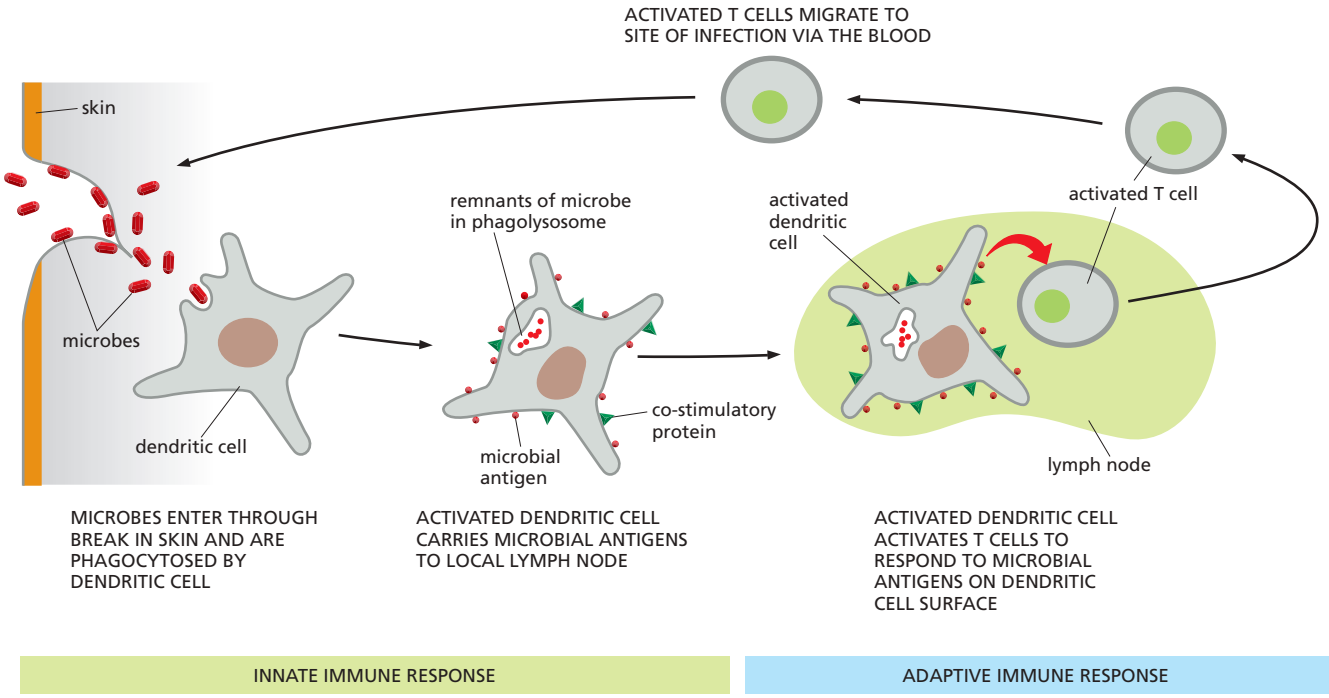


Figure 25–5 How the innate immune system can help activate the adaptive immune system. Dendritic cells ingest invading microbes or their products at the site of an infection. The microbial PAMPs activate the dendritic cells to express *co-stimulatory proteins* on their surface and to migrate in lymphatic vessels to a nearby lymph node. In the lymph node, the activated dendritic cells activate the small fraction of T cells that express a receptor for the microbial antigens displayed on the dendritic cell surface. These T cells proliferate and some then migrate to the site of infection, where they help eliminate the microbes, by either helping to activate macrophages or killing infected cells (not shown).

microbe-associated molecules that are not present in the host organism, called *microbe-associated immunostimulants*. Because they often occur in repeating patterns, they are also called *pathogen-associated molecular patterns (PAMPs)*. PAMPs include repeated patterns of molecular structure in microbial nucleic acids, lipids, polysaccharides, and proteins.

Some of the pattern recognition receptors are present on the surface of professional phagocytic cells (phagocytes) such as macrophages and neutrophils, where they mediate the uptake of pathogens, which are then delivered to lysosomes for destruction. Others are secreted and bind to the surface of pathogens, marking them for destruction by either phagocytes or a system of blood proteins collectively called the *complement system* (discussed in Chapter 24). Still others, including the *Toll-like receptors (TLRs)* discussed in Chapter 24, activate intracellular signaling pathways that lead to the secretion of extracellular signal molecules that promote inflammation and help activate adaptive immune responses.

The cells of the vertebrate innate immune system that respond to PAMPs and activate adaptive immune responses most efficiently are **dendritic cells**. Present in most tissues, dendritic cells express high levels of TLRs and other pattern recognition receptors, and they function by presenting microbial antigens to T cells in peripheral lymphoid organs. In most cases, they recognize and phagocytose invading microbes or their products or fragments of infected cells at a site of infection and then migrate with their prey to a nearby lymph node; in other cases, they pick up microbes or their products directly in a peripheral lymphoid organ such as the spleen. In either case, the microbial PAMPs activate the dendritic cells so that they, in turn, can directly activate the T cells in peripheral lymphoid organs to respond to the microbial antigens displayed on the dendritic cell surface. Once activated, some of the T cells then migrate to the site of infection, where they help destroy the microbes (**Figure 25–5**). Other activated T cells remain in the lymphoid organ, where they help keep the dendritic cells active, help activate other T cells, and help activate B cells to make antibodies against the microbial antigens.

Thus, innate immune responses are activated mainly at sites of infection (or injury), whereas adaptive immune responses are activated mainly in peripheral lymphoid organs such as lymph nodes and spleen. Both types of responses work together to eliminate invading pathogens and foreign macromolecules.

B Lymphocytes Develop in the Bone Marrow; T Lymphocytes Develop in the Thymus

T cells and B cells derive their names from the organs in which they develop. T cells develop in the *thymus*, and B cells, in mammals, develop in the *bone marrow* in adults or the liver in fetuses.

Both T and B cells are thought to develop from the same *common lymphoid progenitor cells*. The common lymphoid progenitor cells themselves derive from multipotential *hemopoietic stem cells*, which give rise to all of the blood cells, including red blood cells, white blood cells, and platelets. These stem cells (discussed in Chapter 23) are located primarily in *hemopoietic* tissues—mainly the liver in fetuses and the bone marrow in adults.

T cells develop in the thymus from common lymphoid progenitor cells that migrate there from the hemopoietic tissues via the blood. In most mammals, including humans and mice, B cells develop from common lymphoid progenitor cells in the hemopoietic tissues themselves (**Figure 25–6**). Because they are sites where lymphocytes develop from precursor cells, the thymus and hemopoietic tissues are referred to as **central** (**primary**) **lymphoid organs** (see Figure 25–3).

As we discuss later, most lymphocytes die in the central lymphoid organs soon after they develop, without ever functioning. Others, however, mature and migrate via the blood to the **peripheral** (**secondary**) **lymphoid organs**—mainly, the lymph nodes, spleen, and epithelium-associated lymphoid tissues in the gastrointestinal tract, respiratory tract, and skin (see Figure 25–3). It is in these peripheral lymphoid organs that foreign antigens activate T and B cells (see Figure 25–6).

T and B cells become morphologically distinguishable from each other only after they have been activated by antigen. Resting T and B cells look very similar, even in an electron microscope. Both are small, only marginally bigger than red blood cells, and contain little cytoplasm (**Figure 25–7**A). After activation by an antigen, both proliferate and mature into *effector cells*. Effector B cells secrete antibodies. In their most mature form, called *plasma cells*, they are filled with an extensive rough endoplasmic reticulum that is busily making antibodies (Figure 25–7B). In contrast, effector T cells (Figure 5–7C) contain very little endoplasmic reticulum and do not secrete antibodies; instead, they secrete a variety of signal proteins called **cytokines**, which act as local mediators.

There are three main classes of T cells—cytotoxic T cells, helper T cells, and regulatory (suppressor) T cells. *Cytotoxic T cells* directly kill infected host cells. *Helper T cells* help activate macrophages, dendritic cells, B cells, and cytotoxic T cells by secreting a variety of cytokines and displaying a variety of co-stimulatory proteins on their surface. *Regulatory T cells* are thought to use similar strategies

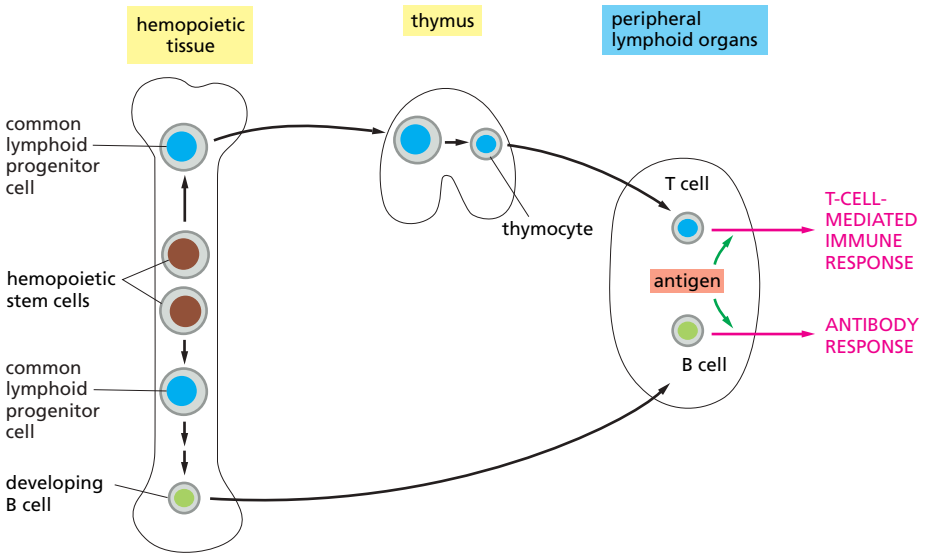
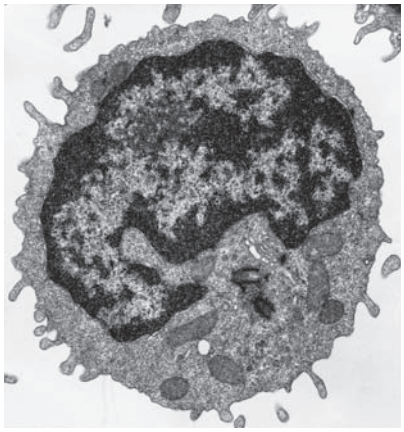


Figure 25–6 The development of T and B cells. The central lymphoid organs, where lymphocytes develop from common lymphoid progenitor cells, are labeled in *yellow boxes*. The common lymphoid progenitor cells develop from multipotent hemopoietic stem cells in the bone marrow. Some of the common lymphoid progenitor cells develop locally in the bone marrow into immature B cells, while others migrate to the thymus (via the bloodstream) where they develop into thymocytes (developing T cells). T cells and B cells are activated by foreign antigens mainly in peripheral lymphoid organs, such as lymph nodes or the spleen.



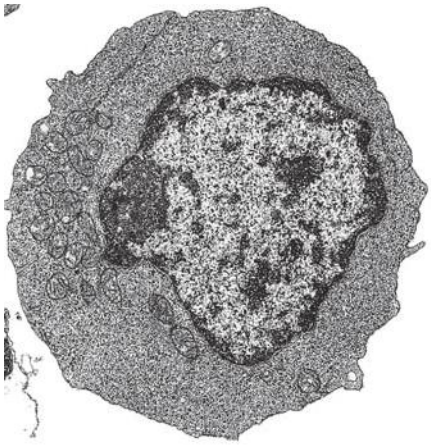
(A) resting T or B cell

1 μm



(B) effector B cell (plasma cell)

1 μm



(C) effector T cell

1 μm

to inhibit the function of helper T cells, cytotoxic T cells, and dendritic cells. Thus, whereas B cells can act over long distances by secreting antibodies that are widely distributed by the bloodstream, T cells can migrate to distant sites, but, once there, they act only locally on neighboring cells.

Figure 25–7 Electron micrographs of resting and effector lymphocytes. (A) This resting lymphocyte could be either a T cell or a B cell, as these cells are difficult to distinguish morphologically until they have been activated to become effector cells. (B) An effector B cell (a plasma cell). It is filled with an extensive rough endoplasmic reticulum (ER), which is distended with antibody molecules. (C) An effector T cell, which has relatively little rough ER but is filled with free ribosomes. The three cells are shown at the same magnification. (A, courtesy of Dorothy Zucker-Franklin; B, courtesy of Carlo Grossi; A and B, from D. Zucker-Franklin et al., Atlas of Blood Cells: Function and Pathology, 2nd ed. Milan, Italy: Edi. Ermes, 1988; C, courtesy of Stefanello de Petris.)

The Adaptive Immune System Works by Clonal Selection

The most remarkable feature of the adaptive immune system is that it can respond to millions of different foreign antigens in a highly specific way. Human B cells, for example, can make more than 10^{12} different antibody molecules that react specifically with the antigen that induced their production. How do B cells produce such a diversity of specific antibodies? The answer began to emerge in the 1950s with the formulation of the **clonal selection theory**. According to this theory, an animal first randomly generates a vast diversity of lymphocytes and then selects for activation those lymphocytes that can react against the foreign antigens that the animal actually encounters. As each lymphocyte develops in a central lymphoid organ, it becomes committed to react with a particular antigen before ever being exposed to the antigen. It expresses this commitment in the form of cell-surface receptor proteins that specifically bind the antigen. When a lymphocyte encounters its antigen in a peripheral lymphoid organ, the binding of the antigen to the receptors activates the lymphocyte, causing it to proliferate, thereby producing many more cells with the same receptor—a process called *clonal expansion* (as cells derived from a common ancestor cell are referred to as a *clone*). The encounter with antigen also causes the cells to differentiate into *effector cells*. An antigen therefore selectively stimulates those cells that express complementary antigen-specific receptors and are thus already committed to respond to it (**Figure 25–8**). This arrangement is what makes adaptive immune responses antigen-specific.

Compelling evidence supports the main tenets of the clonal selection theory. But how can the adaptive immune system produce lymphocytes that collectively display such an enormous diversity of receptors, including ones that recognize synthetic molecules that never occur in nature? We shall see later that, in humans, the antigen-specific receptors on both T and B cells are encoded by genes that are assembled from a series of gene segments by a special form of genetic recombination that occurs early in a lymphocyte's development, before it has encountered antigen. This assembly process generates an enormous diversity of receptors and lymphocytes, thereby enabling the immune system to respond to an almost unlimited variety of antigens.

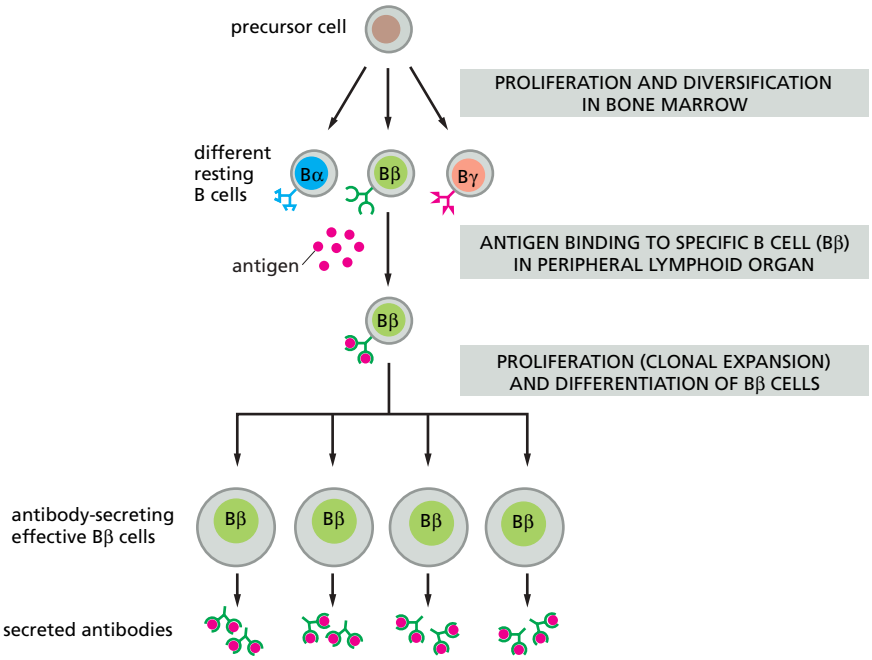


Figure 25–8 The clonal selection theory. An antigen activates only those lymphocytes that are already committed to respond to it. A cell committed to respond to a particular antigen displays cell-surface receptors that specifically recognize the antigen. The human immune system is thought to consist of many millions of different lymphocyte clones, with cells within a clone expressing the same unique receptor. Before their first encounter with antigen, a clone would usually contain only one or a small number of cells. A particular antigen may activate hundreds of different clones. Although only B cells are shown here, T cells operate in a similar way. Note that the receptors on B cells are antibody molecules and that those on the B cells labeled "Bβ" in this diagram bind the same antigen as do the antibodies secreted by the effector "Bβ" cells.

Most Antigens Activate Many Different Lymphocyte Clones

Most large molecules, including virtually all proteins and many polysaccharides, can act as antigens. Those parts of an antigen that bind to the antigen-binding site on either an antibody molecule or a lymphocyte receptor are called **antigenic determinants** (or *epitopes*). Most antigens have a variety of antigenic determinants that can stimulate the production of antibodies, specific T cell responses, or both. Some determinants of an antigen produce a greater response than others, so that the reaction to them may dominate the overall response. Such determinants are called *immunodominant*.

Any antigenic determinant is likely to activate many lymphocyte clones, each of which produces an antigen-binding site with its own characteristic affinity for the determinant. Even a relatively simple structure, like the *dinitrophenyl (DNP)* group in **Figure 25–9**, can be "looked at" in many ways. When it is coupled to a protein, as shown in the figure, it usually stimulates the production of hundreds of species of anti-DNP antibodies, each made by a different B cell clone. Such responses are said to be *polyclonal*. In *oligoclonal* responses, only a few clones are activated, and in *monoclonal* responses only a single B or T cell clone is activated. Monoclonal antibodies are widely used as tools in biology and medicine, but they have to be produced in a special way (see Figure 8–8), as the responses to most antigens are polyclonal.

Immunological Memory Involves Both Clonal Expansion and Lymphocyte Differentiation

The adaptive immune system, like the nervous system, can remember prior experiences. This is why we develop lifelong immunity to many common infectious diseases after our initial exposure to the pathogen, and it is why vaccination works. The same phenomenon can be demonstrated in experimental animals. If an animal is immunized once with antigen A, an immune response (antibody, T-cell-mediated, or both) appears after several days, rises rapidly and

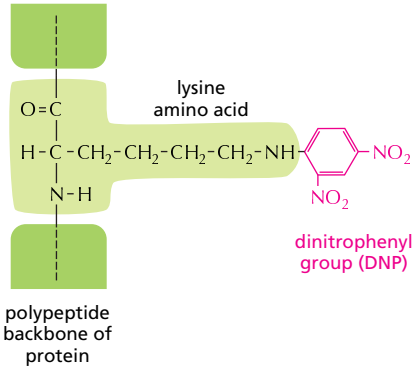


Figure 25–9 The dinitrophenyl (DNP) group. Although it is too small to induce an immune response on its own, when it is coupled covalently to a lysine side chain on a protein, as illustrated, DNP stimulates the production of hundreds of different species of antibodies that all bind specifically to it.

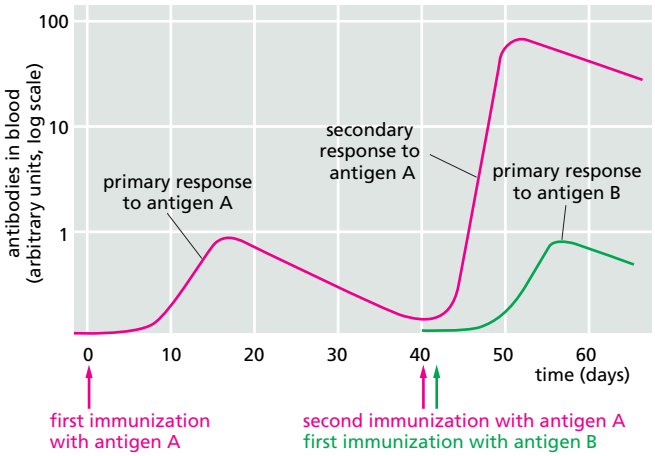


Figure 25–10 Primary and secondary antibody responses. The secondary response induced by a second exposure to antigen A is faster and greater than the primary response and is specific for A, indicating that the adaptive immune system has specifically remembered its previous encounter with antigen A. The same type of immunological memory is observed in T-cell-mediated responses. As we discuss later, the types of antibodies produced in the secondary response are different from those produced in the primary response, and these antibodies bind the antigen more tightly.

exponentially, and then, more gradually, declines. This is the characteristic course of a **primary immune response**, occurring on an animal's first exposure to an antigen. If, after some weeks, months, or even years have elapsed, the animal is immunized again with antigen A, it will usually produce a **secondary immune response** that differs from the primary response: the lag period is shorter, and the response is greater and more efficient. These differences indicate that the animal has "remembered" its first exposure to antigen A. If the animal is given a different antigen (for example, antigen B) instead of a second immunization with antigen A, the response is typical of a primary, and not a secondary, immune response. The secondary response must therefore reflect antigen-specific **immunological memory** for antigen A (**Figure 25–10**).

The clonal selection theory provides a useful conceptual framework for understanding the cellular basis of immunological memory. In an adult animal, the peripheral lymphoid organs contain a mixture of lymphocytes in at least three stages of maturation: *naïve cells, effector cells*, and *memory cells*. When **naïve cells** encounter their antigen for the first time, the antigen stimulates some of them to proliferate and differentiate into **effector cells**, which then carry out an immune response (effector B cells secrete antibody, while effector T cells either kill infected cells or influence the response of other cells). Some of the antigen-stimulated naïve cells multiply and differentiate into **memory cells**, which do not themselves carry out immune responses but are more easily and more quickly induced to become effector cells by a later encounter with the same antigen. When they encounter their antigen, memory cells (like naïve cells), give rise to either effector cells or more memory cells (**Figure 25–11**).

Thus, the primary response generates immunological memory because of clonal expansion, whereby the proliferation of antigen-stimulated naïve cells creates many memory cells, as well as because these memory cells are able to respond more sensitively, rapidly, and effectively to the same antigen than do naïve cells. And, unlike most effector cells, which die within days or weeks, memory cells can persist for the lifetime of the animal, even in the absence of their specific antigen, thereby providing lifelong immunological memory.

As we discuss later, memory B cells produce antibodies of different classes and of much higher affinity for antigen than those produced by naïve B cells. This is the main reason that secondary antibody responses are much more effective at eliminating pathogens than are primary antibody responses.

Although most effector T and B cells die after an immune response is over, some survive as effector cells and help provide long-term protection against the

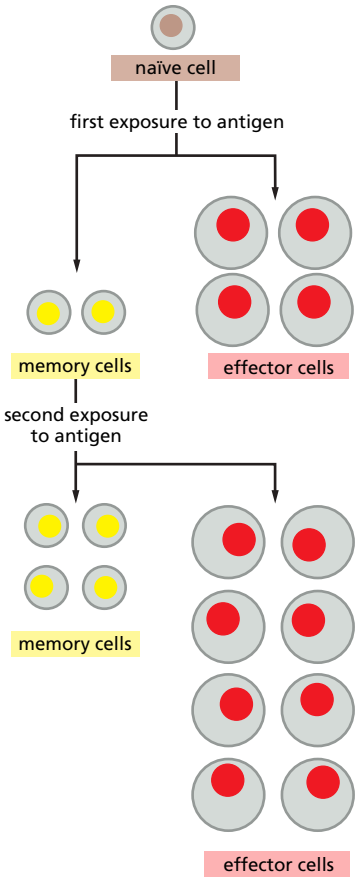


Figure 25–11 A model for the cellular basis of immunological memory. When stimulated by their specific antigen, naïve cells proliferate and differentiate. Most become effector cells, which function and then usually die, while others become memory cells. During a subsequent exposure to the same antigen, the memory cells respond more readily, rapidly, and efficiently than did the naïve cells: they proliferate and give rise to effector cells and to more memory cells. Memory T cells might also develop from effector cells (not shown).

pathogen. A small proportion of the plasma cells produced in a primary B cell response, for example, can survive for many months in the bone marrow, where they continue to secrete their specific antibodies into the bloodstream.

Immunological Tolerance Ensures That Self Antigens Are Not Normally Attacked

As discussed in Chapter 24, cells of the innate immune system use pattern recognition receptors to distinguish pathogens from the normal molecules of the host. The adaptive immune system has a far more difficult recognition task: it must be able to respond specifically to an almost unlimited number of foreign macromolecules, while avoiding responding to the large number of molecules made by the host organism itself. How is this possible? It helps that self molecules do not induce the innate immune reactions required to activate the adaptive immune system. But even when an infection or tissue injury triggers innate reactions, the vast excess of self molecules present normally fail to induce an adaptive immune response. Why not?

One answer is that the adaptive immune system has "learned" not to respond to self antigens. Transplantation experiments provide one line of evidence for this learning process. When tissues are transplanted from one individual to another (and the two individuals are not identical twins), the immune system of the recipient usually recognizes the donor cells as foreign and destroys them. (For reasons we discuss later, the foreign antigens on the donor cells are so powerful that they can stimulate adaptive immune responses in the absence of infection, injury, or an adjuvant.) If, however, we introduce cells from one strain of mouse into a newborn mouse of another strain, some of these cells survive for most of the recipient animal's life, and the recipient will now accept a graft from the original donor, even though it rejects "third-party" grafts. Apparently, nonself antigens can, in particular circumstances, induce the immune system to become specifically unresponsive to them. This antigen-specific unresponsiveness to foreign antigens is known as *acquired immunological tolerance* (**Figure 25–12**).

The unresponsiveness of an animal's adaptive immune system to its own macromolecules (*natural immunological tolerance*, or **self-tolerance**) is acquired in the same way. Normal mice, for example, cannot make an immune response against one of their own protein components of the complement system called C5 (discussed in Chapter 24). But, mutant mice that lack the gene encoding C5 (but are otherwise genetically identical to the normal mice) can make a strong immune response to this blood protein when immunized with it. Similarly, humans that lack a normal gene that codes for the clotting protein Factor VIII (and therefore bleed excessively) make antibodies against the protein when it is administered to them to control bleeding.

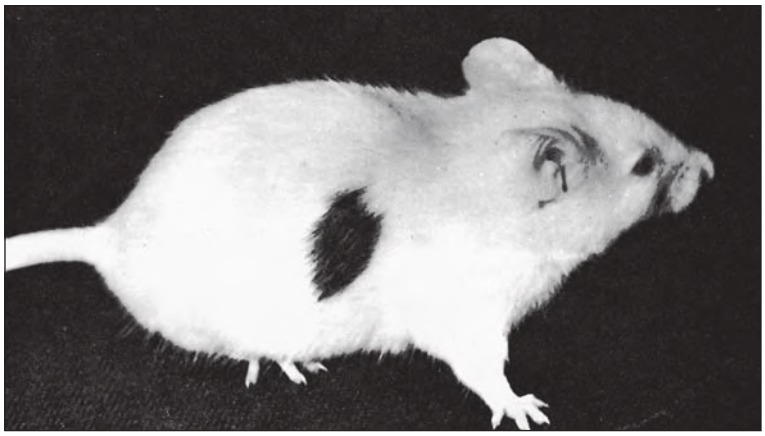
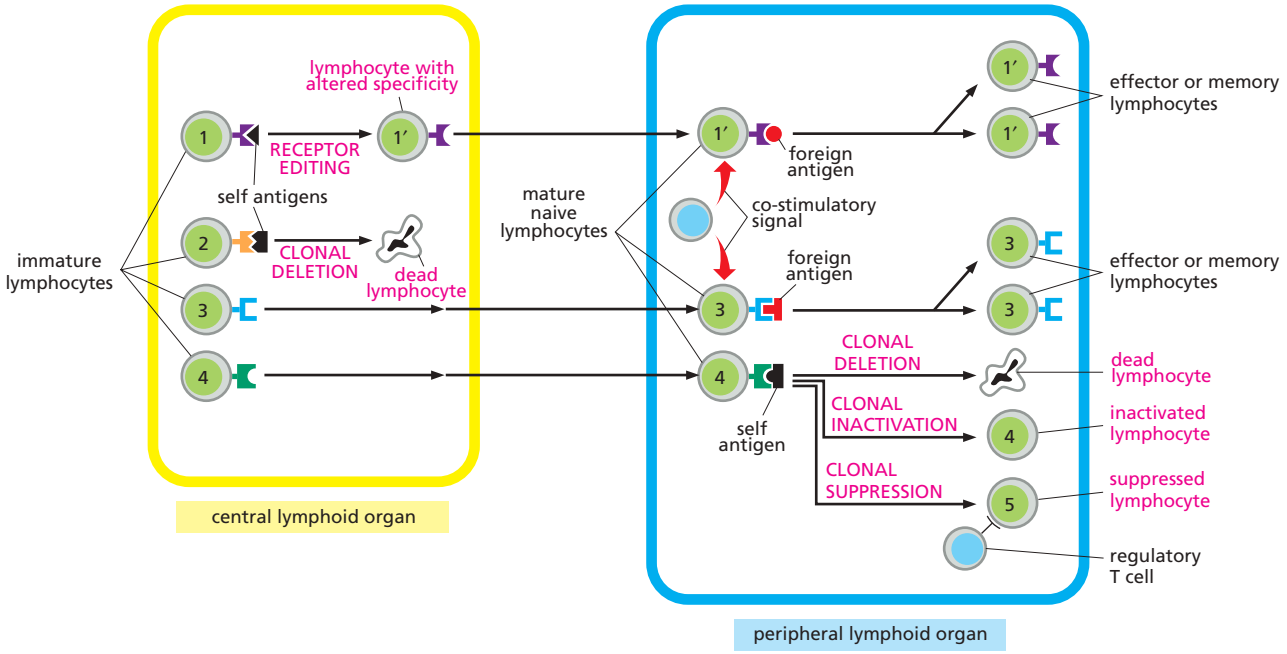


Figure 25–12 Acquired immunological tolerance. The skin graft seen here was transplanted from an adult brown mouse to an adult white mouse. It has survived for many weeks only because the white mouse, at the time of its birth, received an injection of bone marrow cells from the brown mouse and therefore became immunologically tolerant to them. Some of the bone marrow cells (and their progeny) from the brown mouse persist in the adult white mouse and continue to induce tolerance in newly formed lymphocytes that would otherwise react against the brown skin. (Courtesy of Leslie Brent, from I. Roitt, Essential Immunology, 6th ed. Oxford, UK: Blackwell Scientific, 1988.)



The natural immunological tolerance for a particular self molecule persists only for as long as the molecule remains present in the body. If a self molecule such as C5 is experimentally removed from an adult mouse, the animal gains the ability to respond to it after a few weeks or months. Thus, the immune system is genetically capable of responding to self molecules but learns not to do so.

Self-tolerance depends on a number of distinct mechanisms:

1. In *receptor editing,* developing lymphocytes that recognize self molecules *(self-reactive lymphocytes)* change their antigen receptors so that they no longer recognize self antigens.
2. In *clonal deletion,* self-reactive lymphocytes die by apoptosis when they bind their self antigen.
3. In *clonal inactivation* (also called clonal anergy), self-reactive lymphocytes become functionally inactivated when they encounter their self antigen.
4. In *clonal suppression,* regulatory T cells suppress the activity of self-reactive lymphocytes.

Some of these mechanisms—especially the first two, clonal deletion and receptor editing—operate in central lymphoid organs when newly formed self-reactive lymphocytes first encounter their self antigens, and they are largely responsible for the process of *central tolerance.* Clonal inactivation and clonal suppression, by contrast, operate mainly when lymphocytes encounter their self antigens in peripheral lymphoid organs, and they are responsible for the process of *peripheral tolerance.* Clonal deletion and clonal inactivation, however, are known to operate both centrally and peripherally (**Figure 25–13**).

Why does the binding of a self antigen lead to tolerance rather than activation? The answer is still not completely known. As we discuss later, to activate a lymphocyte in a peripheral lymphoid organ, the cell must do more than bind its antigen: it must also receive membrane-bound and secreted co-stimulatory signals (the secreted signals are various cytokines). Both types of signals are provided by a helper T cell in the case of a B lymphocyte and by an activated dendritic cell in the case of a T lymphocyte. Because the production of these signals is usually triggered by exposure to a pathogen, a self-reactive lymphocyte normally encounters its self antigen in the absence of such signals. Under these conditions, a B cell interacting with its antigen or a T cell interacting with its antigen on the surface of a nonactivated dendritic cell will not only fail to be activated, it will often be rendered tolerant—being either killed, inactivated, or actively suppressed by a regulatory T cell (see Figure 25–13). As we discuss later, in peripheral lymphoid organs, both T cell tolerance and activation usually occur on the surface of a dendritic cell.

Figure 25–13 Mechanisms of immunological tolerance to self antigens. When a self-reactive immature lymphocyte binds its self antigen in the central lymphoid organ where the cell is produced, it may alter its antigen receptor so that it is no longer self-reactive (cell 1). This process is called receptor editing and is thought to occur mainly in developing B cells. Alternatively, the cell may die by apoptosis, a process called clonal deletion (cell 2). Because these two forms of tolerance (shown on the left) occur in central lymphoid organs, they are called *central tolerance.*

When a self-reactive naïve lymphocyte escapes tolerance in the central lymphoid organ and binds its self antigen in a peripheral lymphoid organ (cell 4), it will generally not be activated, because the binding usually occurs in the absence of appropriate co-stimulatory signals; instead, the cell may die by apoptosis (often after a period of proliferation), be inactivated, or be suppressed by regulatory T cells (if the self-reactive lymphocyte is an effector T cell). These forms of tolerance shown on the right are called *peripheral tolerance.*

The tolerance mechanisms sometimes break down, causing T or B cells (or both) to react against the organism's own tissue antigens. *Myasthenia gravis* is an example of such an **autoimmune disease**. Affected individuals make antibodies against the acetylcholine receptors on their own skeletal muscle cells. These antibodies interfere with the normal functioning of the receptors so that the patients become weak and may die because they cannot breathe. Similarly, in *childhood (type 1) diabetes*, immune reactions against insulin-secreting cells in the pancreas kill these cells, leading to severe insulin deficiency.

For the most part, the mechanisms responsible for the breakdown of tolerance to self antigens in autoimmune diseases are unknown. It is thought, however, that activation of the innate immune system by infection or tissue injury may help trigger anti-self responses in individuals with defects in their self-tolerance mechanisms, leading to autoimmunity.

Lymphocytes Continuously Circulate Through Peripheral Lymphoid Organs

Pathogens generally enter the body through an epithelial surface, usually through the skin, gut, or respiratory tract. To induce an adaptive immune response, microbial antigens must travel from these entry points to a peripheral lymphoid organ, such as a lymph node or the spleen, the sites where lymphocytes are activated (see Figure 25–5). The route and destination depend on the site of entry. Lymphatic vessels (see Figure 25–3) carry antigens that enter through the skin or respiratory tract to local lymph nodes; antigens that enter through the gut end up in gut-associated peripheral lymphoid organs such as Peyer's patches; and the spleen filters out antigens that enter the blood. As discussed earlier, in many cases, dendritic cells will carry the antigen from the site of infection to the peripheral lymphoid organ, where they play a crucial part in activating T cells (see Figure 25–5).

But only a tiny fraction of the total lymphocyte population can recognize a particular microbial antigen in a peripheral lymph organ (estimated to be between 1/10,000 and 1/100,000 of each class of lymphocyte). How do these rare cells find an antigen-presenting cell displaying their antigen? The answer is that the lymphocytes continuously circulate between one peripheral lymphoid organ and another via the lymph and blood. In a lymph node, for example, lymphocytes continually leave the bloodstream by squeezing out between specialized endothelial cells lining small veins called *postcapillary venules*. After percolating through the node, they accumulate in small lymphatic vessels that leave the node and connect with other lymphatic vessels that pass through other lymph nodes downstream (see Figure 25–3). Passing into larger and larger vessels, the lymphocytes eventually enter the main lymphatic vessel (the *thoracic duct*), which carries them back into the blood (**Figure 25–14**).

The continuous recirculation between the blood and lymph ends only if a lymphocyte is activated by its specific antigen in a peripheral lymphoid organ. Now the lymphocyte remains in the peripheral lymphoid organ, where it proliferates and differentiates into either effector cells or memory cells. Many of the effector T cells leave the lymphoid organ via the lymph and migrate through the

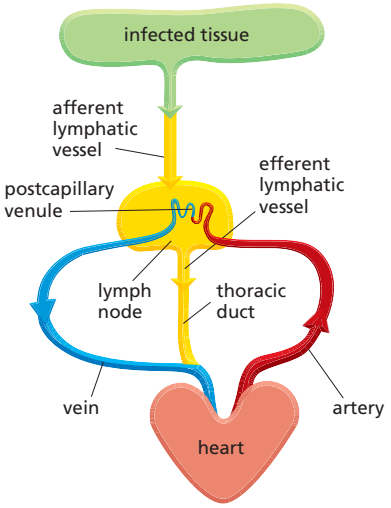


Figure 25–14 The path followed by lymphocytes as they continuously circulate between the lymph and blood. The circulation through a lymph node *(yellow)* is shown here. Microbial antigens are usually carried into the lymph node by dendritic cells (not shown), which enter the node via afferent lymphatic vessels draining an infected tissue *(green)*. T and B cells, by contrast, enter the lymph node via an artery and migrate out of the bloodstream through postcapillary venules. Unless they encounter their antigen, the T and B cells leave the lymph node via efferent lymphatic vessels, which eventually join the thoracic duct. The thoracic duct empties into a large vein carrying blood to the heart to complete the circulation process for T and B cells. A typical circulation cycle for these lymphocytes takes about 12–24 hours.

blood to the site of infection (see Figure 25–5), whereas others stay in the lymphoid organ and help activate B cells or other T cells there. Some effector B cells (plasma cells) remain in the peripheral lymphoid organ and secrete antibodies into the blood for days until they die; others migrate to the bone marrow, where they secrete antibodies into the blood for months or years. The memory T and B cells produced join the recirculating pool of lymphocytes.

Lymphocyte recirculation depends on specific interactions between the lymphocyte cell surface and the surface of the endothelial cells lining the blood vessels in the peripheral lymphoid organs. Many cell types in the blood come into contact with the specialized endothelial cells lining the postcapillary venules in lymph nodes, but only lymphocytes adhere and then migrate out of the bloodstream into the nodes. The lymphocytes initially adhere to the endothelial cells via *homing receptors* that bind to specific ligands (often called *counterreceptors*) on the endothelial cell surface. Lymphocyte migration into lymph nodes depends on a homing receptor protein called *L-selectin*, a member of the selectin family of cell-surface lectins. This protein binds to specific sugar groups on a counterreceptor that is expressed exclusively on the surface of the specialized endothelial cells lining the postcapillary venules in lymph nodes, causing the lymphocytes to adhere weakly to the endothelial cells and to roll slowly along their surface. The rolling continues until another, much stronger adhesion system is called into play by the chemoattractant proteins (called *chemokines;* see below) secreted by endothelial cells. This strong adhesion is mediated by members of the *integrin* family of cell adhesion molecules, which become activated on the lymphocyte surface. Now the lymphocytes stop rolling and crawl out of the blood vessel into the lymph node (**Figure 25–15**). Both selectins and integrins are discussed in Chapter 19.

Chemokines are a type of cytokine. They are small, secreted, positively charged proteins that have a crucial role in guiding the migrations of various types of cells, including white blood cells. <ACCG> They are all structurally related and bind to the surface of endothelial cells, as well as to negatively charged proteoglycans of the extracellular matrix in organs. By binding to G-protein-coupled chemokine receptors (discussed in Chapter 15) on the surface of specific blood cells, chemokines attract these cells from the bloodstream into an organ, guide them to specific locations within the organ, and then help stop migration. (Unfortunately, the AIDS virus, HIV, also binds to certain chemokine receptors, as well as to the CD4 co-receptor that we discuss later, and thus allows the virus to infect white blood cells.) The T and B cells initially enter the same region of a lymph node, but then different chemokines guide them to separate regions of the node—T cells to the *paracortex* and B cells to *lymphoid follicles* (**Figure 25–16**).

Unless they encounter their antigen, both T and B cells soon leave the lymph node via efferent lymphatic vessels. If they encounter their antigen, however, they are stimulated to display adhesion receptors that trap the cells in the node; the cells accumulate at the junction between the T cell and B cell areas, where the rare specific T and B cells can interact, leading to their proliferation and differentiation

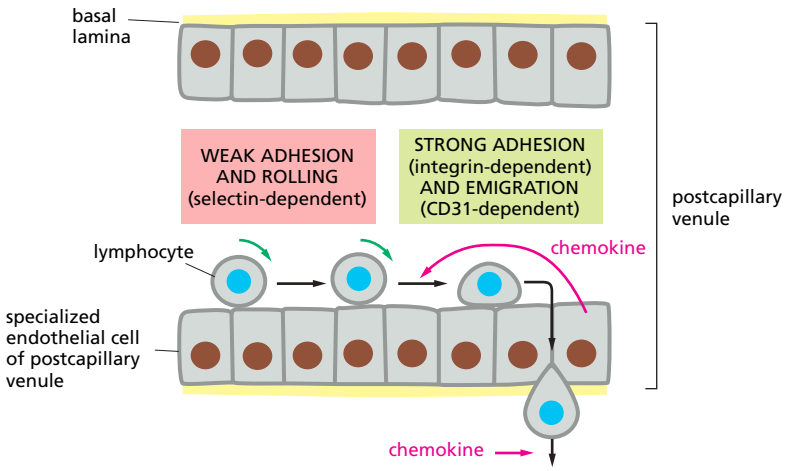


Figure 25–15 Migration of a lymphocyte out of the bloodstream into a lymph node. <CCCC> A circulating lymphocyte adheres weakly to the surface of the specialized endothelial cells lining a postcapillary venule in a lymph node. This initial adhesion is mediated by L-selectin on the lymphocyte surface. The adhesion is sufficiently weak to enable the lymphocyte to roll along the surface of the endothelial cells, pushed along by the flow of blood. Stimulated by chemokines secreted by the endothelial cells (curved *red* arrow), the lymphocyte rapidly activates a stronger adhesion system, mediated by an integrin. This strong adhesion enables the cell to stop rolling. The lymphocyte then uses an Ig-like cell adhesion protein (CD31) to bind to the junctions between adjacent endothelial cells and migrate out of the venule. CD31 is located both on the surface of the lymphocyte and at the junctions between the endothelial cells. The subsequent migration of the lymphocytes in the lymph node depends on chemokines produced within the node (straight *red* arrow).

The migration of other white blood cells out of the bloodstream into sites of infection occurs in a similar way.

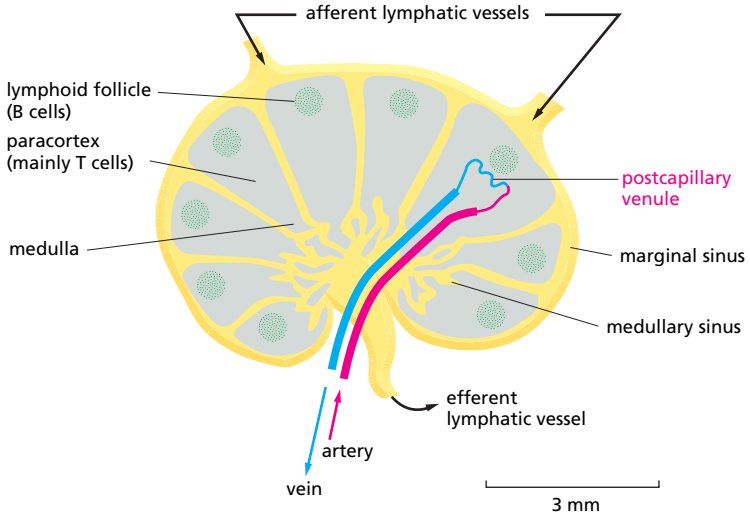


Figure 25–16 A simplified drawing of a human lymph node. B cells are primarily clustered in structures called lymphoid follicles, whereas T cells are found mainly in the paracortex. Chemokines attract both types of lymphocytes into the lymph node from the blood via postcapillary venules (see Figure 25–15). T and B cells then migrate to their respective areas, attracted by different chemokines. If they do not encounter their specific antigen, both T cells and B cells then enter the medullary sinuses and leave the node via the efferent lymphatic vessel. This vessel ultimately empties into the bloodstream, allowing the lymphocytes to begin another cycle of circulation through a peripheral lymphoid organ (see Figure 25–14).

If they encounter their specific antigen, T and B cells are retained in the node and are activated to become effector cells or memory cells; T cells and B cells responding to the same pathogen can interact in and around lymphoid follicles.

into either effector cells or memory cells. Many of the effector cells leave the node, expressing different chemokine receptors that help guide them to their new destinations—T cells to sites of infection and B cells to the bone marrow.

Summary

Innate immune responses are triggered at sites of infection by pathogen-associated molecular patterns (PAMPs), which are recognized by pattern recognition receptors made by cells of the innate immune system. In addition to fighting infection directly, these innate immune responses help activate adaptive immune responses in peripheral lymphoid organs. Unlike innate immune responses, adaptive responses display immunological memory and thereby provide specific and long-lasting protection against the particular pathogen that induced them.

The adaptive immune system is composed of many millions of lymphocyte clones, with the cells in each clone sharing a unique cell-surface receptor that enables them to bind a particular antigen. The binding of antigen to these receptors, however, is usually not sufficient to stimulate a lymphocyte to proliferate and differentiate into an effector cell that can help eliminate the pathogen. Membrane-bound co-stimulatory signals and a variety of secreted signals (cytokines) provided by another specialized cell in a peripheral lymphoid organ are also required. Helper T cells provide such signals for B cells, while dendritic cells usually provide them for T cells. Effector B cells secrete antibodies, which can act over long distances to help eliminate extracellular pathogens and their toxins. Effector T cells, by contrast, act locally to either kill infected host cells or help other cells to eliminate the pathogen. As part of the adaptive immune response, some lymphocytes proliferate and differentiate into memory cells, which are able to respond faster and more efficiently the next time the same pathogen invades. Both B and T cells circulate continuously between one peripheral lymphoid organ and another via the blood and lymph. Only if they encounter their specific foreign antigen in a peripheral lymphoid organ do they stop migrating, proliferate, and differentiate into effector cells or memory cells. Lymphocytes that would react against self molecules either alter their receptors or are eliminated, inactivated, or suppressed by regulatory T cells, so that the adaptive immune system normally avoids attacking the molecules and cells of the host.

B CELLS AND ANTIBODIES

Vertebrates inevitably die of infection if they are unable to make antibodies. Antibodies defend us against infection by binding to viruses and microbial toxins, thereby inactivating them (see Figure 25–2). When antibodies bind to invading pathogens, they also recruit some of the components of the innate immune system, including various types of white blood cells and components of the

complement system (discussed in Chapter 24). The white blood cells and a ctivated complement components work together to attack the invaders.

Synthesized exclusively by B cells, antibodies are produced in billions of forms, each with a different amino acid sequence. Collectively called **immunoglobulins** (abbreviated as **Ig**), they are among the most abundant protein components in the blood, constituting about 20% of the total protein in plasma by weight. Mammals make five classes of antibodies, each of which mediates a characteristic biological response following antigen binding. In this section, we discuss the structure and function of antibodies and how they interact with antigen.

B Cells Make Antibodies as Both Cell-Surface Antigen Receptors and Secreted Proteins

All antibody molecules made by an individual B cell have the same antigen-binding site. The first antibodies made by a newly formed B cell are not secreted but are instead inserted into the plasma membrane, where they serve as receptors for antigen. Each B cell has approximately 10^5 such receptors in its plasma membrane. As we discuss later, each of these receptors is stably associated with a complex of transmembrane proteins that activate intracellular signaling pathways when antigen on the outside of the cell binds to the receptor.

Each B cell clone produces a single species of antibody, with a unique antigen-binding site. When an antigen (with the aid of a helper T cell) activates a naïve or a memory B cell, that B cell proliferates and differentiates into an antibody-secreting effector cell. Such effector cells make and secrete large amounts of soluble (rather than membrane-bound) antibody, which has the same unique antigen-binding site as the cell-surface antibody that served earlier as the antigen receptor (**Figure 25–17**). Effector B cells can begin secreting antibody while they are still small lymphocytes, but the end stage of their maturation pathway is a large plasma cell (see Figure 25–7B), which continuously secretes antibodies at the astonishing rate of about 5000 molecules per second. Although most plasma cells die after several days, some survive in the bone marrow for months or years and continue to secrete antibodies into the blood, helping to provide long-term protection against the pathogen that stimulated their production.

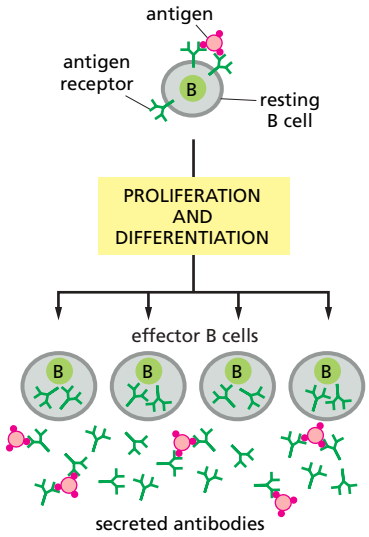


Figure 25–17 The membrane-bound and secreted antibodies made by a B cell clone. When an antigen (aided by a helper T cell—not shown) binds to and thereby activates either a naïve or a memory B cell, the cell proliferates and differentiates into effector cells. The effector cells produce and secrete antibodies with a unique antigen-binding site, which is the same as that of their original membrane-bound antibody that served as their antigen receptors.

A Typical Antibody Has Two Identical Antigen-Binding Sites

The simplest antibodies are Y-shaped molecules with two identical antigen-binding sites, one at the tip of each arm of the Y (**Figure 25–18**). Because of their two antigen-binding sites, they are described as *bivalent*. As long as an antigen has three or more antigenic determinants, bivalent antibody molecules can cross-link it into a large lattice (**Figure 25–19**) that macrophages can readily phagocytose and degrade. The efficiency of antigen binding and cross-linking is greatly increased by the flexible *hinge region* in most antibodies, which allows the distance between the two antigen-binding sites to vary (**Figure 25–20**).

The protective effect of antibodies is not due simply to their ability to bind and cross-link antigen. The tail of the Y-shaped molecule mediates many other activities of antibodies. As we discuss later, antibodies with the same antigen-binding sites can have any one of several different tail regions. Each type of tail region gives the antibody different functional properties, such as the ability to activate the complement system, to bind to phagocytic cells, or to cross the placenta from mother to fetus.

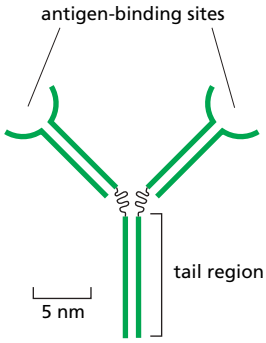


Figure 25–18 A simple representation of an antibody molecule. Note that its two antigen-binding sites are identical.

An Antibody Molecule Is Composed of Heavy and Light Chains

The basic structural unit of an antibody molecule consists of four polypeptide chains, two identical **light (L) chains** (each containing about 220 amino acids) and two identical **heavy (H) chains** (each usually containing about 440 amino

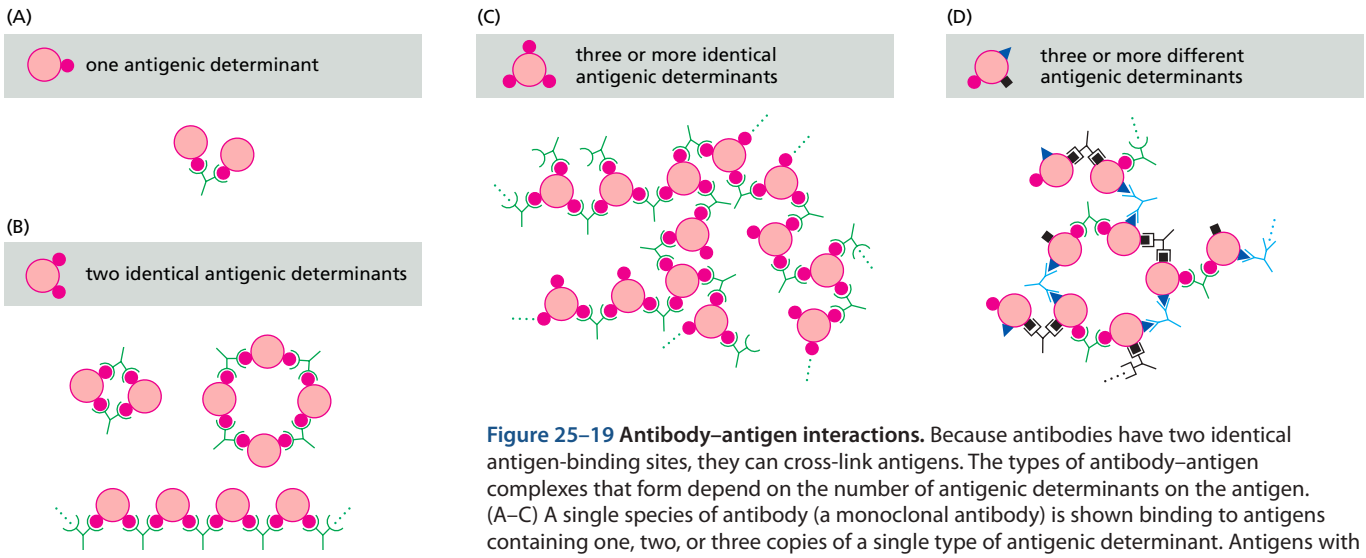


Figure 25–19 Antibody–antigen interactions. Because antibodies have two identical antigen-binding sites, they can cross-link antigens. The types of antibody–antigen complexes that form depend on the number of antigenic determinants on the antigen. (A–C) A single species of antibody (a monoclonal antibody) is shown binding to antigens containing one, two, or three copies of a single type of antigenic determinant. Antigens with two identical antigenic determinants can form small cyclic complexes or linear chains with the antibodies, while antigens with three or more identical antigenic determinants can form large three-dimensional lattices that readily precipitate out of solution. (D) Most antigens have many different antigenic determinants (see Figure 25–29A), and different antibodies that recognize different determinants can cooperate in cross-linking the antigen into large three-dimensional lattices.

acids). A combination of noncovalent and covalent (disulfide) bonds holds the four chains together. The molecule is composed of two identical halves, each with the same antigen-binding site. Both light and heavy chains usually cooperate to form the antigen-binding surface (**Figure 25–21**).

There Are Five Classes of Antibody Heavy Chains, Each with Different Biological Properties

In mammals, there are five *classes* of antibodies, IgA, IgD, IgE, IgG, and IgM, each with its own class of heavy chain—α, δ, ε, γ, and μ, respectively. IgA molecules have α chains, IgG molecules have γ chains, and so on. In addition, there are a number of subclasses of IgG and IgA immunoglobulins; for example, there are four human IgG subclasses (IgG1, IgG2, IgG3, and IgG4), having γ_1, γ_2, γ_3, and γ_4 heavy chains, respectively. The various heavy chains give a distinctive conformation to the hinge and tail regions of antibodies, so that each class (and subclass) has characteristic properties of its own.

IgM, which has μ heavy chains, is always the first class of antibody that a developing B cell makes, although many B cells eventually switch to making other classes of antibody when an antigen stimulates them (discussed below). The first cells in the B cell lineage that make Ig are *pro-B cells*, which make only μ chains. They give rise to *pre-B cells*, in which the μ chains associate with so-called *surrogate light chains* (substituting for genuine light chains) and insert into the plasma membrane. Signaling from this pre-B cell receptor is required for the cell to progress to the next stage of development, where it makes bona fide light chains. The light chains combine with the μ chains, replacing the surrogate light chains, to form four-chain IgM molecules (each with two μ chains and two light chains). These molecules then insert into the plasma membrane, where they function as receptors for antigen. At this point, the cell is called an *immature naïve B cell*. After leaving the bone marrow, the cell starts to produce cell-surface **IgD** molecules as well, with the same antigen-binding site as the IgM molecules. It is now called a *mature naïve B cell*. It is this cell that can respond to foreign antigen in peripheral lymphoid organs (**Figure 25–22**).

IgM is not only the first class of antibody to appear on the surface of a developing B cell. It is also the major class secreted into the blood in the early stages of a primary antibody response, on first exposure to an antigen. (Unlike IgM, IgD

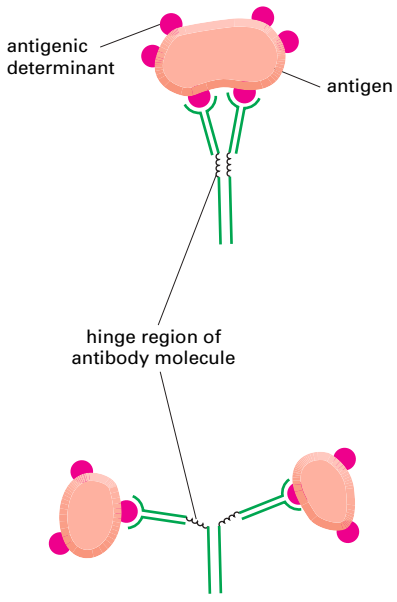


Figure 25–20 The hinge region of an antibody molecule. Because of its flexibility, the hinge region improves the efficiency of antigen binding and cross-linking.

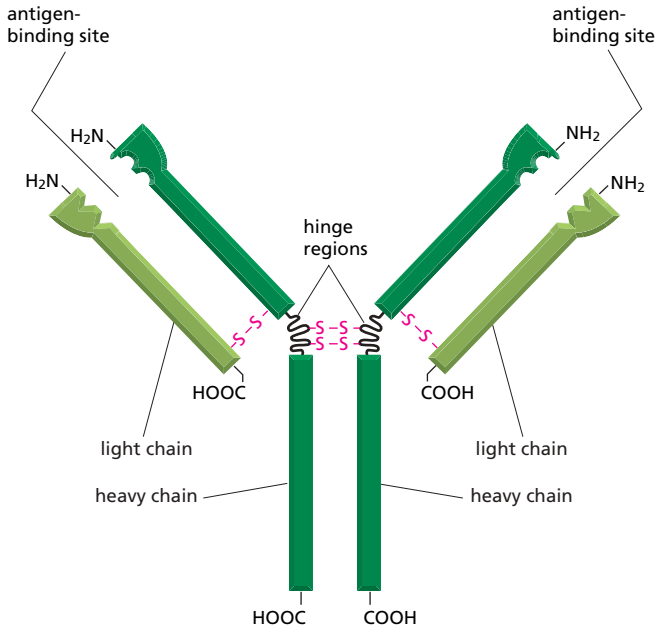


Figure 25–21 A schematic drawing of a bivalent antibody molecule. It is composed of four polypeptide chains—two identical heavy chains and two identical light chains. The two antigen-binding sites are identical, each formed by the N-terminal region of a light chain and the N-terminal region of a heavy chain. The two heavy chains also form both the tail and hinge region of the antibody.

molecules are secreted in only small amounts and seem to function mainly as cell-surface receptors for antigen.) In its secreted form, IgM is a pentamer composed of five four-chain units, giving it a total of 10 antigen-binding sites. Each pentamer contains one copy of another polypeptide chain, called a *J (joining) chain*. The J chain is produced by IgM-secreting cells and is covalently inserted between two adjacent tail regions (**Figure 25–23**).

When an antigen with multiple identical antigenic determinants (see Figure 25–19) binds to a single secreted pentameric IgM molecule, it alters the structure of the pentamer, allowing it to activate the complement system. As discussed in Chapter 24, when the antigen is on the surface of an invading pathogen, this activation of complement can either mark the pathogen for phagocytosis or kill it directly. As we discuss later, complement activation can also greatly increase the immune response to an antigen: the binding of an activated complement component to an antibody–antigen complex, for example, can increase the ability of the antigen to stimulate a B cell response more than a thousand fold (see Figure 25–71A).

The major class of immunoglobulin in the blood is **IgG**, which is a four-chain monomer (see Figure 25–21) produced in large quantities during secondary antibody responses. Besides activating complement, the tail region of an IgG molecule binds to specific receptors on macrophages and neutrophils.

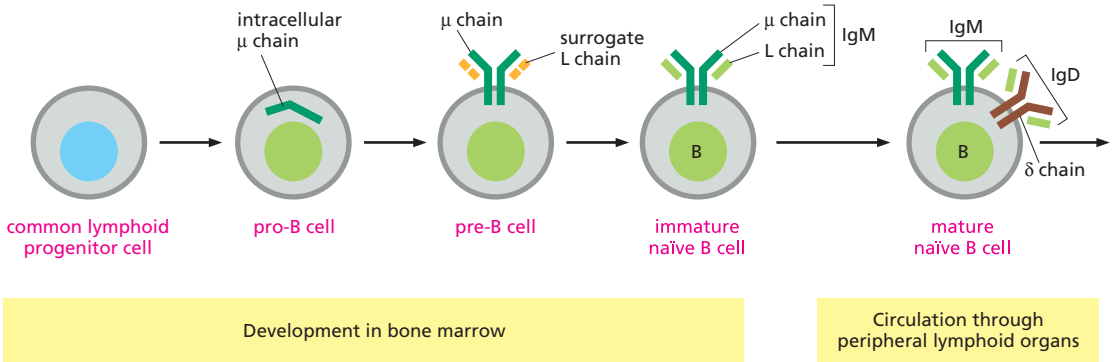


Figure 25–22 The main stages in B cell development. All of the stages shown occur independently of antigen. The pro-B cell makes μ chains, but they remain in the endoplasmic reticulum until surrogate light chains are made. Although not shown, all of the cell-surface Ig molecules are associated with transmembrane proteins that help convey signals to the cell interior (see Figure 25–70). When they are activated by their specific foreign antigen and helper T cells in peripheral lymphoid organs, mature naïve B cells proliferate and differentiate into either antibody-secreting cells or memory cells (not shown).

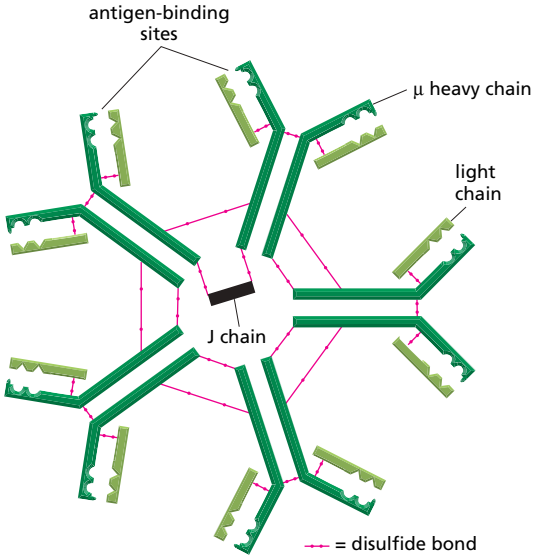


Figure 25–23 A pentameric IgM molecule. Disulfide bonds *(red)* hold the five four-chain units together. A single J chain, which has a structure similar to that of a single Ig domain (discussed later), is covalently attached by disulfide bonds to the tails of two μ heavy chains. The J chain is required for pentamer formation. The addition of each successive four-chain IgM unit requires a J chain, which is then discarded, except for the last one, which is retained. Note that IgM molecules do not have hinge regions.

Largely by means of such **Fc receptors** (so-named because antibody tails are called *Fc* regions), these phagocytic cells bind, ingest, and destroy infecting microorganisms that have become coated with the IgG antibodies produced in response to the infection (**Figure 25–24**).

Some IgG subclasses are the only antibodies that can pass from mother to fetus via the placenta. Cells of the placenta that are in contact with maternal blood have Fc receptors that bind these blood-borne IgG molecules and direct their passage to the fetus. The antibody molecules bound to the receptors are first taken into the placental cells by receptor-mediated endocytosis. They are then transported across the cell in vesicles and released by exocytosis into the fetal blood (a process called *transcytosis*—see Figure 25–26). Because other classes of antibodies do not bind to these particular Fc receptors, they cannot pass across the placenta. Later, IgG is secreted into the mother's milk and is then

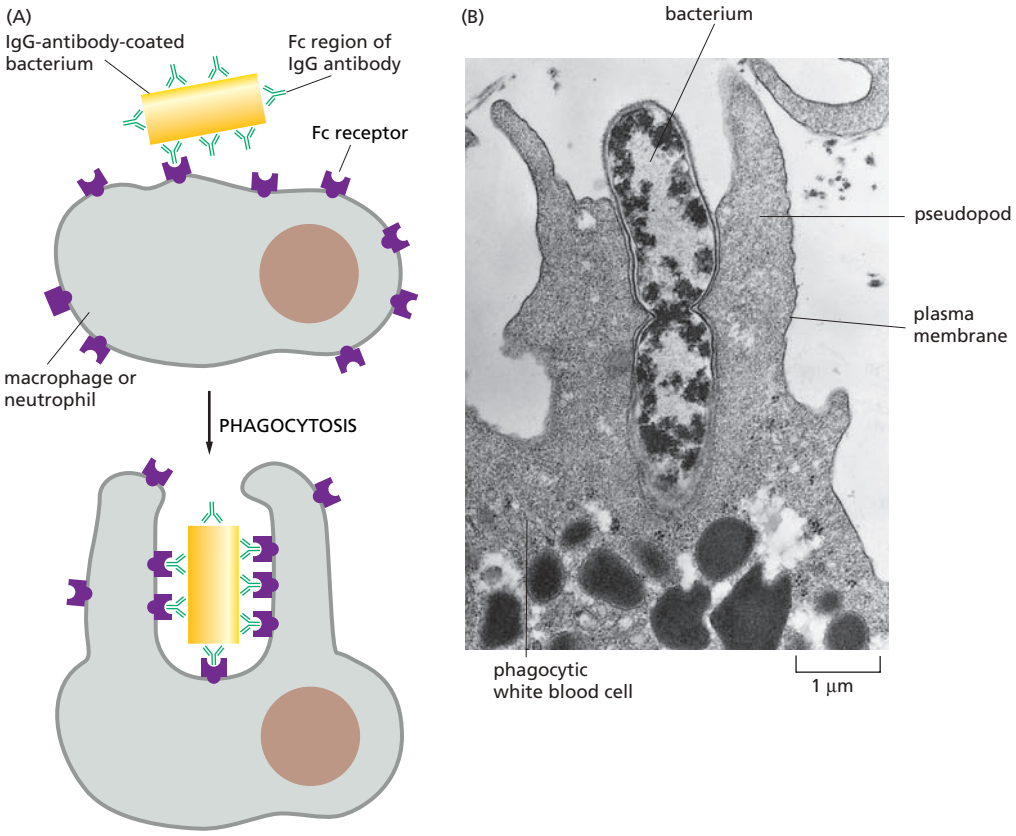


Figure 25–24 Antibody-activated phagocytosis. (A) An IgG-antibody-coated bacterium is efficiently phagocytosed by a macrophage or neutrophil, which has cell-surface receptors that bind the tail (Fc) region of IgG molecules. The binding of the antibody-coated bacterium to these Fc receptors activates the phagocytic process. The tail of an antibody molecule is called an Fc region because, when antibodies are cleaved with the proteolytic enzyme papain, the fragments containing the tail region readily crystallize. (B) Electron micrograph of a neutrophil phagocytosing an IgG-coated bacterium, which is in the process of dividing. The process in which antibody (or complement) coating of a pathogen increases the efficiency with which the pathogen is phagocytosed is called *opsonization.* (B, courtesy of Dorothy F. Bainton, from R.C. Williams, Jr. and H.H. Fudenberg, Phagocytic Mechanisms in Health and Disease. New York: Intercontinental Medical Book Corporation, 1971.)

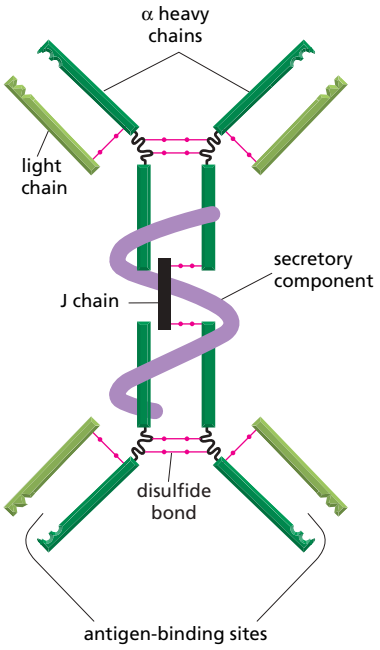


Figure 25–25 A highly schematized diagram of a dimeric IgA molecule found in secretions. In addition to the two IgA monomers, there is a single J chain and an additional polypeptide chain called the *secretory component*, which is derived from the Fc receptor (see Figure 25–26) and is thought to protect the IgA molecules from digestion by proteolytic enzymes in secretions.

taken up from the gut of the neonate into the blood by transcytosis, providing protection for the baby against infection.

IgA is the principal class of antibody in secretions, including saliva, tears, milk, and respiratory and intestinal secretions. IgA is a four-chain monomer in the blood which is assembled into a dimer by the addition of two other polypeptide chains before it is released into secretions (**Figure 25–25**). It is transported through secretory epithelial cells from the extracellular fluid into the secreted fluid by transcytosis mediated by another type of Fc receptor that is unique to secretory epithelia (**Figure 25–26**). This Fc receptor can also transport IgM into secretions (but less efficiently), which is probably why individuals with a selective IgA deficiency, the most common form of antibody deficiency, are only mildly affected by the defect.

The tail region of **IgE** molecules, which are four-chain monomers, binds with unusually high affinity (K_a ~ 10^{10} liters/mole) to yet another class of Fc receptors. These receptors are located on the surface of *mast cells* in tissues and of *basophils* in the blood. The IgE molecules bound to them function as passively acquired receptors for antigen. Antigen binding triggers the mast cell or basophil to secrete a variety of cytokines and biologically active amines, especially *histamine* (**Figure 25–27**). The histamine causes blood vessels to dilate and become leaky, which in turn helps white blood cells, antibodies, and complement components to enter sites where mast cells have been activated. The release of amines from mast cells and basophils is largely responsible for the symptoms of such *allergic* reactions as hay fever, asthma, and hives. In addition, mast cells secrete factors that attract and activate white blood cells called *eosinophils*. Eosinophils also have Fc receptors that bind IgE molecules, and they can kill extracellular parasitic worms, especially if the worms are coated with IgE antibodies.

In addition to the five classes of heavy chains found in antibody molecules, higher vertebrates have two types of light chains, **κ** and **λ**, which seem to be functionally indistinguishable. Either type of light chain may be associated with any of the heavy chains. An individual antibody molecule, however, always contains identical light chains and identical heavy chains: an IgG molecule, for instance, may have either κ or λ light chains, but not one of each. As a result, an antibody's antigen-binding sites are always identical. Such symmetry is crucial for the cross-linking function of secreted antibodies (see Figure 25–19).

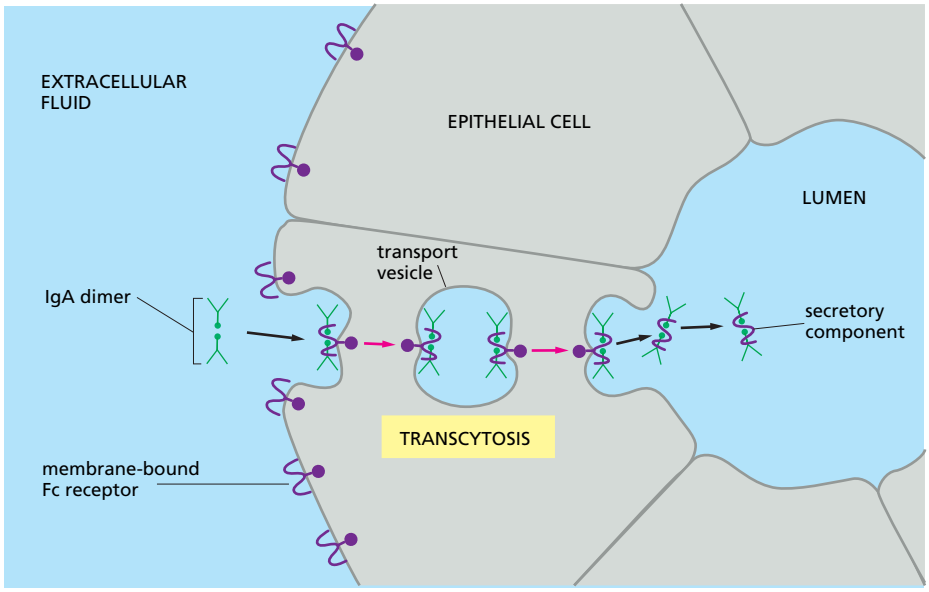


Figure 25–26 The mechanism of transport of a dimeric IgA molecule across an epithelial cell. The IgA molecule, as a J-chain-containing dimer, binds to a transmembrane receptor protein on the nonluminal surface of a secretory epithelial cell. (The J chain has been omitted in this diagram for clarity.) The receptor–IgA complexes are ingested by receptor-mediated endocytosis, transferred across the epithelial cell cytoplasm in vesicles, and secreted into the lumen on the opposite side of the cell by exocytosis. When exposed to the lumen, the part of the Fc receptor protein that is bound to the IgA dimer (the *secretory component*) is cleaved from its transmembrane tail, thereby releasing the antibody in the form shown in Figure 25–25.

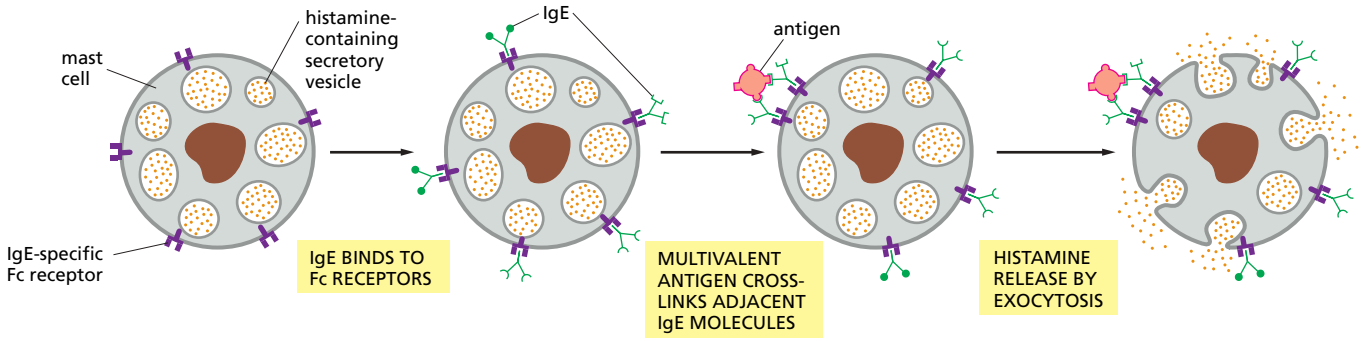


Figure 25–27 The role of IgE in histamine secretion by mast cells. A mast cell (or a basophil) binds IgE molecules after they are secreted by effector B cells. The soluble IgE antibodies bind to Fc receptor proteins on the mast cell surface that specifically recognize the Fc region of these antibodies. The bound IgE molecules serve as cell-surface receptors for antigen. Thus, unlike B cells, each mast cell (and basophil) has a set of cell-surface antibodies with a wide variety of antigen-binding sites. When an antigen molecule binds to these membrane-bound IgE antibodies so as to cross-link them to their neighbors, it signals the mast cell to release its histamine and other local mediators by exocytosis.

All classes of antibody can be made in a membrane-bound form, as well as in a soluble, secreted form. The two forms differ only in the C-terminus of their heavy chain. The heavy chains of membrane-bound antibody molecules have a transmembrane hydrophobic C-terminus, which anchors them in the lipid bilayer of the B cell's plasma membrane. The heavy chains of secreted antibody molecules, by contrast, have instead a hydrophilic C-terminus, which allows them to escape from the cell. The switch in the character of the antibody molecules made occurs because the activation of B cells by antigen (and helper T cells) induces a change in the way in which the H-chain RNA transcripts are made and processed in the nucleus (see Figure 7–99).

The properties of the various classes of antibodies in humans are summarized in **Table 25–1**.

The Strength of an Antibody–Antigen Interaction Depends on Both the Number and the Affinity of the Antigen-Binding Sites

The binding of an antigen to an antibody, like the binding of a substrate to an enzyme, is reversible. The sum of many relatively weak non-covalent forces, including hydrogen bonds, hydrophobic and van der Waals forces, and ionic interactions determine the strength of the interaction. These weak forces are effective only when the antigen molecule is close enough to allow some of its atoms to fit into complementary recesses on the surface of the antibody. The complementary regions of a four-chain antibody unit are its two identical antigen-binding sites; the corresponding region on the antigen is an antigenic determinant (**Figure 25–28**). Most antigenic macromolecules have many different antigenic determinants and are said to be *multivalent;* if two or more of the determinants are identical (as in a polymer with a repeating structure), the antigen is said to be *polyvalent* (**Figure 25–29**).

Table 25–1 Properties of the Major Classes of Antibodies in Humans

PROPERTIES	CLASS OF ANTIBODY				
	IgM	IgD	IgG	IgA	IgE
Heavy chains	μ	δ	γ	α	ε
Light chains	κ or λ	κ or λ	κ or λ	κ or λ	κ or λ
Number of four-chain units	5	1	1	1 or 2	1
Percentage of total Ig in blood	10	<1	75	15	<1
Activates complement	++++	–	++	–	–
Crosses placenta	–	–	+	–	–
Binds to macrophages and neutrophils	–	–	+	–	–
Binds to mast cells and basophils	–	–	–	–	+

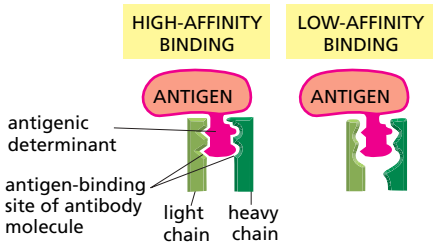


Figure 25–28 Antigen binding to antibody. In this highly schematized diagram, an antigenic determinant on a macromolecule is shown interacting with one of the antigen-binding sites of two different antibody molecules, one of high affinity and one of low affinity. Various weak noncovalent forces hold the antigenic determinant in the binding site, and the site with the better fit to the antigen has a greater affinity. Note that both the light and heavy chains of the antibody molecule usually contribute to the antigen-binding site.

The reversible binding reaction between an antigen with a single antigenic determinant (denoted Ag) and a single antigen-binding site (denoted Ab) can be expressed as

$$Ag + Ab \leftrightarrow AgAb$$

The equilibrium point depends both on the concentrations of Ab and Ag and on the strength of their interaction. Clearly, a larger fraction of Ab will become associated with Ag as the concentration of Ag increases. The strength of the interaction is generally expressed as the **affinity constant** (K_a) (see Figure 3–43), where

$$K_a = [AgAb]/[Ag][Ab]$$

(the square brackets indicate the concentration of each component at equilibrium).

One can determine the affinity constant, also known as the association constant, by measuring the concentration of free Ag required to fill half of the antigen-binding sites on the antibody. When half the sites are filled, [AgAb] = [Ab] and $K_a = 1/[Ag]$. Thus, the reciprocal of the antigen concentration that produces half the maximum binding is equal to the affinity constant of the antibody for the antigen. Common values range from as low as 5×10^4 to as high as 10^{11} liters/mole.

The **affinity** of an antibody for an antigenic determinant describes the strength of binding of a single copy of the antigenic determinant to a single antigen-binding site, and it is independent of the number of antigen-binding sites. When, however, a polyvalent antigen, carrying multiple copies of the same antigenic determinant, combines with a polyvalent IgM antibody (see Figure 25–23), the binding strength is greatly increased because all of the antigen–antibody bonds must be broken simultaneously before the antigen and antibody can dissociate. Even a bivalent IgG molecule can bind at least 100 times more strongly to a polyvalent antigen if both antigen-binding sites are engaged than if only one site is engaged. The total binding strength of a bivalent or polyvalent antibody with a polyvalent antigen is referred to as the **avidity** of the interaction.

If the affinity of the antigen-binding sites in an IgG and an IgM molecule is the same, the IgM molecule (with 10 binding sites) will have a much greater avidity for a polyvalent antigen than an IgG molecule (which has two binding sites). This difference in avidity, often 10^4-fold or more, is important because antibodies produced early in an immune response usually have much lower affinities than those produced later. Because of its high total avidity, IgM—the major Ig class produced early in primary immune responses—can function effectively even when each of its binding sites has only a low affinity.

So far, we have considered the general structure and function of antibodies. Next, we look at the details of their structure, as revealed by studies of their amino acid sequence and three-dimensional structure.

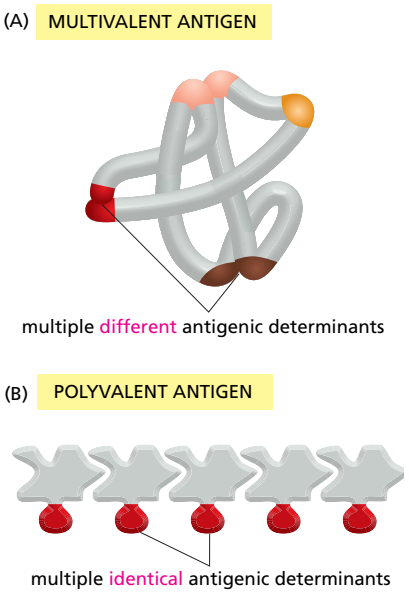


Figure 25–29 Molecules with multiple antigenic determinants. (A) A globular protein is shown with a number of *different* antigenic determinants. Different regions of a polypeptide chain usually come together in the folded structure to form each antigenic determinant on the surface of the protein, as shown for three of the four determinants. (B) A polymeric structure is shown with many *identical* antigenic determinants.

Antibody Light and Heavy Chains Consist of Constant and Variable Regions

Comparison of the amino acid sequences of different antibody molecules reveals a striking feature with important genetic implications. Both light and heavy chains have a variable sequence at their N-terminal ends but a constant sequence at their C-terminal ends. Consequently, when we compare the amino acid sequences of many different κ chains, the C-terminal halves are the same or show only minor differences, whereas the N-terminal halves all differ. Light chains have a **constant region** about 110 amino acids long and a **variable region**

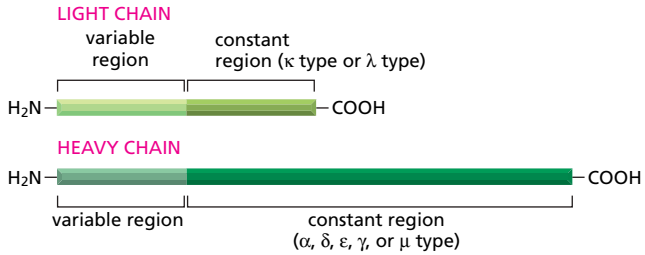


Figure 25–30 Constant and variable regions of immunoglobulin chains. The variable regions of the light and heavy chains form the antigen-binding sites, while the constant regions of the heavy chains determine the other biological properties of an antibody.

of the same size. The variable region of the heavy chains is also about 110 amino acids long, but the constant region is about three or four times longer (330 or 440 amino acids), depending on the class (**Figure 25–30**).

It is the N-terminal ends of the light and heavy chains that come together to form the antigen-binding site, and the variability of their amino acid sequences provides the structural basis for the diversity of antigen-binding sites. The greatest diversity occurs in three small **hypervariable regions** in the variable regions of both light and heavy chains; the remaining parts of the variable region, known as *framework regions*, are relatively constant.

Only about 5–10 amino acids in each hypervariable region form the actual antigen-binding site (**Figure 25–31**). As a result, the size of the antigenic determinant that an antibody recognizes is generally comparably small. It can consist of fewer than 10 amino acids on the surface of a globular protein, for example.

The Light and Heavy Chains Are Composed of Repeating Ig Domains

Both light and heavy chains are made up of repeating segments—each about 110 amino acids long and each containing one intrachain disulfide bond. Each repeating segment folds independently to form a compact functional unit called an **immunoglobulin (Ig) domain**. As shown in **Figure 25–32**, a light chain consists of one variable (V_L) and one constant (C_L) domain (equivalent to the variable and constant regions shown in the top half of Figure 25–30). V_L pairs with the variable (V_H) domain of the heavy chain to form the antigen-binding region. C_L pairs with the first constant domain of the heavy chain (C_H1), and the remaining constant domains of the heavy chains form the Fc region, which determines the other biological properties of the antibody. Most heavy chains have three constant domains (C_H1, C_H2, and C_H3), but those of IgM and IgE antibodies have four.

The similarity in their domains suggests that antibody chains arose during evolution by a series of gene duplications, beginning with a primordial gene coding for a single 110 amino acid domain of unknown function. Each domain of the constant region of a heavy chain is encoded by a separate coding sequence (exon), which supports this hypothesis (**Figure 25–33**).

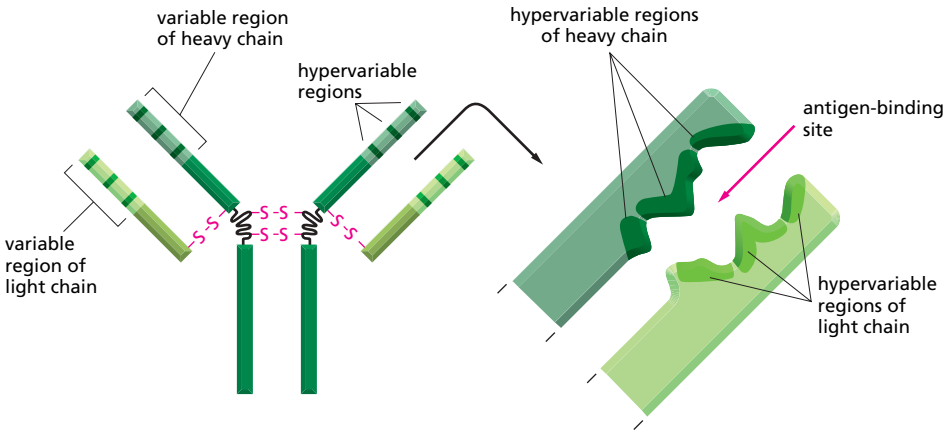


Figure 25–31 Antibody hypervariable regions. Highly schematized drawing of how the three hypervariable regions in each light and heavy chain together form the antigen-binding site of an antibody molecule.

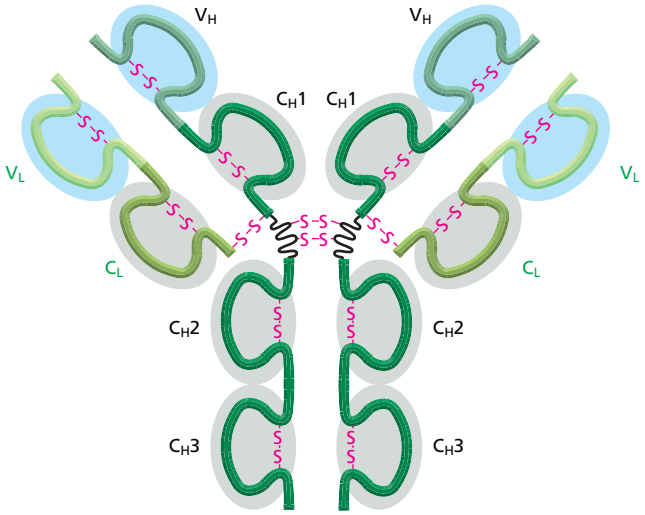


Figure 25–32 Immunoglobulin domains. The light and heavy chains in an antibody molecule are each folded into similar repeating domains. The variable domains (shaded in *blue*) of the light and heavy chains (V_L and V_H) make up the antigen-binding sites, while the constant domains of the heavy chains (mainly C_H2 and C_H3) determine the other biological properties of the molecule. The heavy chains of IgM and IgE antibodies do not have a hinge region and have an extra constant domain (C_H4). Hydrophobic interactions between domains on adjacent chains help hold the chains together in the antibody molecule: V_L binds to V_H, C_L binds to C_H1, and so on (see Figure 25–34). All antibodies are glycosylated on their C_H2 domains (not shown); the attached oligosaccharide chains vary from antibody to antibody and influence the biological properties of the antibody.

An Antigen-Binding Site Is Constructed from Hypervariable Loops

Many antibody fragments, as well as some intact antibody molecules, have been studied by x-ray crystallography. From these examples, we can understand the way in which billions of different antigen-binding sites are constructed on a common structural theme.

As illustrated in **Figure 25–34**, each Ig domain has a very similar three-dimensional structure consisting of a sandwich of two β sheets held together by a disulfide bond. As we discuss later, many other proteins on the surface of lymphocytes and other cells, many of which function as cell–cell adhesion molecules (discussed in Chapter 19), contain similar domains and hence are members of a very large *immunoglobulin (Ig) superfamily* of proteins.

The variable domains of antibody molecules are unique in that each has its particular set of three hypervariable regions, which are arranged in three *hypervariable loops* (see Figure 25–34). The hypervariable loops of both the light and heavy variable domains cluster together to form the antigen-binding site. Because the variable domain of an antibody molecule consists of a highly conserved rigid framework, with hypervariable loops attached at one end, changes in only the lengths and amino acid sequences of the hypervariable loops can generate an enormous diversity of antigen-binding sites. The overall three-dimensional structure necessary for antibody function remains constant.

X-ray analyses of crystals of antibody fragments bound to an antigenic determinant reveal exactly how the hypervariable loops of the light and heavy variable domains cooperate to form an antigen-binding surface in particular

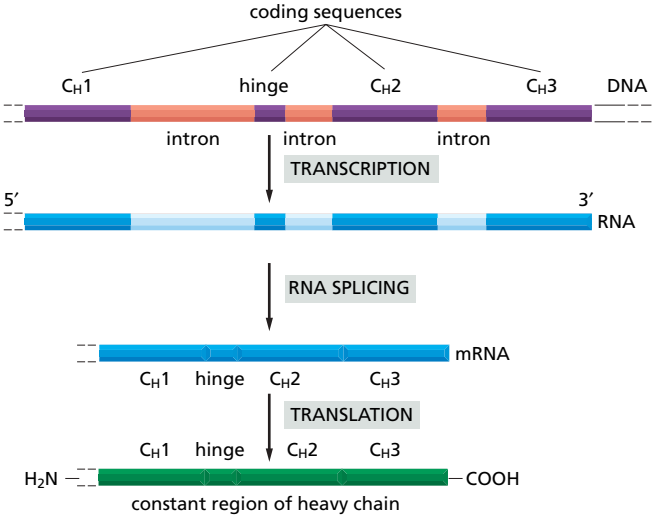


Figure 25–33 The organization of the DNA sequences that encode the constant region of an antibody heavy chain, such as that found in IgG. The coding sequences (exons) for each domain and for the hinge region are separated by noncoding sequences (introns). The intron sequences are removed by splicing the primary RNA transcripts to form mRNA. The presence of introns in the DNA is thought to have facilitated accidental duplications of DNA segments that gave rise to the antibody genes during evolution (discussed in Chapter 4). The DNA and RNA sequences that encode the variable region of the heavy chain are not shown.

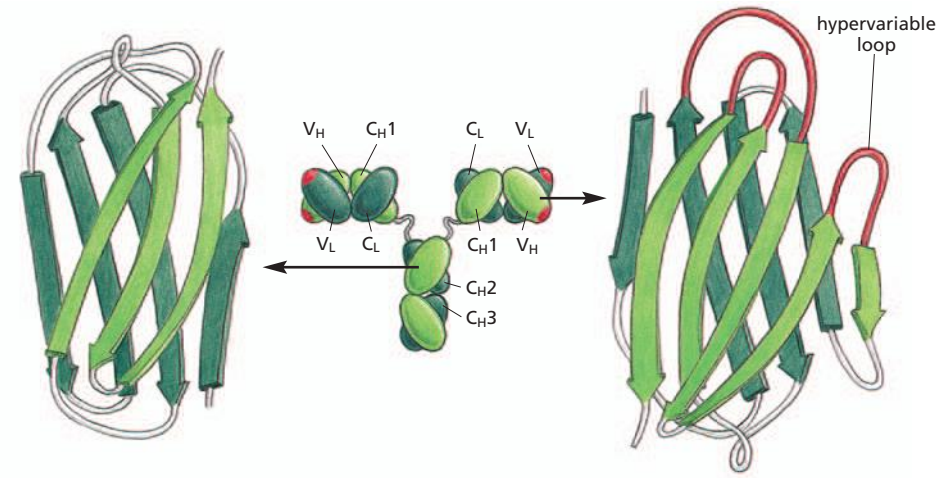


Figure 25–34 The folded structure of an IgG antibody molecule, based on x-ray crystallography studies. <GCCG> The structure of the whole protein is shown in the middle, while the structure of a constant domain is shown on the left and that of a variable domain on the right. Both domains consist of two β sheets, which are joined by a disulfide bond (not shown). Note that all the hypervariable regions *(red)* form loops at the far end of the variable domain, where they come together to form part of the antigen-binding site (see also Figure 3–41).

cases. The dimensions and shape of each different site vary depending on the conformations of the polypeptide chain in the hypervariable loops, which in turn are determined by the sequences of the amino acid side chains in the loops. The shapes of binding sites vary greatly—from pockets, to grooves, to undulating flatter surfaces, and even to protrusions—depending on the antibody (**Figure 25–35**). Smaller ligands tend to bind to deeper pockets, whereas larger ones tend to bind to flatter surfaces. In addition, the binding site can alter its shape after antigen binding to fit the ligand better.

Now that we have discussed the structure and functions of antibodies, we are ready to consider the crucial question that puzzled immunologists for many years—what are the genetic mechanisms that enable each of us to make many billions of different antibody molecules?

Summary

Antibodies defend vertebrates against infection by inactivating viruses and microbial toxins and by recruiting the complement system and various types of white blood cells to kill invading pathogens. A typical antibody molecule is Y-shaped, with two identical antigen-binding sites at the tips of the Y, plus binding sites for complement components and various cell-surface receptors on the tail of the Y.

Each B cell clone makes antibody molecules with a unique antigen-binding site. Initially, during B cell development in the bone marrow, the antibody molecules are inserted into the plasma membrane, where they serve as receptors for antigen. In peripheral lymphoid organs, antigen binding to these receptors, together with co-stimulatory signals provided by helper T cells, activates the B cells to proliferate and differentiate into either memory cells or antibody-secreting effector cells. The effector cells secrete large amounts of antibodies with the same unique antigen-binding site as the membrane-bound antibodies.

A typical antibody molecule is composed of four polypeptide chains, two identical heavy chains and two identical light chains. Parts of both the heavy and light chains usually combine to form the antigen-binding sites. There are five classes of antibodies (IgA, IgD, IgE, IgG, and IgM), each with a distinctive heavy chain (α, δ, ε, γ, and μ, respectively). The heavy chains also form the tail (Fc region) of the antibody, which determines what other proteins will bind to the antibody and therefore what biological properties the antibody class has. Either type of light chain (κ or λ) can be associated with any class of heavy chain; this choice has no effect on the properties of the antibody, except for its specificity for antigen.

Each light and heavy chain is composed of a number of Ig domains—β sheet structures constructed from about 110 amino acids. A light chain has one variable (V_L) and one constant (C_L) domain, while a heavy chain has one variable (V_H) and either three or four constant (C_H) domains. The amino acid sequence variation in the variable domains of both light and heavy chains is concentrated in several small hypervariable regions, which protrude as loops at one end of these domains to form the antigen-binding site.

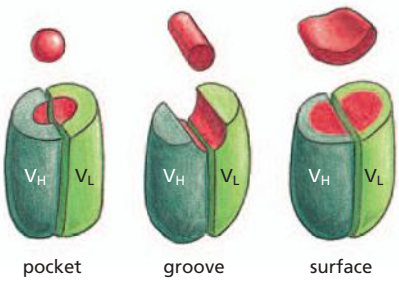


Figure 25–35 The variety of antigen-binding surfaces in antibodies. The hypervariable loops of different V_L and V_H domains can combine to form a large variety of binding surfaces. The antigenic determinants and the antigen-binding site of the antibodies are shown in *red*. Only one antigen-binding site is shown for each antibody.

THE GENERATION OF ANTIBODY DIVERSITY

Even in the absence of antigen stimulation, a human can probably make more than 10^{12} different antibody molecules—its preimmune, **primary antibody repertoire**. The primary repertoire consists of IgM and IgD antibodies and is apparently large enough to ensure that there will be an antigen-binding site to fit almost any potential antigenic determinant, albeit with low affinity. (The antigen-binding sites of many antibodies can cross-react with a variety of related but different antigenic determinants, making this primary antibody defense force even more formidable.)

After stimulation by antigen (and helper T cells), B cells can switch from making IgM and IgD to making other classes of antibodies—a process called *class switching*. In addition, the affinity of these antibodies for their antigen progressively increases over time—a process called *affinity maturation*. Thus, antigen stimulation generates a **secondary antibody repertoire**, with a greatly increased diversity of both Ig classes and antigen-binding sites.

Antibodies are proteins, and proteins are encoded by genes. Antibody diversity therefore poses a special genetic problem: how can an animal make more antibodies than there are genes in its genome? (The human genome, for example, contains only about 25,000 genes.) This problem is not quite as formidable as it might first appear. Recall that the variable regions of the light and heavy chains of antibodies usually combine to form the antigen-binding site. Thus, an animal with 1000 genes encoding light chains and 1000 genes encoding heavy chains could, in principle, combine their products in 1000×1000 different ways to make 10^6 different antigen-binding sites (although, in reality, not every light chain can combine with every heavy chain to make an antigen-binding site). Nonetheless, unique genetic mechanisms have evolved to enable adaptive immune systems to generate an almost unlimited number of different light and heavy chains in a remarkably economical way.

Not all vertebrates use the same genetic mechanisms to diversify antibodies, and there are even substantial differences in the mechanisms used by different mammals. We discuss the mechanisms used by mice and humans, in which antibody diversity is generated in two steps. First, before antigen stimulation, developing B cells join together separate *gene segments* in DNA in order to create the genes that encode the primary repertoire of low-affinity IgM and IgD antibodies. Second, after antigen stimulation, the assembled antibody-coding genes can undergo two further changes—mutations that can increase the affinity of the antigen-binding site and DNA rearrangements that switch the class of antibody made. Together, these changes produce the secondary repertoire of high-affinity IgG, IgA, and IgE antibodies.

We begin this section by discussing the mechanisms that B cells use to produce the primary antibody repertoire and then discuss the mechanisms that they use to produce the secondary repertoire.

Antibody Genes Are Assembled From Separate Gene Segments During B Cell Development

Mice and humans produce their primary antibody repertoire by joining separate antibody **gene segments** together during B cell development. Each type of antibody chain—κ light chains, λ light chains, and heavy chains—is encoded by a separate locus on a separate chromosome. Each locus contains a large number of gene segments encoding the V region of an antibody chain, and one or more gene segments encoding the C region. During the development of a B cell in the bone marrow (or fetal liver), a complete coding sequence for each of the two antibody chains to be synthesized is assembled by site-specific genetic recombination (discussed in Chapter 5). In addition to bringing together the separate gene segments of the antibody gene, these rearrangements also activate transcription from the gene promoter through changes in the relative positions of the enhancers and silencers acting on the gene. Thus, a complete antibody chain can be synthesized only after the DNA has been rearranged.

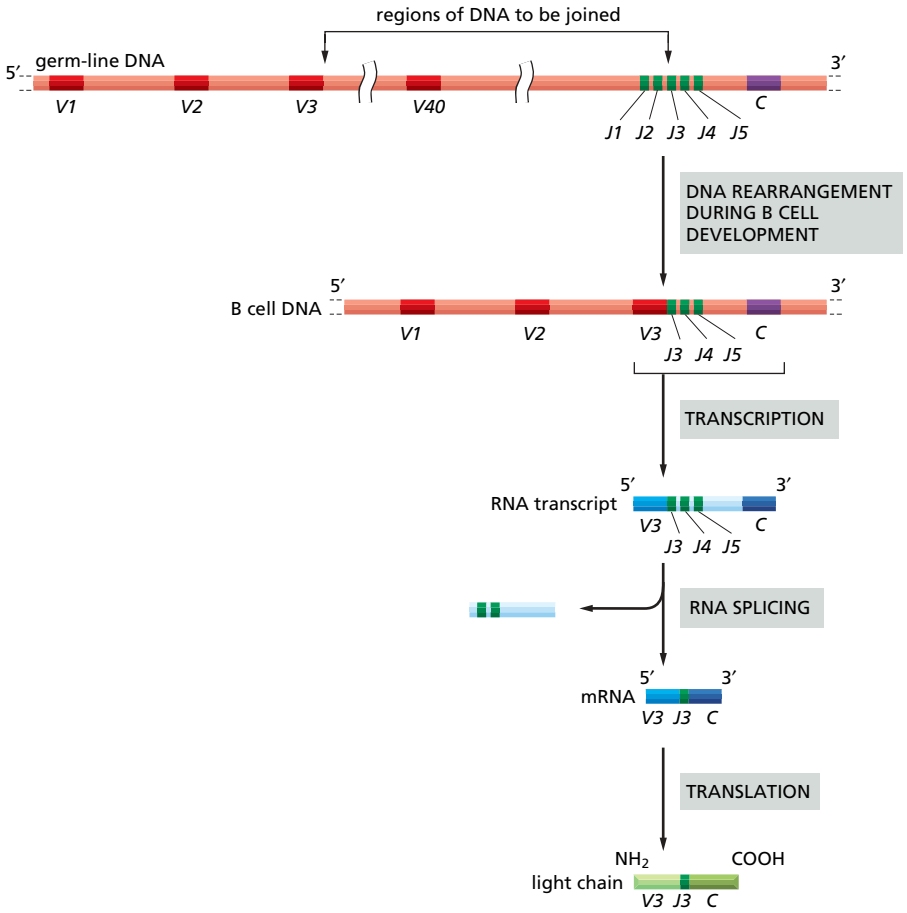


Figure 25–36 The *V–J* joining process involved in making a human κ light chain. In the "germ-line" DNA (where the antibody genes are not rearranged and are therefore not being expressed), the cluster of five *J* gene segments is separated from the *C*-region coding sequence by a short intron and from the 40 *V* gene segments by thousands of nucleotide pairs. During the development of a B cell, a randomly chosen *V* gene segment *(V3* in this case) is moved to lie precisely next to one of the *J* gene segments *(J3* in this case). The "extra" *J* gene segments *(J4* and *J5)* and the intron sequence are transcribed (along with the joined *V3* and *J3* gene segments and the C-region coding sequence) and then removed by RNA splicing to generate mRNA molecules with contiguous *V3, J3,* and *C* sequences, as shown. These mRNAs are then translated into κ light chains. A *J* gene segment encodes the C-terminal 15 or so amino acids of the V region, and a short sequence containing the *V–J* segment junction encodes the third hypervariable region of the light chain, which is the most variable part of the V region.

Each light-chain V region is encoded by a DNA sequence assembled from two gene segments—a long ***V* gene segment** and a short *joining,* or ***J* gene segment** (not to be confused with the protein *J chain* (see Figure 25–23), which is encoded elsewhere in the genome). **Figure 25–36** illustrates the sequence of events involved in the production of a human κ light-chain polypeptide from its separate gene segments. Each heavy-chain V region is similarly constructed by combining gene segments, but here an additional *diversity segment,* or ***D* gene segment**, is also required (**Figure 25–37**).

The large number of inherited *V, J,* and *D* gene segments available for encoding antibody chains contributes substantially to antibody diversity, and the combinatorial joining of these segments (called *combinatorial diversification*) greatly increases this contribution. Any of the 40 *V* segments in the human κ light-chain locus, for example, can be joined to any of the 5 *J* segments (see Figure 25–36), so that this locus can encode at least 200 (40 × 5) different κ-chain V regions. Similarly, any of the 40 *V* segments in the human heavy-chain locus can be joined to any of the 25 *D* segments and to any of the 6 *J* segments to encode at least 6000 (40 × 25 × 6) different heavy-chain V regions.

The combinatorial diversification resulting from the assembly of different combinations of inherited *V, J,* and *D* gene segments is an important mechanism for diversifying the antigen-binding sites of antibodies. By this mechanism alone, called *V(D)J recombination,* a human can produce 320 different V_L regions (200 κ and 120 λ) and 6000 different V_H regions. In principle, these could then be combined to make about 1.9×10^6 (320 × 6000) different antigen-binding sites. In addition, as we discuss next, the joining mechanism itself greatly increases this number of possibilities (probably more than 10^8-fold), making the primary antibody repertoire much larger than the total number of B cells (about 10^{12}) in a human.

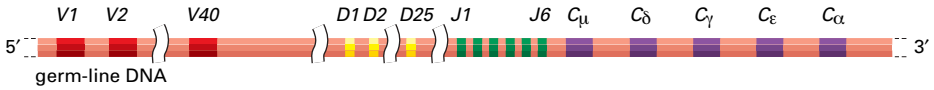


Figure 25–37 The human heavy-chain locus. There are 40 *V* segments, 25 *D* segments, 6 *J* segments, and an ordered cluster of *C*-region coding sequences, each cluster encoding a different class of heavy chain. The *D* segment (and part of the *J* segment) encodes amino acids in the third hypervariable region, which is the most variable part of the heavy-chain V region. The genetic mechanisms involved in producing a heavy chain are the same as those shown in Figure 25–36 for light chains, except that two DNA rearrangement steps are required instead of one. First a *D* segment joins to a *J* segment, and then a *V* segment joins to the rearranged *DJ* segment. The figure is not drawn to scale and omits detail: for example, the total length of the heavy chain locus is over 2 megabases.

Imprecise Joining of Gene Segments Greatly Increases the Diversity of V Regions

In the process of **V(D)J recombination**, site-specific recombination joins separate antibody gene segments together to form a functional V_L- or V_H-region coding sequence. Conserved *recombination signal sequences* flank each gene segment and serve as recognition sites for the joining process, ensuring that only appropriate gene segments recombine. Thus, for example, a light-chain *V* segment will always join to a *J* segment but not to another *V* segment. An enzyme complex called the *V(D)J recombinase* mediates joining. This complex contains two proteins that are specific to developing lymphocytes, as well as enzymes that help repair damaged DNA in all our cells.

Two closely linked genes called *Rag1* and *Rag2* (*Rag* = *r*ecombination *a*ctivating *g*enes) encode the lymphocyte-specific proteins of the V(D)J recombinase, RAG1 and RAG2. To mediate V(D)J joining, the two proteins come together to form a complex (called **RAG**), which functions as an endonuclease, introducing double-strand breaks precisely between the gene segments to be joined and their flanking recombination signal sequences. RAG then initiates the rejoining process by recruiting enzymes involved in DNA double-strand repair in all cells (**Figure 25–38**). Mice or humans deficient in either of the two *Rag* genes or in nonhomologous end joining are highly susceptible to infection because they are unable to carry out V(D)J recombination and consequently do not have functional B or T cells, a condition called *severe combined immunodeficiency (SCID)*. (As we discuss later, T cells use the same V(D)J recombinase to assemble the gene segments that encode their antigen-specific receptors.)

During the joining of antibody (and T cell receptor) gene segments, as in nonhomologous end-joining (see Figure 5–51A), a variable number of nucleotides are often lost from the ends of the recombining gene segments, and one or more randomly chosen nucleotides may also be inserted. This random loss and gain of nucleotides at joining sites is called **junctional diversification**, and it enormously increases the diversity of V-region coding sequences created by V(D)J recombination, specifically in the third hypervariable region. This

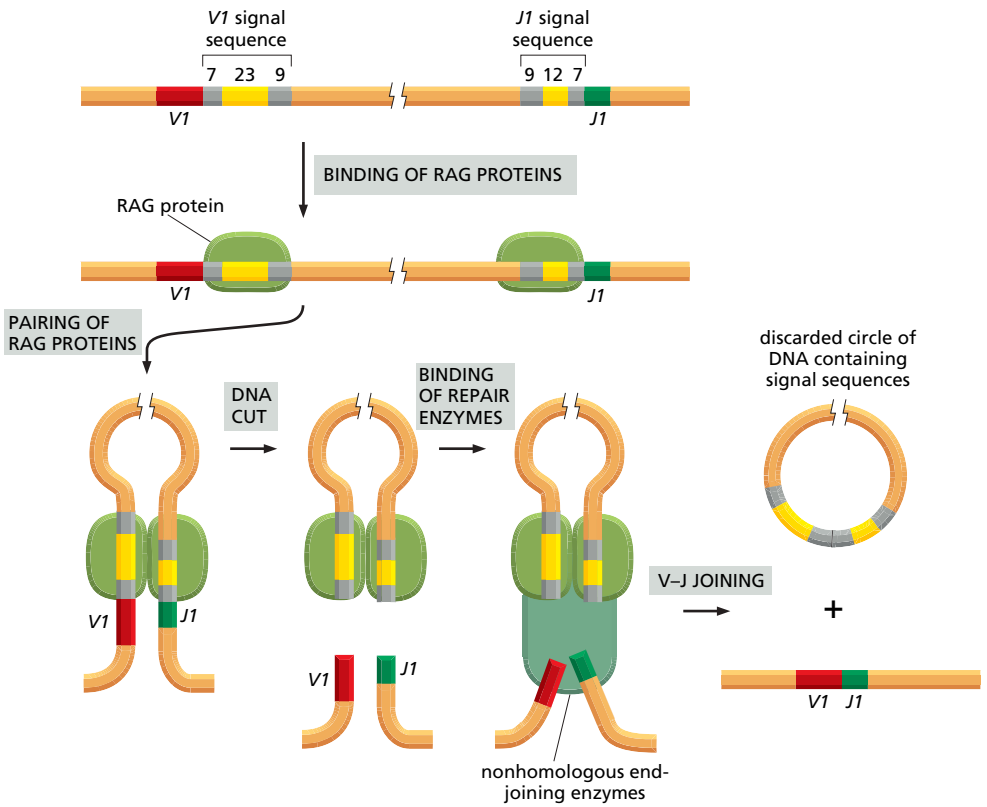


Figure 25–38 The role of recombination signal sequences in RAG-mediated gene segment joining. In the case shown, V1 is joined to J1 in a light-chain locus. Two types of DNA signal sequences are involved in V(D)J recombination, and recombination can only occur between different types: both have the same 7-base-pair (bp) sequence at one end and the same 9-bp sequence at the other end, but in one type the ends are separated by a 12-bp spacer and in the other they are separated by a 23-bp spacer, as shown. When one RAG protein binds to a 12-bp spacer and another to a 23-bp spacer, and the two RAG proteins bind to each other, the two different signal sequences are juxtaposed. The RAG complex then cuts the two signal sequences at their 7-bp ends, and DNA repair enzymes join the cut V1 and J1 segments together. The signal sequences are also joined together and discarded as a small circle of DNA that contains all the DNA originally located between V1 and J1.

The same process and signal sequences are used to join V, D, and J gene segments in a heavy-chain locus. The arrangement of signal sequences and the "12/23 rule" just described ensure that only appropriate gene segments recombine.

increased diversification comes at a price, however. In many cases, it will shift the reading frame to produce a nonfunctional gene. Because roughly two in every three rearrangements are "nonproductive" in this way, many developing B cells never make a functional antibody molecule and consequently die in the bone marrow.

B cells making functional antibody molecules that bind strongly to self antigens in the bone marrow would be dangerous. Such B cells maintain expression of the RAG proteins and can undergo a second round of V(D)J recombination in a light-chain locus (usually a κ locus), thereby changing the specificity of the cell-surface antibody they make—a process referred to as **receptor editing**. To provide a further layer of protection, clonal deletion eliminates those self-reactive B cells that fail to change their specificity (see Figure 25–13).

The Control of V(D)J Recombination Ensures That B Cells Are Monospecific

B cells are *monospecific*. That is, all the antibodies that any one B cell produces have identical antigen-binding sites. This property enables antibodies to cross-link antigens into large aggregates, thereby promoting antigen elimination (see Figure 25–19). It also means that an activated B cell secretes antibodies with the same specificity as that of its membrane-bound antibody receptor, guaranteeing the specificity of antibody responses (see Figure 25–17).

To achieve monospecificity, each B cell must make only one type of V_L region and one type of V_H region. Since B cells, like other somatic cells, are diploid, each cell has six loci encoding antibody chains: two heavy-chain loci (one from each parent) and four light-chain loci (one κ and one λ from each parent). If DNA rearrangements occurred independently in each heavy-chain locus and each light-chain locus, a single B cell could make up to eight different antibodies, each with a different antigen-binding site.

In fact, however, each B cell uses only two of the six antibody loci: one of the two heavy-chain loci and one of the four light-chain loci. Thus, each B cell must choose not only between its κ and λ light-chain loci, but also between its maternal and paternal light-chain and heavy-chain loci. This second choice is called **allelic exclusion**. Allelic exclusion also occurs in the expression of some genes that encode T cell receptors and genes that encode olfactory receptors in the nose (discussed in Chapter 15). However, for most proteins that are encoded by autosomal genes, both the maternal and paternal gene copies in a cell are expressed about equally.

Allelic exclusion and κ versus λ light-chain choice during B cell development depend on negative feedback regulation of the V(D)J recombination process. A functional rearrangement in one antibody locus suppresses rearrangements in all remaining loci that encode the same type of antibody chain (**Figure 25–39**). In B cell clones isolated from transgenic mice expressing a rearranged μ-chain gene, for example, the rearrangement of all of the endogenous heavy-chain genes is usually suppressed. Comparable results have been obtained for light chains. The suppression does not occur if the product of the rearranged gene fails to assemble into a receptor that inserts into the plasma membrane. It has therefore been proposed that either the receptor assembly process itself or extracellular signals that act on the receptor suppress further gene rearrangements.

Although no biological differences between the constant regions of κ and λ light chains have been discovered, there is an advantage in having two separate loci encoding light-chain variable regions. Having two separate loci increases the chance that a pre-B cell that has successfully assembled a V_H-region coding sequence will then successfully assemble a V_L-region coding sequence to become a B cell. This chance is further increased because, before a developing pre-B cell produces ordinary light chains, it makes surrogate light chains (see Figure 25–22), which assemble with μ heavy chains. The resulting receptors are displayed on the cell surface and allow the cell to proliferate, producing large numbers of progeny cells, some of which are likely to succeed in producing bona fide light chains.

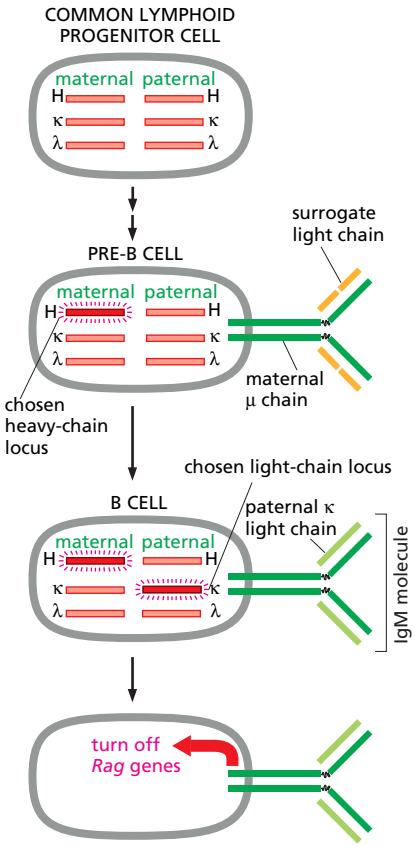


Figure 25–39 Selection of antibody loci during B cell development in the bone marrow. To produce antibodies with only one type of antigen-binding site, a developing B cell must use only one L-chain locus and one H-chain locus. Although the choice between maternal and paternal loci is thought to be random, the assembly of V-region coding sequences in a developing B cell proceeds in an orderly sequence, one segment at a time, usually beginning with the H-chain locus. In this locus, *D* segments first join to J_H segments on both parental chromosomes; then V_H to DJ_H joining occurs on one of these chromosomes (not shown). If this rearrangement produces a functional gene, the resulting production of complete μ chains (always the first H chains made) leads to their expression on the cell surface in association with surrogate light chains (see Figure 25–22). The cell now shuts down all further rearrangements of V_H-region-encoding gene segments and initiates V_L rearrangement. V_L rearrangement usually occurs first in a κ locus, and only if that fails does it occur at the other κ locus or at a λ locus. If, at any point, "in-phase" V_L-to-J_L joining leads to the production of light chains, these combine with preexisting μ chains to form IgM antibody molecules, which insert into the plasma membrane. The IgM cell-surface receptors are thought to enable the newly formed B cell to receive extracellular signals that shut down all further V(D)J recombination, by turning off the expression of the *Rag1* and *Rag2* genes.

If a developing B cell makes a receptor with high affinity for a self antigen, *Rag* gene expression is maintained and the cell undergoes another round of V(D)J recombination in a light-chain locus (called receptor editing—see Figure 25–13), thereby changing the specificity of its receptor (not shown). If a cell fails to assemble both a functional V_H-region and a functional V_L-region coding sequence, it is unable to make antibody molecules and dies by apoptosis (not shown).

The production of a functional B cell is a complex and highly selective process: in the end, all B cells that fail to produce intact antibody molecules die by apoptosis.

We turn now from the mechanisms responsible for generating the primary antibody repertoire before antigen stimulation to those responsible for generating the secondary antibody repertoire after antigen stimulation. We begin with the remarkable Darwinian-like mechanism responsible for increasing the affinity of the antigen-binding sites of antibodies for their specific antigen.

Antigen-Driven Somatic Hypermutation Fine-Tunes Antibody Responses

As mentioned earlier, with the passage of time after immunization, there is usually a progressive increase in the affinity of the antibodies produced against the immunizing antigen. This phenomenon, known as **affinity maturation**, is due to the accumulation of point mutations in both heavy-chain and light-chain V-region coding sequences. The mutations occur long after the coding regions have been assembled. After B cells have been stimulated by antigen and helper T cells in a peripheral lymphoid organ, some of the activated B cells proliferate rapidly in the lymphoid follicles (see Figure 25–16) and form structures called *germinal centers*. Here, the B cells mutate at the rate of about one mutation per V-region coding sequence per cell generation. Because this is about a million times greater than the spontaneous mutation rate in other genes and occurs in somatic cells rather than germ cells (discussed in Chapter 21), the process is called **somatic hypermutation**.

Very few of the altered antibodies generated by hypermutation will have an increased affinity for the antigen. Because the same antibody genes produce the antigen receptors on the B cell surface, the antigen will stimulate preferentially those few B cells that do make such antibodies with increased affinity for the antigen. Clones of these altered B cells will preferentially survive and proliferate, especially as the amount of antigen decreases to very low levels late in the response. Most other B cells in the germinal center will die by apoptosis. Thus, as a result of repeated cycles of somatic hypermutation, followed by antigen-driven proliferation of selected clones of effector and memory B cells, antibodies of increasingly higher affinity become abundant during an immune response, providing progressively better protection against the pathogen. (In

some mammals, including sheep and cows, a similar somatic hypermutation also plays a major part in diversifying the primary antibody repertoire before B cells encounter their antigen.)

A breakthrough in understanding the molecular mechanism of somatic hypermutation came with the identification of an enzyme that is required for the process. It is called **activation-induced deaminase** (**AID**) because it is expressed specifically in activated B cells and deaminates cytosine (C) to uracil (U) in transcribed V-region coding DNA. The deamination produces U:G mismatches in the DNA double helix, and the repair of these mismatches produces various types of mutations, depending on the repair pathway used (**Figure 25–40**). Somatic hypermutation affects only actively transcribed V-region coding sequences, possibly because the AID enzyme is specifically loaded onto RNA transcripts (discussed in Chapter 7). AID is also required when activated B cells switch from IgM production to the production of other classes of antibody, as we now discuss.

B Cells Can Switch the Class of Antibody They Make

As discussed earlier, all B cells begin their antibody-synthesizing lives by making IgM molecules and inserting them into the plasma membrane as receptors for antigen. After the B cells leave the bone marrow, but before they interact with antigen, they begin making both IgM and IgD molecules as membrane-bound antigen receptors, both with the same antigen-binding sites (see Figure 25–22). Stimulation by antigen and helper T cells activates many of these cells to become IgM-secreting effector cells, so that IgM antibodies dominate the primary antibody response. Later in the immune response, however, when activated B cells are undergoing somatic hypermutation, the combination of antigen and helper-T-cell-derived cytokines stimulates many of the B cells to switch from making membrane-bound IgM and IgD to making IgG, IgA, or IgE antibodies—a process called **class switching**. Some of these cells become memory cells that express the corresponding class of antibody molecules on their surface, while others become effector cells that secrete the antibodies. The IgG, IgA, and IgE molecules are collectively referred to as *secondary classes* of antibodies, because they are produced only after antigen stimulation, dominate secondary antibody responses, and make up the secondary antibody repertoire. As we saw earlier, each different class of antibody is specialized to attack pathogens in different ways and in different sites.

The constant region of an antibody heavy chain determines the class of the antibody. Thus, the ability of B cells to switch the class of antibody they make without changing the antigen-binding site implies that the same assembled V_H-region coding sequence (which specifies the antigen-binding part of the heavy chain) can sequentially associate with different C_H-coding sequences. This has important functional implications. It means that, in an individual animal, a

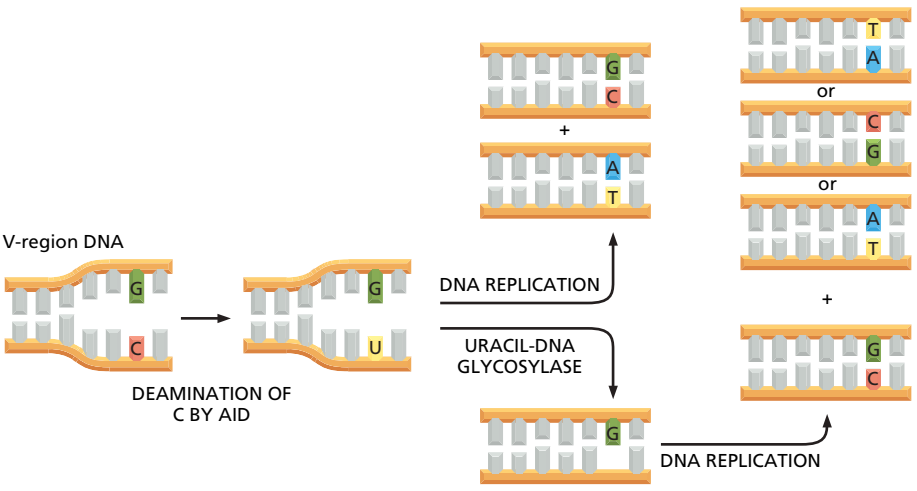


Figure 25–40 Some ways in which AID can cause mutations during somatic hypermutation. AID deaminates some cytosines to uracil in transcribed V-region coding DNA, causing U:G mismatches, which lead to mutations in various ways. Some mutations occur when DNA containing unprocessed U:G mismatches are replicated (see Figure 5–49A). Others occur when the uracil is removed by uracil-DNA glycosylase before the DNA is replicated, as this generates a position on one of the DNA template strands that lacks a base for DNA polymerase to copy. Still others (not shown) occur when the area around the U:G mismatch is excised by the mismatch repair system (discussed in Chapter 5), producing a gap that can be repaired by error-prone DNA polymerases, thereby generating mutations at A:T as well as C:G pairs.

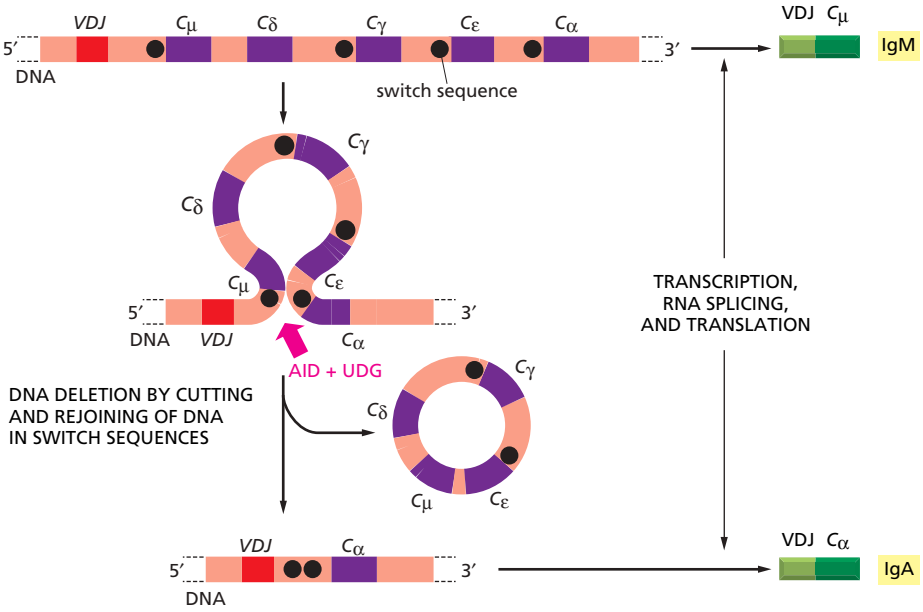


Figure 25–41 An example of the DNA rearrangement that occurs in class switch recombination. A B cell making an IgM antibody from an assembled *VDJ* DNA sequence is stimulated to switch to making an IgA antibody. In the process, it deletes the DNA between the *VDJ* sequence and the C_α-coding sequence. Specific DNA sequences *(switch sequences)* located upstream of each C_H-coding sequence (except C_δ) recombine with each other, with the deletion of the intervening DNA. As discussed in the text, the recombination process depends on AID and uracil-DNA glycosylase (UDG), the same enzymes that are involved in somatic hypermutation (see Figure 25–40).

particular antigen-binding site that has been selected by environmental antigens can be distributed among the various classes of antibodies, thereby acquiring the different biological properties of each class.

When a B cell switches from making IgM and IgD to one of the secondary classes of antibody, an irreversible change at the DNA level occurs—a process called **class-switch recombination**. It entails the deletion of all the C_H-coding sequences between the assembled VDJ-coding sequence and the particular C_H-coding sequence that the cell is destined to express. Class-switch recombination differs from V(D)J recombination in several ways. (1) It happens after antigen stimulation, mainly in germinal centers, and depends on helper T cells. (2) It uses different recombination signal sequences, called *switch sequences*, which are composed of short motifs tandemly repeated over several kilobases. (3) It involves cutting and joining the switch sequences, which are noncoding sequences, and so the coding sequence is unaffected (**Figure 25–41**). (4) Most importantly, the molecular mechanism is different. It depends on AID, which is also involved in somatic hypermutation, rather than on RAG, which is responsible for V(D)J recombination.

The cytokines that activate class switching induce the production of transcription factors that activate transcription from the relevant switch sequences, allowing AID to bind to these sequences. Once bound, AID initiates switch recombination by deaminating some cytosines to uracil in the vicinity of these switch sequences. Excision of these uracils by uracil-DNA glycosylase (see Figure 25–40) is thought to lead somehow to double-strand breaks in the participating switch regions, which are then joined by a form of nonhomologous end-joining (discussed in Chapter 5).

Thus, whereas the primary antibody repertoire in mice and humans is generated by V(D)J joining mediated by RAG, the secondary antibody repertoire is generated by somatic hypermutation and class-switch recombination, both of which are mediated by AID. **Figure 25–42** summarizes the main mechanisms involved in diversifying antibodies that we have discussed in this chapter.

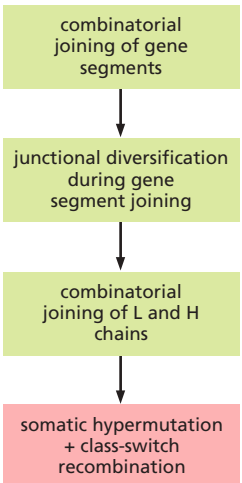


Figure 25–42 The main mechanisms of antibody diversification in mice and humans. Those shaded in *green* occur during B cell development in the bone marrow (or fetal liver), whereas the mechanisms shaded in *red* occur when B cells are stimulated by foreign antigen and helper T cells in peripheral lymphoid organs, either late in a primary response or in a secondary response.

Summary

Antibodies are encoded by three loci on separate chromosomes, each of which produces a different polypeptide chain. One encodes κ light chains, one encodes λ light chains, and one encodes heavy chains. Each antibody locus contains separate gene segments that code for different parts of the variable region of the particular antibody chain. Each light-chain locus contains one or more constant- (C-) region coding sequences and sets of variable (V) and joining (J) gene segments. The heavy-chain locus contains sets of C-region coding sequences and sets of V, diversity (D), and J gene segments.

During B cell development in the bone marrow (or fetal liver), before antigen stimulation, separate gene segments are brought together by site-specific recombination that depends on the RAG complex. A V_L gene segment recombines with a J_L gene segment to produce a DNA sequence coding for the V region of a light chain, and a V_H gene segment recombines with a D and a J_H gene segment to produce a DNA sequence coding for the V region of a heavy chain. Each of the assembled V-region coding sequences is then co-transcribed with the appropriate C-region sequence to produce an RNA molecule that codes for the complete polypeptide chain. Once a B cell makes functional heavy and light chains that form antigen-binding sites (that do not bind a self antigen with high affinity), it turns off the V(D)J recombination process, thereby ensuring that the cell makes only one species of antigen-binding site.

By randomly combining inherited gene segments that code for V_L and V_H regions during B cell development, humans can make hundreds of different light chains and thousands of different heavy chains. Because the antigen-binding site is formed where the hypervariable loops of the V_L and V_H come together in the final antibody, the heavy and light chains can potentially pair to form antibodies with millions of different antigen-binding sites. The loss and gain of nucleotides at the site of gene-segment joining increases this number enormously. These antibodies, made by RAG-dependent V(D)J recombination before antigen stimulation, are low-affinity IgM and IgD antibodies, and they constitute the primary antibody repertoire.

Antibodies are further diversified following antigen stimulation in peripheral lymphoid organs by the AID- and helper-T-cell-dependent processes of somatic hypermutation and class-switch recombination, which produce the high-affinity IgG, IgA, and IgE antibodies that constitute the secondary antibody repertoire. Class switching allows the same antigen-binding site to be incorporated into antibodies that have different biological properties.

T CELLS AND MHC PROTEINS

Like antibody responses, T-cell-mediated immune responses are exquisitely antigen-specific, and they are at least as important as antibodies in defending vertebrates against infection. Indeed, most adaptive immune responses, including most antibody responses, require helper T cells for their initiation. Most importantly, unlike B cells, T cells can help eliminate pathogens that would otherwise be invisible inside host cells. Much of the rest of this chapter is concerned with how T cells accomplish this feat.

T cell responses differ from B cell responses in at least two crucial ways. First, T cells are activated by foreign antigen to proliferate and differentiate into effector cells only when the antigen is displayed on the surface of *antigen-presenting cells*, usually dendritic cells in peripheral lymphoid organs. T cells require antigen-presenting cells for activation because the form of antigen they recognize is different from that recognized by B cells. Whereas B cells recognize intact protein antigens, for example, T cells recognize fragments of protein antigens that have been partly degraded inside the antigen-presenting cell. Special proteins, called *MHC proteins* (introduced in Chapter 24), bind to the peptide fragments and carry them to the surface of the antigen-presenting cell, where T cells can recognize them.

The second difference is that, once activated, effector T cells act only at short range, either within a secondary lymphoid organ or after they have migrated into a site of infection. Effector B cells, by contrast, secrete antibodies that can act far away. Effector T cells interact directly with another host cell in the body, which they either kill (as in the case of an infected host cell, for example) or signal in some way (as in the case of a B cell or macrophage, for example). We shall refer to such host cells as *target cells*. However, because these traget cells must display an antigen bound to an MHC protein on their surface for a T cell to recognize them, they are also antigen-presenting cells.

There are three main classes of T cells—cytotoxic T cells, helper T cells, and regulatory (suppressor) T cells. Effector *cytotoxic T cells* directly kill cells that are infected with a virus or some other intracellular pathogen. Effector *helper T cells*

help stimulate the responses of other cells—mainly macrophages, dendritic cells, B cells, and cytotoxic T cells. Effector *regulatory T cells* suppress the activity of other cells, especially of self-reactive effector T cells.

In this section, we describe these three main classes of T cells and their respective functions. We discuss how they recognize foreign antigens on the surface of antigen-presenting cells and target cells and consider the crucial part played by MHC proteins in the recognition process. Finally, we describe how T cells are selected during their development in the thymus to ensure that only cells with potentially useful receptors survive and mature. We begin by considering the nature of the cell-surface receptors that T cells use to recognize antigen.

T Cell Receptors (TCRs) Are Antibodylike Heterodimers

Because T cell responses depend on direct contact with an antigen-presenting cell or a target cell, **T cell receptors** (**TCRs**), unlike antibodies made by B cells, exist only in membrane-bound form and are not secreted. For this reason, TCRs were difficult to isolate, and it was not until the 1980s that researchers identified their molecular structure. TCRs resemble antibodies. They are composed of two disulfide-linked polypeptide chains, each of which contains two Ig-like domains, one variable and one constant (**Figure 25–43**A). Moreover, the three-dimensional structure of the extracellular part of a TCR has been determined by x-ray diffraction, and it looks very much like one arm of a Y-shaped antibody molecule (Figure 25–43B).

On most T cells, the TCRs have one α chain and one β chain. The genetic loci that encode the α and β chains are located on different chromosomes. Like an antibody heavy-chain locus, the TCR loci contain separate *V, D,* and *J* gene segments (or just *V* and *J* gene segments in the case of the α chain locus), which are brought together by site-specific recombination during T cell development in the thymus. With one exception, T cells use the same mechanisms to generate TCR diversity as B cells use to generate antibody diversity. Indeed, they use the same V(D)J recombinase, including the RAG proteins discussed earlier. The mechanism that does not operate in TCR diversification is antigen-driven somatic hypermutation. Thus, the affinity of the receptors tends to be low (K_a ~ 10^5–10^7 liters/mole), although T cells with the highest affinities are preferentially selected by antigen to persist as memory cells. T cells can partly compensate for their low affinity by increased avidity, which results when multiple TCRs bind simultaneously to multiple membrane-bound ligands (the

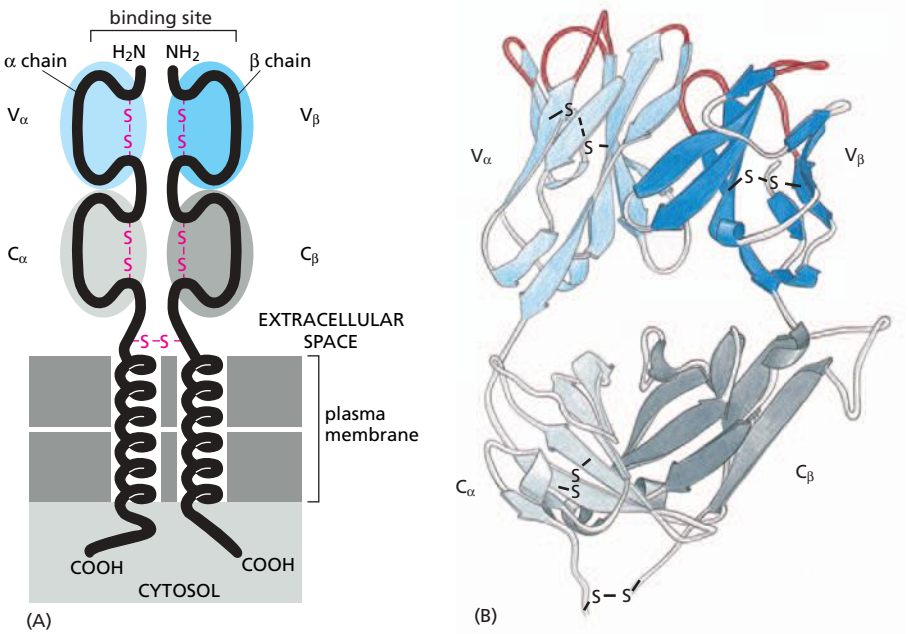


Figure 25–43 A T cell receptor (TCR) heterodimer. (A) Schematic drawing showing that the receptor is composed of an α and a β polypeptide chain. Each chain is about 280 amino acids long and has a large extracellular part that is folded into two Ig-like domains—one variable (V) and one constant (C). The antigen-binding site is formed by a V_α and a V_β domain (shaded in *blue*). Unlike antibodies, which have two binding sites for antigen, TCRs have only one. The αβ heterodimer is noncovalently associated with a large set of invariant membrane-bound proteins (not shown), which help activate the T cell when the TCRs bind to antigen. A typical T cell has about 30,000 such receptor complexes on its surface. (B) The three-dimensional structure of the extracellular part of a TCR. The antigen-binding site is formed by the hypervariable loops of both the V_α and V_β domains *(red),* and it is similar in its overall dimensions and geometry to the antigen-binding site of an antibody molecule. (B, based on K.C. Garcia et al., *Science* 274:209–219, 1996. With permission from AAAS.)

peptide–MHC complexes that we describe later). Moreover, various co-receptors and cell–cell adhesion proteins greatly strengthen the binding of a T cell to an antigen-presenting cell or a target cell.

A minority of T cells, instead of making α and β chains, make a different but related type of receptor heterodimer, composed of γ chains and δ chains. Although these cells normally make up 5–10% of the T cells in human blood, they can be the dominant T cell population in epithelia (in the skin and gut, for example). The functions of these cells are less well understood than those of T cells expressing α/β TCRs, and we will not discuss them further.

As with antigen receptors on B cells, the TCRs are tightly associated in the plasma membrane with a number of invariant membrane-bound proteins that are involved in passing the signal from an antigen-activated receptor to the cell interior (see Figure 25–66). We will discuss these proteins in more detail later. First, we must consider the special ways in which T cells recognize foreign antigen on the surface of an antigen-presenting cell.

Antigen Presentation by Dendritic Cells Can Either Activate or Tolerize T Cells

Naïve cytotoxic or helper T cells must be activated to proliferate and differentiate into effector cells before they can kill or help their target cells, respectively. This activation occurs in peripheral lymphoid organs on the surface of activated **dendritic cells** (**Figure 25–44**) that display foreign antigen complexed with MHC proteins on their surface, along with co-stimulatory proteins. By contrast, memory T cells can be activated by other types of antigen-presenting cells, including macrophages and B cells, as well as by dendritic cells.

Various kinds of dendritic cells interact with T cells, but they all have a single known function, which is to present antigens that either activate or suppress the T cells. Dendritic cells are located in tissues throughout the body, including the central and peripheral lymphoid organs. Wherever they encounter invading microbes, they endocytose the pathogens or their products. If the encounter is not in a lymphoid organ, the dendritic cells carry the foreign antigens via the lymph to local lymph nodes or gut-associated lymphoid organs. The encounter with a pathogen activates pattern recognition receptors of the dendritic cell, which is thereby induced to mature from an antigen-capturing cell to an activated antigen-presenting cell that can activate T cells (see Figure 25–5). Dendritic cells have to be activated in order to activate naïve T cells, and they can also be activated by tissue injury or by effector helper T cells. Tissue injury is thought to activate dendritic cells by the release of heat shock proteins and uric acid crystals when cells die by necrosis rather than by apoptosis (discussed in Chapter 18).

Activated dendritic cells display three types of protein molecules on their surface that have a role in activating a T cell to become an effector cell or a memory cell (**Figure 25–45**): (1) *MHC proteins*, which present foreign antigen to the TCR, (2) *co-stimulatory proteins,* which bind to complementary receptors on the T cell surface, and (3) *cell–cell adhesion molecules,* which enable a T cell to bind to the antigen-presenting cell for long enough to become activated, which is usually hours. In addition, activated dendritic cells secrete a variety of cytokines that can influence the type of effector helper T cell that develops (discussed later), as well as where the T cell migrates after it has been stimulated. T cells activated by dendritic cells isolated from gut-associated Peyer's patches (see Figure 25–3), for example, but not by those isolated from lymph nodes, migrate to the small intestine where their antigens are likely to be located.

Nonactivated dendritic cells also have important roles. They help induce self-reactive T cells to become tolerant, both in the thymus and in other organs; such dendritic cells present self antigens in the absence of the co-stimulatory molecules required to activate naïve T cells. They induce tolerance in at least two ways: they can stimulate abortive responses in the T cell that lead to either inactivation or apoptosis, and they can activate regulatory T cells to suppress the activity of another T cell.

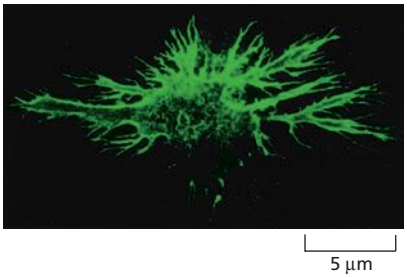


Figure 25–44 Immunofluorescence micrograph of a dendritic cell in culture. These antigen-presenting cells derive their name from their long processes, or "dendrites." The cell has been labeled with a monoclonal antibody that recognizes a surface antigen on these cells. (Courtesy of David Katz.)

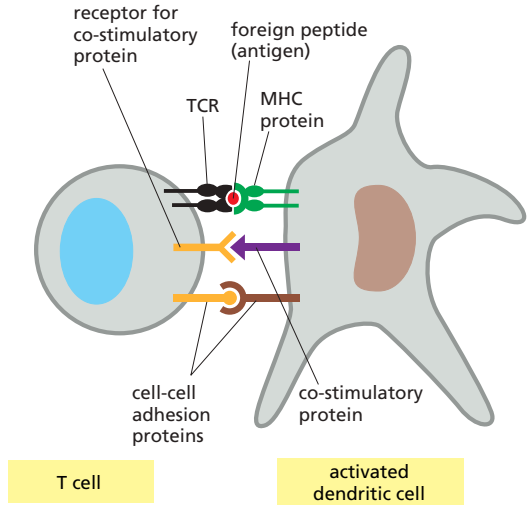


Figure 25–45 Three types of proteins on the surface of an activated dendritic cell involved in activating a T cell. The invariant polypeptide chains that are always stably associated with the T cell receptor (TCR) are not shown.

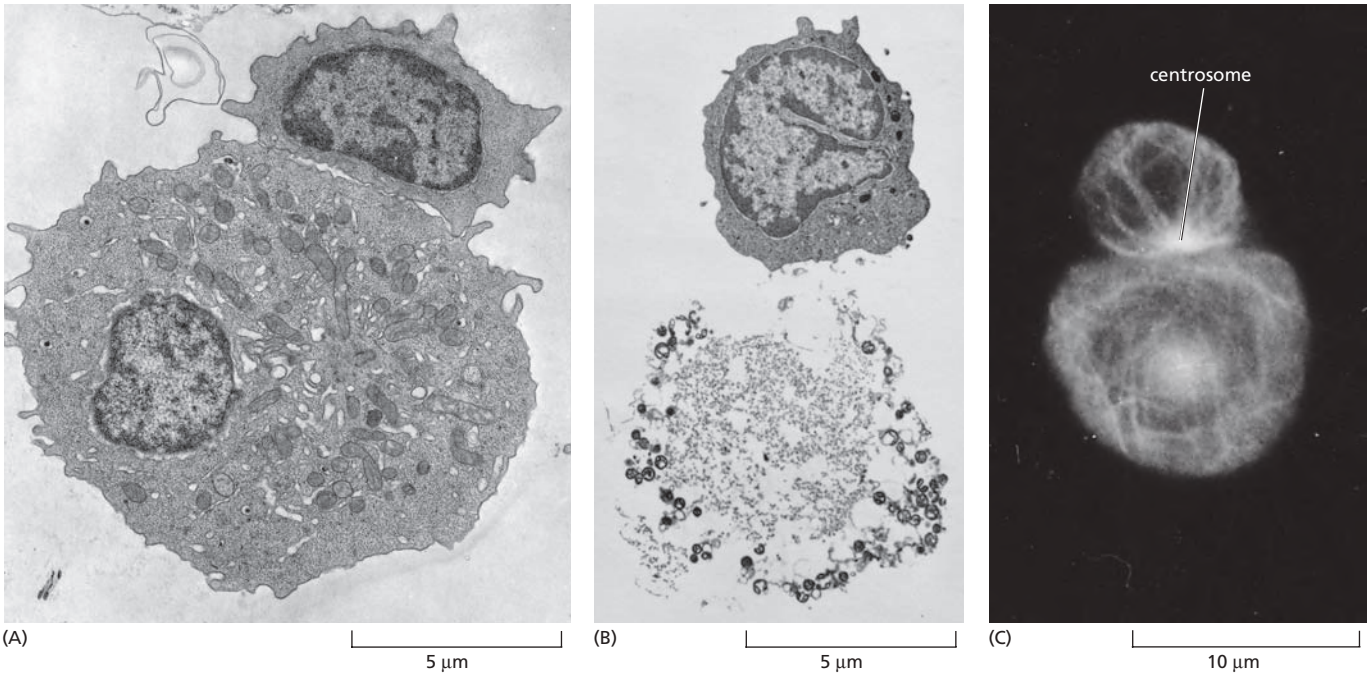
Before discussing the role of MHC proteins in presenting antigen to T cells, we consider the functions of the three major classes of T cells.

Effector Cytotoxic T Cells Induce Infected Target Cells to Kill Themselves

Cytotoxic T cells protect vertebrates against intracellular pathogens such as viruses and some bacteria and parasites that multiply in the host-cell cytoplasm, where they are sheltered from antibody-mediated attack. Cytotoxic T cells do this by killing the infected cell before the microbes can proliferate and escape from the infected cell to infect neighboring cells. As we discuss later, the intracellular microbes can be recognized by T cells because vertebrate cells have mechanisms for displaying fragments of their intracellular proteins on the cell surface, where they are bound to MHC proteins.

Once a cytotoxic T cell has been activated by an infected antigen-presenting cell to become an effector cell, it can kill any target cell harboring the same pathogen. Using its TCR, the effector cytotoxic T cell first recognizes a microbial antigen bound to an MHC protein on the surface of an infected target cell. This causes the T cell to reorganize its cytoskeleton and focus its killing apparatus on the target (**Figure 25–46**). Focus is achieved when the TCRs actively aggregate—

Figure 25–46 Effector cytotoxic T cells killing target cells in culture. <GTCA> (A) Electron micrograph showing an effector cytotoxic T cell binding to a target cell. The cytotoxic T cells were obtained from mice immunized with the target cells, which are foreign tumor cells. (B) Electron micrograph showing a cytotoxic T cell and a tumor cell that the T cell has killed. In an animal, as opposed to in a culture dish, the killed target cell would be phagocytosed by neighboring cells long before it disintegrated in the way that it has here. (C) Immuno-fluorescence micrograph of a T cell and tumor cell after staining with anti-tubulin antibodies. Note that the centrosome in the T cell is located at the point of cell–cell contact with the target cell—an immunological synapse. The secretory granules (not visible) in the T cell are initially transported along microtubules to the centrosome, which then moves to the synapse, delivering the granules to where they can release their contents. See also Figure 16–103. (A and B, from D. Zagury et al., *Eur. J. Immunol.* 5:818–822, 1975. With permission from John Wiley & Sons, Inc. C, reproduced from B. Geiger, D. Rosen and G. Berke, *J. Cell Biol.* 95:137–143, 1982. With permission from The Rockefeller University Press.)



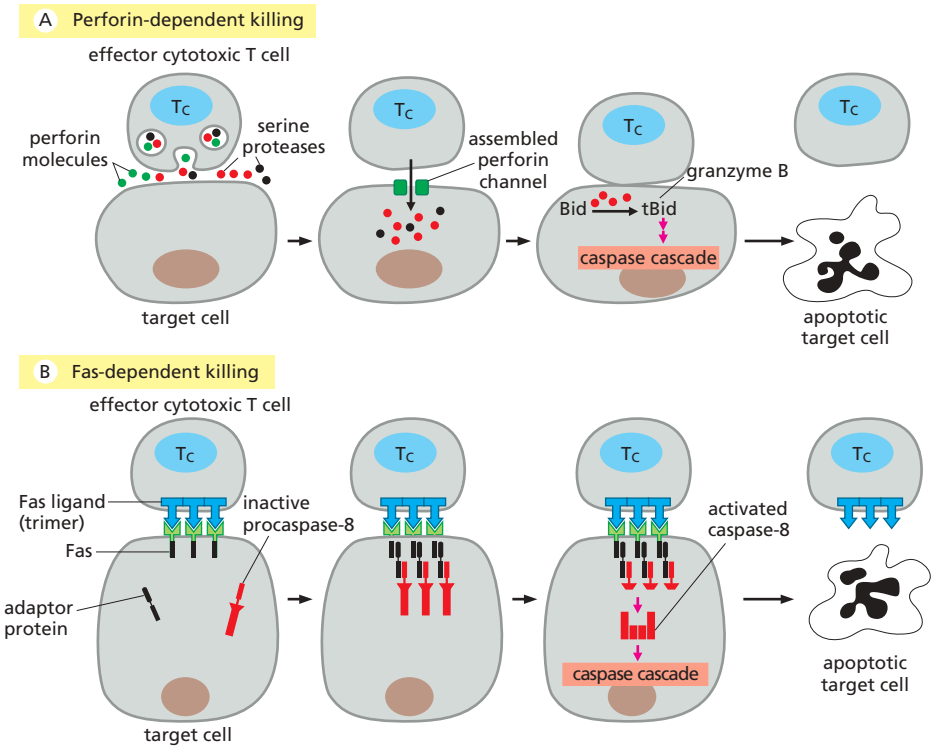


Figure 25–47 Two strategies by which effector cytotoxic T cells kill their target cells. In both cases, the T cell has to contact the target cell to kill it, and a single cytotoxic T cell can kill multiple target cells in sequence. (A) The cytotoxic T cell (T_C) releases perforin and proteolytic enzymes onto the surface of an infected target cell by localized exocytosis. The high concentration of Ca^{2+} in the extracellular fluid causes the perforin to assemble into transmembrane channels in the target cell plasma membrane. The channels are thought to allow the proteolytic enzymes to enter the target cell cytosol. One of the enzymes, granzyme B, cleaves the Bid protein to produce the truncated form tBid, which releases cytochrome *c* from mitochondria to initiate a caspase cascade leading to apoptosis. (B) The homotrimeric Fas ligand on the surface of the cytotoxic T cell binds to and activates the Fas protein on the surface of a target cell. The cytosolic tail of Fas contains a *death domain,* which, when activated, binds to an adaptor protein, which in turn recruits a specific procaspase *(procaspase-8).* Clustered procaspase-8 molecules become activated and initiate a proteolytic caspase cascade leading to apoptosis (see Figure 18–6).

along with various co-receptors, adhesion molecules, and signaling proteins—at the T cell/target cell interface, forming an *immunological synapse.* A similar synapse forms when an effector helper T cell interacts with its target cell. In this way, effector T cells avoid delivering their signals to neighboring cells.

Once bound to its target cell, an effector cytotoxic T cell can employ one of two strategies to kill the target, both of which operate by inducing the target cell to kill itself by undergoing apoptosis (discussed in Chapter 18). In killing an infected target cell, the cytotoxic T cell usually releases a pore-forming protein called **perforin**, which is homologous to the complement component C9 (see Figure 24–49). The cytotoxic T cell stores perforin in secretory vesicles and releases it by local exocytosis at the point of contact with the target cell. Perforin then polymerizes in the target cell plasma membrane to form transmembrane channels. The secretory vesicles also contain serine proteases, which are thought to enter the target cell cytosol through the perforin channels. One of the proteases, called *granzyme B,* activates a pro-apoptotic Bcl2 protein called *Bid* by producing a truncated form of the protein called *tBid;* tBid then releases cytochrome *c* from mitochondria, triggering a proteolytic caspase cascade that kills the cell by apoptosis (discussed in Chapter 18) (**Figure 25–47**A). Mice with an inactivated perforin gene cannot generate pathogen-specific cytotoxic T cells, and they show an increased susceptibility to certain viral and intracellular bacterial infections.

In the second killing strategy, the cytotoxic T cell activates a death-inducing caspase cascade in the target cell less directly. A homotrimeric protein on the cytotoxic T cell surface called **Fas ligand** binds to transmembrane receptor proteins on the target cell called **Fas**. The binding alters the Fas proteins so that their clustered cytosolic tails recruit procaspase-8 into a complex via an adaptor protein. The recruited procaspase-8 molecules thereby become activated and initiate a caspase cascade that leads to apoptosis (Figure 25–47B).

Effector Helper T Cells Help Activate Other Cells of the Innate and Adaptive Immune Systems

In contrast to cytotoxic T cells, **helper T cells** are crucial for defense against both extracellular and intracellular pathogens. They help stimulate B cells to make

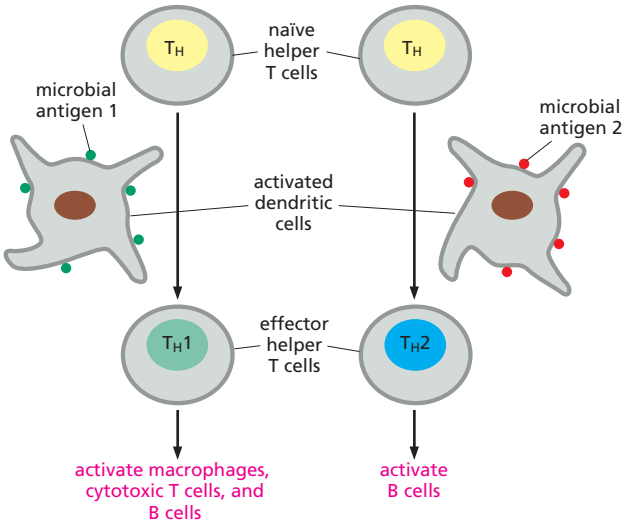


Figure 25–48 Differentiation of naïve helper T cells into either T_H1 or T_H2 effector helper cells in a peripheral lymphoid organ. The nature of the activated dendritic cell and the characteristics of the pathogen that activated it mainly determine which type of effector helper cell develops.

antibodies that help inactivate or eliminate extracellular pathogens and their toxic products. They also activate macrophages to destroy any intracellular pathogens multiplying within the macrophage's phagosomes, and they help activate cytotoxic T cells to kill infected target cells. They can also stimulate a dendritic cell to maintain it in an activated state.

Once an antigen-presenting cell activates a helper T cell to become an effector cell, the helper cell can then help activate other cells. It does this both by secreting a variety of co-stimulatory cytokines and by displaying co-stimulatory proteins on its surface. When activated by its binding to an antigen on a dendritic cell, a naïve helper T cell usually differentiates into either of two distinct types of effector helper cell, called T_H1 and T_H2. *T_H1 cells* are mainly involved in immunity to intracellular microbes and help activate macrophages, cytotoxic T cells, and B cells. *T_H2 cells* are mainly involved in immunity to extracellular pathogens, especially multicellular parasites, and they help activate B cells to make antibodies against the pathogen (**Figure 25–48**). As we discuss later, the nature of the invading pathogen and the types of innate immune responses it elicits largely determine which type of helper T cell develops. This, in turn, determines the nature of the adaptive immune responses mobilized to fight the invaders.

In some cases, a naïve helper T cell that encounters its antigen in a peripheral lymphoid organ develops into an effector cell that suppresses rather than helps an immune response. Such *regulatory T cells,* however, mostly develop in the thymus, as a distinct class of T cell, as we now discuss.

Regulatory T Cells Suppress the Activity of Other T Cells

Regulatory T cells have been difficult to study and characterize, largely because, until recently, there were no good markers to identify them. Indeed, for many years immunologists questioned whether such cells existed or not. They were originally identified by their ability to suppress the activity of other lymphocytes and were therefore called *suppressor T cells.* When markers became available, they were renamed *regulatory T cells* and shown to suppress the activity of effector helper and cytotoxic T cells and of dendritic cells. Although they make up less than 10% of the T cells in the blood and peripheral lymphoid organs, regulatory T cells play a crucial part in immunological self tolerance by suppressing the activity of self-reactive effector helper and cytotoxic T cells. They also help prevent excessive T cell responses to microbial antigens in chronic infections. In both these ways, they help prevent adaptive immune responses from damaging host tissues.

A breakthrough in understanding regulatory T cells was the discovery that they alone express the transcription factor *Foxp3,* which serves as both an unambiguous marker of these cells and a master controller of their development. When the gene encoding this protein is inactivated in mice or humans, for example, the individuals specifically fail to produce regulatory T cells and

develop an early and fatal autoimmune disease involving multiple organs. It is still uncertain how regulatory T cells suppress the action of effector T cells or dendritic cells, but one pathway is thought to involve the secretion of the inhibitory cytokines *TGFβ* and *interleukin 10 (IL10)*.

We now turn to the crucial role of MHC proteins in presenting antigen to T cells.

T Cells Recognize Foreign Peptides Bound to MHC Proteins

As discussed earlier, both cytotoxic T cells and helper T cells are initially activated in peripheral lymphoid organs by recognizing foreign antigen on the surface of an antigen-presenting cell, which is a dendritic cell in the case of a naïve T cell. The antigen is usually in the form of peptide fragments that are generated by the degradation of foreign protein antigens inside the antigen-presenting cell. The recognition process depends on the presence of **MHC proteins** in the antigen-presenting cell. These bind fragments, carry them to the cell surface, and present them there—along with co-stimulatory signals—to the T cells. Once activated, effector T cells then recognize the same peptide–MHC complex on the surface of the target cell they influence, which may be a B cell, a cytotoxic T cell, or an infected macrophage in the case of a helper T cell, or any infected host cell in the case of a cytotoxic T cell; for helper T cells, it may also be the dendritic cell itself.

A large complex of genes called the **major histocompatibility complex** (**MHC**) encodes MHC proteins. There are two main classes of MHC proteins, and they are both structurally and functionally distinct. *Class I MHC proteins* mainly present foreign peptides to cytotoxic T cells, and *class II MHC proteins* mainly present foreign peptides to helper and regulatory T cells (**Figure 25–49**).

Before examining the mechanisms by which protein antigens are processed for presentation to T cells, we must look more closely at the MHC proteins themselves, which have such a critical role in T cell function.

MHC Proteins Were Identified in Transplantation Reactions Before Their Functions Were Known

MHC proteins were initially identified as the main antigens recognized in **transplantation reactions**. When organ grafts are exchanged between adult individuals, either of the same species *(allografts)* or of different species *(xenografts)*, they are usually rejected. In the 1950s, skin grafting experiments between different strains of mice demonstrated that *graft rejection* is an adaptive immune response to the foreign antigens on the surface of the grafted cells. Rejection is mediated mainly by T cells, which react against genetically "foreign" versions of cell-surface proteins called *histocompatibility molecules* (from the Greek word *histos,* meaning "tissue"). The MHC proteins encoded by the clustered genes of the major histocompatibility complex (MHC) are by far the most important of these. MHC proteins are expressed on the cells of all higher vertebrates. They were first demonstrated in mice, where they are called *H-2 antigens (h*istocompatibility-*2*

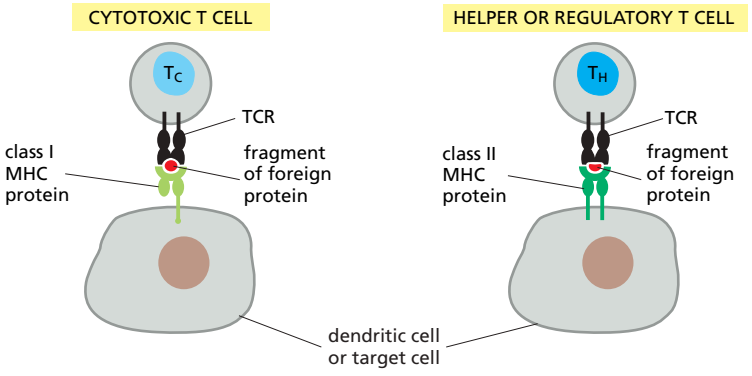


Figure 25–49 Recognition by T cells of foreign peptides bound to MHC proteins. Cytotoxic T cells recognize foreign peptides in association with class I MHC proteins, whereas helper T cells and regulatory T cells recognize foreign peptides in association with class II MHC proteins. In both cases, the T cell recognizes the peptide–MHC complexes on the surface of a dendritic cell or a target cell.

antigens). In humans they are called *HLA antigens (h*uman-*l*eucocyte-*a*ssociated antigens) because they were first demonstrated on leucocytes (white blood cells).

Three remarkable properties of MHC proteins baffled immunologists for many years. First, MHC proteins are overwhelmingly the antigens that are recognized in T-cell-mediated transplantation reactions. Second, an unusually large fraction of T cells are able to recognize foreign MHC proteins: whereas fewer than 0.001% of an individual's naïve T cells respond to a typical viral antigen, up to 10% of them respond to the foreign MHC proteins of another individual. Third, some of the genes that code for MHC proteins are the most *polymorphic* known in higher vertebrates. That is, within a species, there are an extraordinarily large number of *alleles* (alternative forms of the same gene) present (in some cases more than 400), without any one allele predominating. As each individual has at least 12 genes encoding MHC proteins (as discussed later), it is very rare for two unrelated individuals to have an identical set of MHC proteins. These differences make it very difficult to match donor and recipient for organ transplantation unless they are closely related.

Of course, a vertebrate does not need to protect itself against invasion by foreign vertebrate cells. So the apparent obsession of its T cells with foreign MHC proteins and the extreme polymorphism of these molecules were a great puzzle. The puzzle was at least partly solved when researchers discovered that (1) MHC proteins bind fragments of foreign proteins and display them on the surface of host cells for T cells to recognize, and (2) T cells respond to foreign MHC proteins in the same way they respond to self MHC proteins that have foreign antigen bound to them.

Class I and Class II MHC Proteins Are Structurally Similar Heterodimers

Class I and class II MHC proteins have very similar overall structures. They are both transmembrane heterodimers with extracellular N-terminal domains that bind antigen for presentation to T cells.

Class I MHC proteins consist of a transmembrane α chain, which is encoded by a class I MHC gene, and a small extracellular protein called *β₂-microglobulin* (**Figure 25–50**A). The β₂-microglobulin does not span the membrane and is

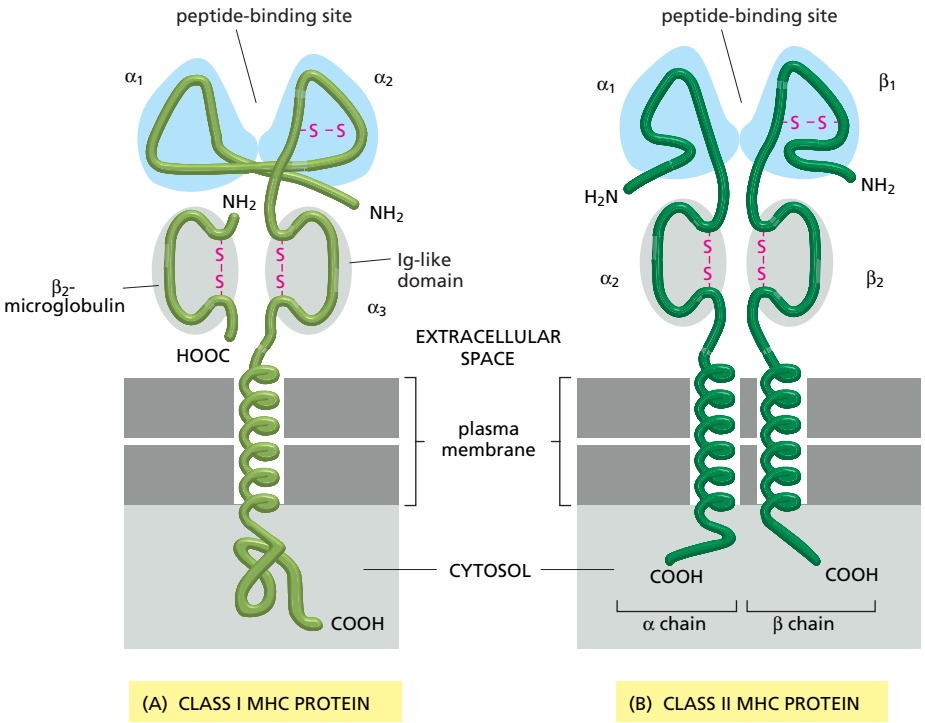


Figure 25–50 Class I and class II MHC proteins. (A) The α chain of the class I molecule has three extracellular domains, α_1, α_2, and α_3, encoded by separate exons. It is noncovalently associated with a smaller polypeptide chain, β_2-microglobulin, which is not encoded within the MHC. The α_3 domain and β_2-microglobulin are Ig-like. While β_2-microglobulin is invariant, the α chain is extremely polymorphic, mainly in the α_1 and α_2 domains. (B) In class II MHC proteins, both chains are polymorphic, mainly in the α_1 and β_1 domains; the α_2 and β_2 domains are Ig-like. Thus, there are striking similarities between class I and class II MHC proteins. In both, the two outermost domains (shaded in *blue*) are polymorphic and interact to form a groove that binds peptide fragments of foreign proteins and presents them to T cells.

encoded by a gene that does not lie in the MHC gene cluster. The α chain is folded into three extracellular globular domains (α_1, α_2, and α_3), and the α_3 domain and the β_2-microglobulin, which are closest to the membrane, are both similar to an Ig domain. The two N-terminal domains of the α chain, which are farthest from the membrane, contain the polymorphic (variable) amino acids that T cells recognize in transplantation reactions. These domains bind a peptide and present it to cytotoxic T cells.

Like class I MHC proteins, **class II MHC proteins** are heterodimers with two conserved Ig-like domains close to the membrane, and two polymorphic (variable) N-terminal domains farthest from the membrane. In these proteins, however, both chains (α and β) are encoded by genes within the MHC, and both span the membrane (Figure 25–50B). The two polymorphic domains bind a peptide and present it to helper or regulatory T cells.

The presence of Ig-like domains in class I and class II proteins suggests that MHC proteins and antibodies have a common evolutionary history. The locations of the genes that encode class I and class II MHC proteins in humans are shown in **Figure 25–51**, where we illustrate how an individual can make six types of class I MHC proteins and more than six types of class II proteins.

In addition to the classic class I MHC proteins, there are many *class-I-MHC-like proteins*, which form dimers with β_2-microglobulin. These proteins are encoded by genes outside the MHC and are much less polymorphic than MHC proteins, but some of them present specific microbial antigens, including some lipids and glycolipids, to T cells. Although the functions of most of them are unknown, some have a role in brain development.

An MHC Protein Binds a Peptide and Interacts with a T Cell Receptor

Any individual can make only a small number of different classic MHC proteins, which together must be able to present peptide fragments from almost any foreign protein to T cells. Thus, unlike an antibody molecule, each MHC protein has to be able to bind a very large number of different peptides. X-ray crystallographic analyses of MHC proteins have revealed the structural basis for this versatility.

As shown in **Figure 25–52**A, a class I MHC protein has a single *peptide-binding site* located at one end of the molecule, facing away from the plasma membrane. This site consists of a deep groove between two long α helices; the groove narrows at both ends so that it is only large enough to accommodate an extended peptide about 8–10 amino acids long. In fact, when a class I MHC protein was first analyzed by x-ray crystallography, this groove contained bound peptides that had co-crystallized with the MHC protein (Figure 25–52B), suggesting that once a peptide binds to this site it does not normally dissociate.

A typical peptide binds in the groove of a class I MHC protein in an extended conformation, with its terminal amino group bound to invariant amino acids of the MHC protein at one end of the groove and its terminal carboxyl group bound to invariant amino acids at the other end of the groove (**Figure 25–53**). Some amino acid side chains of the peptide bind to variable (polymorphic) amino acids of the MHC protein distributed along the groove, while other side chains point outward, in a position to be recognized by TCRs on cytotoxic T cells. Because the invariant amino acids of the MHC protein at the ends of the groove recognize features of the peptide backbone that are common to all peptides, each allelic form of a class I MHC protein can bind a large variety of peptides of

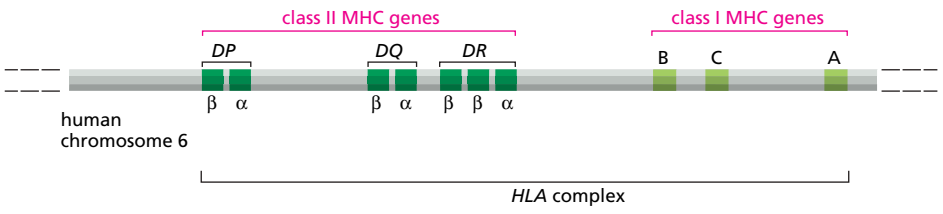


Figure 25–51 Human MHC genes. This simplified schematic drawing shows the location of the genes that encode the transmembrane subunits of class I *(light green)* and class II *(dark green)* MHC proteins. The genes shown encode three types of class I proteins (HLA-A, HLA-B, and HLA-C) and three types of class II MHC proteins (HLA-DP, HLA-DQ, and HLA-DR). An individual can therefore make six types of class I MHC proteins (three encoded by maternal genes and three by paternal genes) and more than six types of class II MHC proteins. The number of class II MHC proteins that can be made is greater than six because there are two DR β genes and because maternally encoded and paternally encoded polypeptide chains can sometimes pair.

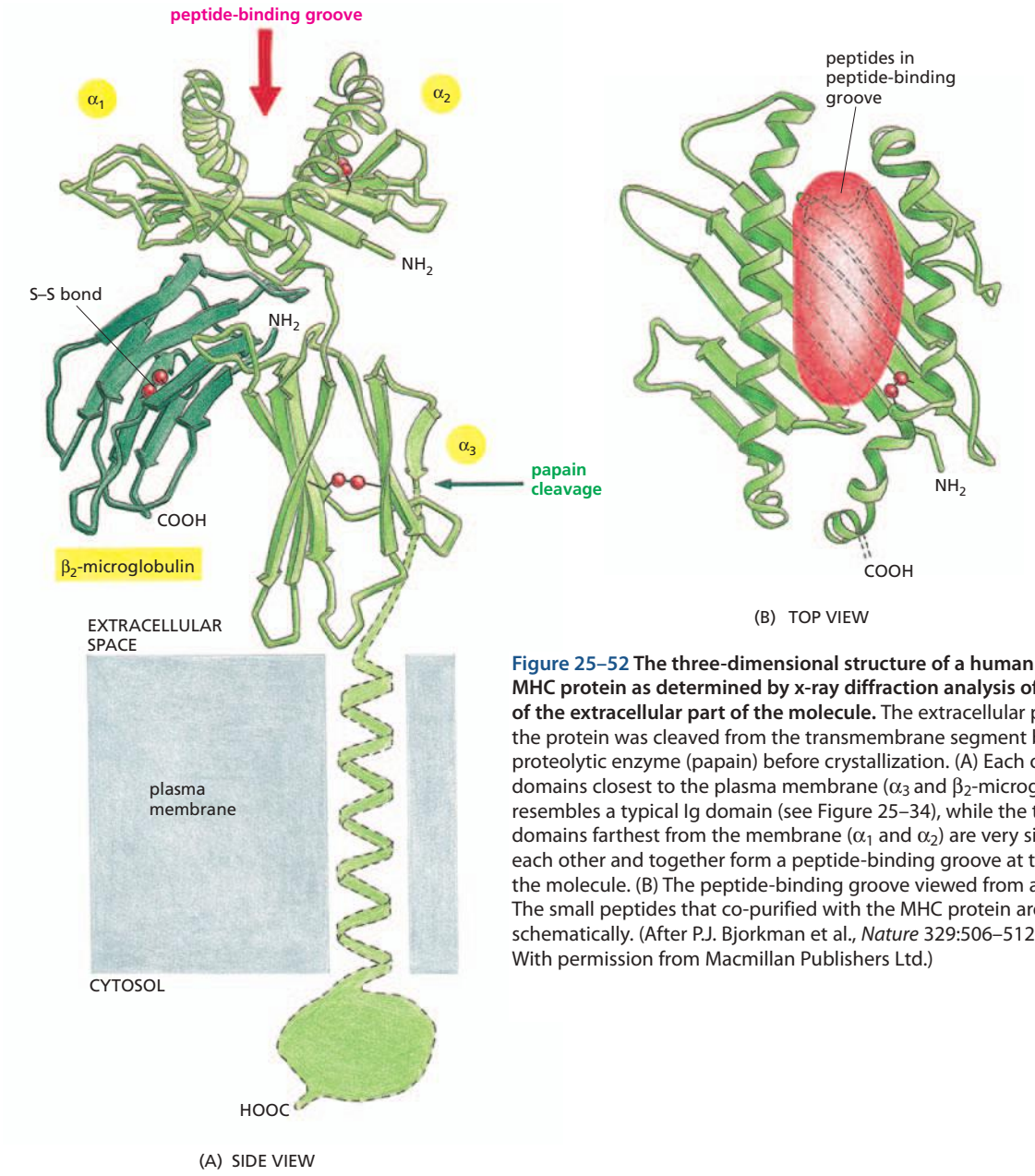


Figure 25–52 The three-dimensional structure of a human class I MHC protein as determined by x-ray diffraction analysis of crystals of the extracellular part of the molecule. The extracellular part of the protein was cleaved from the transmembrane segment by a proteolytic enzyme (papain) before crystallization. (A) Each of the two domains closest to the plasma membrane (α_3 and β_2-microglobulin) resembles a typical Ig domain (see Figure 25–34), while the two domains farthest from the membrane (α_1 and α_2) are very similar to each other and together form a peptide-binding groove at the top of the molecule. (B) The peptide-binding groove viewed from above. The small peptides that co-purified with the MHC protein are shown schematically. (After P.J. Bjorkman et al., *Nature* 329:506–512, 1987. With permission from Macmillan Publishers Ltd.)

diverse sequence. At the same time, the polymorphic MHC amino acids along the groove, which bind specific side chains of the peptide, ensure that each allelic form binds and presents a distinct characteristic set of peptides. Thus, the six types of class I MHC proteins in an individual can present a broad range of foreign peptides to the cytotoxic T cells, but in each individual they do so in slightly different ways.

The three-dimensional structure of class II MHC proteins is very similar to that of class I proteins, but the antigen-binding groove does not narrow at the ends, so it can accommodate longer peptides, which are usually 12–20 amino acids long. Moreover, the peptide is not bound at its ends but is instead held by interactions with invariant amino acids of the MHC protein distributed along the length of the groove (**Figure 25–54**). As in the case of class I MHC proteins, side chains of other amino acids in the peptide either bind to polymorphic MHC amino acids along the groove or point upward to be recognized by TCRs on helper or regulatory T cells. A class II MHC groove can accommodate a more heterogeneous set of peptides than can a class I MHC groove. Thus, although an individual makes only a small number of types of class II proteins, each with its

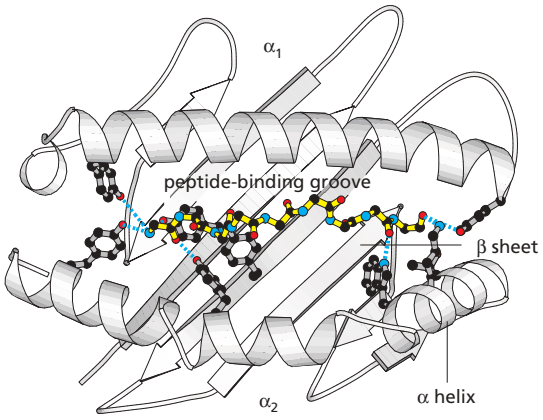


Figure 25–53 A peptide bound in the groove of a class I MHC protein. <AAGT> Schematic drawing of a top view of the groove. A ribbon drawing of the MHC groove is shown in *gray;* it is formed by the α_1 and α_2 domains of the protein (see Figures 25–50A and 25–52A). The backbone of the bound peptide is shown in *yellow,* with the carbon atoms in *black,* oxygen atoms in *red,* and nitrogen atoms in *blue;* the amino terminus of the peptide is to the left. Note that the terminal amino and carboxyl groups of the peptide backbone bind via hydrogen and ionic bonds (shown as *dotted blue lines*) to the side chains of invariant MHC amino acids of the MHC protein towards the ends of the groove (shown in full). Although not illustrated in the drawing, the side chains of some amino acids of the peptide bind to variable (polymorphic) amino acids of the groove, while others point upward and can be recognized by TCRs on a cytotoxic T cell. (Courtesy of Paul Travers.)

own unique peptide-binding groove, together these proteins can bind and present an enormous variety of foreign peptides to helper T cells, which have a crucial role in almost all adaptive immune responses.

X-ray crystallographic analyses of complexes formed between a soluble TCR and a soluble MHC protein with peptide in its binding groove revealed the way in which the TCR recognizes a peptide–MHC complex. Recombinant DNA technology produced the soluble proteins for these experiments. In each case studied, the TCR fits diagonally across the peptide-binding groove and binds through its V_α and V_β hypervariable loops to both the walls of the groove and the peptide (**Figure 25–55**). Soluble peptide–MHC complexes are now widely used to detect T cells with a particular specificity; they are usually cross-linked into tetramers so that they can bind to four TCRs on the T cell surface with strong avidity.

MHC Proteins Help Direct T Cells to Their Appropriate Targets

Virtually all nucleated vertebrate cells express class I MHC proteins. This is presumably because effector cytotoxic T cells must be able to focus on and kill any cell in the body that becomes infected with an intracellular microbe such as a virus. Class II proteins, by contrast, are normally confined largely to cells that take up foreign antigens from the extracellular fluid and interact with helper T cells. These cells also express class I MHC proteins. They include dendritic cells, which initially activate naïve helper T cells, as well as the targets of effector helper T cells, such as macrophages and B cells.

It is important that effector cytotoxic T cells focus their attack mainly on cells that *make* the foreign antigens (such as viral proteins), while helper T cells focus their help mainly on cells that have taken up foreign antigens from the extracellular fluid. Since the former type of target cell is always a menace, while

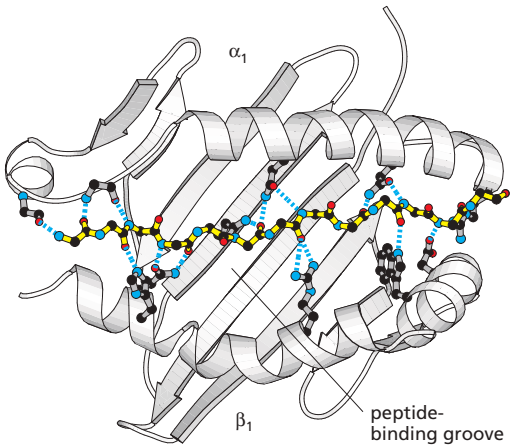


Figure 25–54 A peptide bound in the groove of a class II MHC protein. <GAAA> Schematic drawing similar to that shown in Figure 25–53. The groove is formed by the amino terminal domains of the α and β chains (α_1 and β_1—see Figure 25–50B). Note that the peptide extends beyond the ends of the groove and that its backbone binds through hydrogen bonds distributed along the length of the peptide to invariant amino acid side chains in the groove. (Courtesy of Paul Travers.)

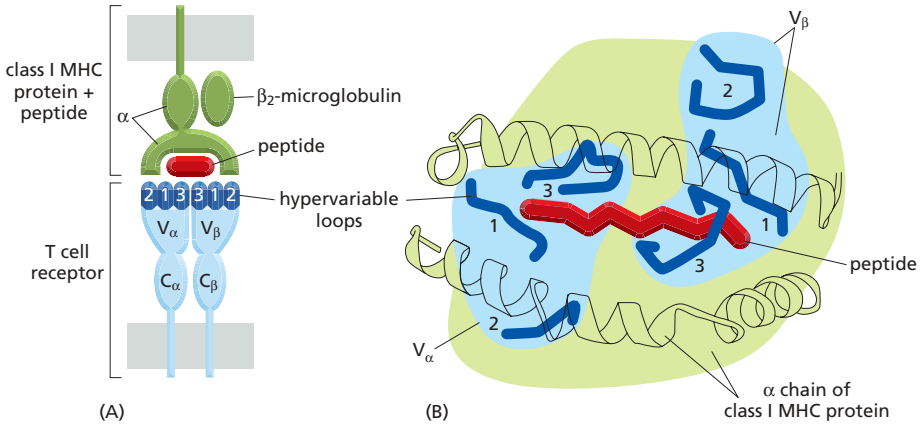


Figure 25–55 The interaction of a T cell receptor with a viral peptide bound to a class I MHC protein. (A) Schematic view of the hypervariable loops of the V_α and V_β domains of the T cell receptor interacting with the peptide and the walls of the peptide-binding groove of the MHC protein. Note that the third hypervariable loops, which are the most variable, primarily interact with the peptide, whereas the other hypervariable loops primarily bind to the walls of the peptide-binding groove. (B) Drawing of the "footprint" of the V domains *(light blue)* and hypervariable loops *(dark blue)* of the receptor over the peptide-binding groove, as determined by x-ray diffraction. The V_α domain covers the amino half of the peptide, while the V_β domain covers the carboxyl half. Note that the receptor is oriented diagonally across the peptide-binding groove. (B, adapted from D.N. Garboczi et al., *Nature* 384:134–141, 1996. With permission from Macmillan Publishers Ltd.)

the latter type is essential for the body's adaptive immune defenses, it is vitally important that T cells do not confuse the two target cells and misdirect their cytotoxic and helper functions. Therefore, in addition to the antigen receptor that recognizes a peptide–MHC complex, each of the three major classes of T cells also expresses a *co-receptor* that recognizes a separate, invariant part of the appropriate class of MHC protein. These two co-receptors, called CD4 and CD8, help direct helper (and regulatory) T cells and cytotoxic T cells, respectively, to their appropriate targets. **Table 25–2** reviews the properties of class I and class II MHC proteins.

CD4 and CD8 Co-Receptors Bind to Invariant Parts of MHC Proteins

The affinity of TCRs for peptide–MHC complexes on an antigen-presenting cell or target cell is usually too low by itself to mediate a functional interaction between the two cells. T cells normally require *accessory receptors* to help stabilize the interaction by increasing the overall strength of the cell–cell adhesion. Unlike TCRs or MHC proteins, the accessory receptors do not bind foreign antigens and are invariant.

When accessory receptors have a direct role in activating the T cell by generating their own intracellular signals, they are called **co-receptors**. The most important and best understood of the co-receptors on T cells are the CD4 and CD8 proteins, both of which are single-pass transmembrane proteins with extracellular Ig-like domains. Like TCRs, they recognize MHC proteins, but, unlike TCRs, they bind to invariant parts of the protein, far away from the peptide-binding groove. **CD4** is expressed on both helper T cells and regulatory T cells and binds to class II MHC proteins, whereas **CD8** is expressed on cytotoxic T cells and binds to class I MHC proteins (**Figure 25–56**). Thus, CD4 and CD8 contribute to T cell recognition by helping the T cell to focus on particular MHC proteins, and thereby on particular types of target cells: the recognition of class I

Table 25–2 Properties of Human Class I and Class II MHC Proteins

	CLASS I	CLASS II
Genetic loci	*HLA-A, HLA-B, HLA-C*	*DP, DQ, DR*
Chain structure	α chain + β_2-microglobulin	α chain + β chain
Cell distribution	most nucleated cells	dendritic cells, B cells, macrophages, thymus epithelial cells, some others
Presents antigen to	cytotoxic T cells	helper T cells, regulatory T cells
Source of peptide fragments	mainly proteins made in cytoplasm	mainly endocytosed plasma membrane and extracellular proteins
Polymorphic domains	$\alpha_1 + \alpha_2$	$\alpha_1 + \beta_1$
Recognition by co-receptor	CD8	CD4

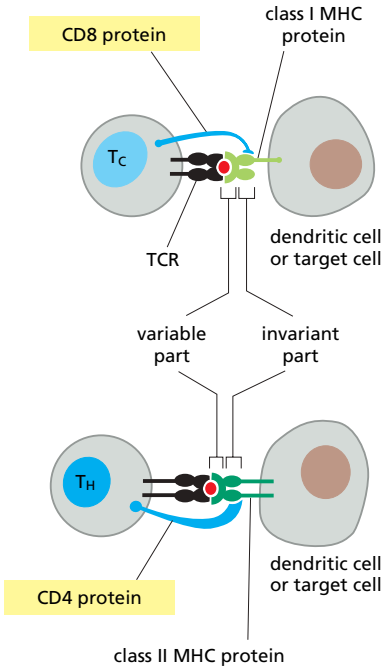


Figure 25–56 CD4 and CD8 co-receptors on the surface of T cells. Cytotoxic T cells (T_C) express CD8, which recognizes class I MHC proteins, whereas helper T cells (T_H) and regulatory T cells (not shown) express CD4, which recognizes class II MHC proteins. Note that the co-receptors bind to the same MHC protein that the TCR has engaged, so that they are brought together with TCRs during the antigen recognition process. Whereas the TCR binds to the variable (polymorphic) parts of the MHC protein that form the peptide-binding groove, the co-receptor binds to the invariant part, well away from the groove.

MHC proteins allows cytotoxic T cells to focus on any host cell, while the recognition of class II MHC proteins allows helper T cells to focus on a small subset of cells—most notably dendritic cells, macrophages, and B cells. The cytoplasmic tail of the CD4 and CD8 proteins is associated with a member of the Src family of cytoplasmic tyrosine kinases called *Lck,* which phosphorylates various intracellular proteins on tyrosines and thereby participates in the activation of the T cell (discussed in Chapter 15). Antibodies to CD4 and CD8 are widely used as tools to help distinguish between the main classes of T cells, in both humans and experimental animals: whereas only cytotoxic T cells express CD8, both helper and regulatory T cells express CD4.

Ironically, the AIDS virus (HIV) uses CD4 molecules (as well as chemokine receptors) to enter helper T cells. AIDS patients are susceptible to infection by microbes that are not normally dangerous because HIV depletes helper T cells. As a result, most AIDS patients die of infection within several years of the onset of symptoms, unless they are treated with a combination of powerful anti-HIV drugs. HIV also uses CD4 and chemokine receptors to enter macrophages, which also have both of these receptors on their surface.

Before a T cell can recognize a foreign protein, the protein has to be processed inside an antigen-presenting cell or target cell so that it can be displayed as peptide–MHC complexes on the cell surface. We now consider how a virus-infected antigen-presenting cell or target cell processes viral proteins for presentation to a cytotoxic T cell. We then discuss how ingested foreign proteins are processed for presentation to a helper T cell.

Cytotoxic T Cells Respond to Fragments of Foreign Cytosolic Proteins in Association with Class I MHC Proteins

An experiment performed in the 1970s provided dramatic evidence that class I MHC proteins are involved in the recognition of viral antigens by cytotoxic T cells. Researchers found that effector cytotoxic T cells from a virus-infected mouse could kill cultured cells infected with the same virus only if these target cells expressed some of the same class I MHC proteins as the infected mouse. This experiment demonstrated that the cytotoxic T cells of any individual can recognize a specific foreign antigen on a target cell only when the target cell expresses at least some of the allelic forms of class I MHC proteins expressed by that individual, a phenomenon known as *MHC restriction* (**Figure 25–57**).

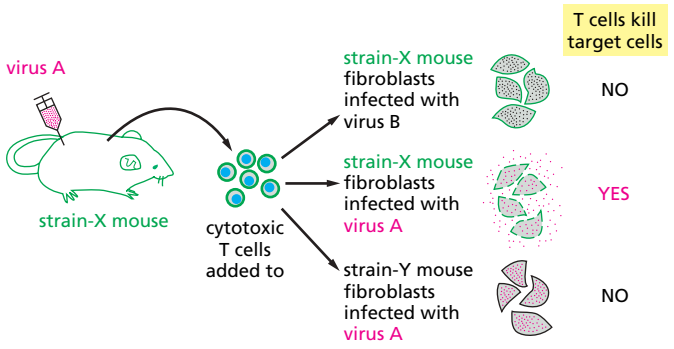


Figure 25–57 The classic experiment showing that an effector cytotoxic T cell recognizes some aspect of the surface of the host target cell in addition to a viral antigen. Mice of strain X are infected with virus A. Seven days later, the spleens of these mice contain effector cytotoxic T cells able to kill virus-infected, strain-X fibroblasts in cell culture. As expected, they kill only fibroblasts infected with virus A and not those infected with virus B. Thus, the cytotoxic T cells are virus-specific. The same T cells, however, are also unable to kill fibroblasts from strain-Y mice infected with the same virus A, indicating that the cytotoxic T cells recognize a genetic difference between the two kinds of fibroblasts and not just the virus. Pinning down the difference required the use of special strains of mice (known as *congenic strains*) that either were genetically identical except for the alleles at their class I MHC loci or were genetically different except for these alleles. In this way, it was found that the killing of infected target cells required that they express at least one of the same class I MHC alleles as expressed by the original infected mouse. This result suggested that class I MHC proteins are somehow necessary to present cell-surface-bound viral antigens to effector cytotoxic T cells and that they do so in a highly specific manner.

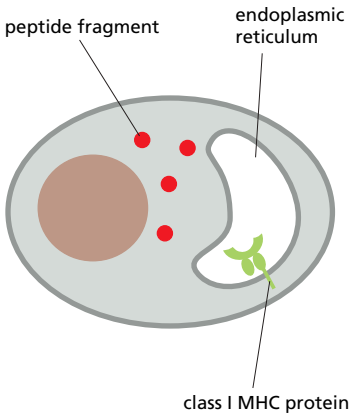


Figure 25–58 The peptide-transport problem. How do peptide fragments get from the cytosol, where they are produced, into the ER lumen, where the peptide-binding grooves of class I MHC proteins are located? A special transport process is required.

Subsequent evidence indicated that, in killing a virus-infected cell, a cytotoxic T cell recognizes degraded fragments of internal proteins of the virus that are bound to class I MHC proteins on the infected cell surface. Because a T cell can recognize tiny amounts of foreign antigen (as few as 1–10 peptide–MHC complexes for T cells with the highest affinity receptors), only a tiny fraction of the fragments generated from viral proteins have to bind to class I MHC proteins and get to the cell surface to trigger an effector cytotoxic T cell to attack.

The internal viral proteins are synthesized in the cytosol of the infected cell. As discussed in Chapter 3, proteolytic degradation in the cytosol is mainly mediated by an ATP- and ubiquitin-dependent mechanism that operates in *proteasomes*—large proteolytic enzyme complexes constructed from many different protein subunits. All proteasomes are probably able to generate peptide fragments of appropriate length to fit in the groove of class I MHC proteins, as even bacterial proteasomes can do so. This suggests that the MHC groove evolved to fit peptides of around this length. Nonetheless, some proteasomes are apparently specialized for producing peptides for class I MHC proteins, as they contain two subunits that are encoded by genes located within the MHC chromosomal region.

How do peptides generated in the cytosol make contact with the peptide-binding groove of class I MHC proteins in the lumen of the endoplasmic reticulum (**Figure 25–58**)? The answer initially came through observations of mutant cells in which class I MHC proteins are not expressed at the cell surface but are instead degraded within the cell. The mutant genes in these cells proved to encode subunits of a protein belonging to the family of *ABC transporters,* which we discuss in Chapter 11. This transporter is located in the ER membrane and uses the energy of ATP hydrolysis to pump peptides from the cytosol into the ER lumen. The genes encoding its two subunits are located in the MHC chromosomal region, and if either gene is inactivated by mutation, cells are unable to supply peptides to class I MHC proteins. It is striking that the class I MHC proteins in such mutant cells are degraded inside the cell without reaching the cell surface. This occurs because peptide binding is normally required for the proper folding of these proteins: until it binds a peptide, a class I MHC protein remains in the ER, tethered to an ABC transporter by a chaperone protein; without a peptide bound, the trapped MHC proteins in the mutant cells eventually undergo proteolysis (**Figure 25–59**).

In all cells, peptide fragments come from the cells' own cytosolic and nuclear proteins that are degraded in the processes of normal protein turnover and quality control mechanisms. (Surprisingly, more than 30% of the proteins made by mammalian cells are apparently faulty and are degraded in proteasomes soon after they are synthesized.) These peptides are constantly being pumped into the ER and are carried to the cell surface by class I MHC proteins. They are not antigenic, however, because the cytotoxic T cells that could recognize them have either been eliminated, inactivated, or suppressed by regulatory T cells in the process of self-tolerance (see Figure 25–13).

When an antigen activates cytotoxic T cells or T_H1 helper T cells to become effector cells, the effector cells secrete the cytokine **interferon-γ** (**IFNγ**), which greatly enhances anti-viral responses. The IFNγ acts on virus-infected host cells in two ways. It blocks viral replication, and it increases the expression of many genes within the MHC chromosomal region. These genes include those that encode class I MHC proteins, the two specialized proteasome subunits, and the two subunits of the peptide transporter located in the ER (**Figure 25–60**). Thus, the machinery in a host cell that is required for presenting viral antigens to cytotoxic T cells is coordinately called into action by IFNγ, creating a positive feedback loop that amplifies the immune response and culminates in the death of the infected cell.

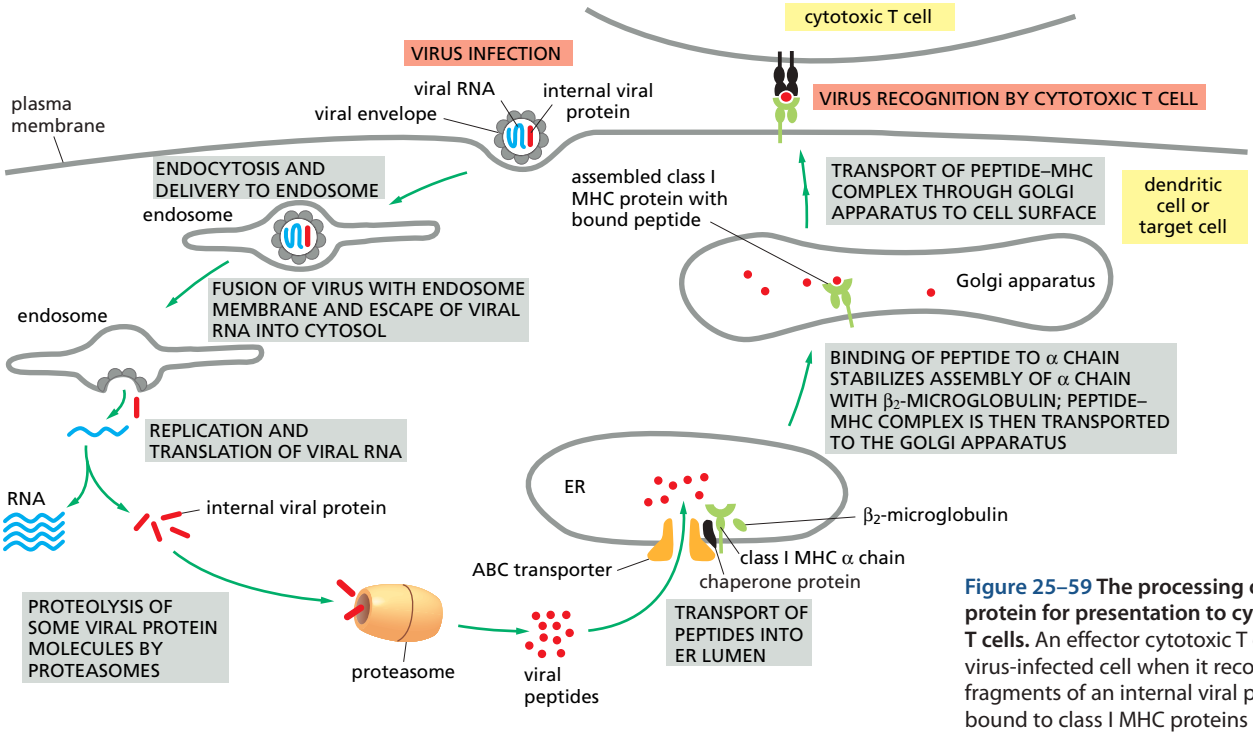


Figure 25–59 The processing of a viral protein for presentation to cytotoxic T cells. An effector cytotoxic T cell kills a virus-infected cell when it recognizes fragments of an internal viral protein bound to class I MHC proteins on the surface of the infected cell. Not all viruses enter the cell in the way that this enveloped RNA virus does, but fragments of internal viral proteins always follow the pathway shown. Only a small proportion of the viral proteins synthesized in the cytosol are degraded and transported to the cell surface, but this is sufficient to attract an attack by a cytotoxic T cell. Several chaperone proteins (only one of which is shown) in the ER lumen aid the folding and assembly of class I MHC proteins. The chaperones bind to the class I MHC α chain and act sequentially. The last one binds the MHC protein to the ABC transporter, as shown. The assembly of class I MHC proteins and their transport to the cell surface requires the binding of peptide.

Helper T Cells Respond to Fragments of Endocytosed Foreign Protein Associated with Class II MHC Proteins

Unlike cytotoxic T cells, helper T cells do not act directly to kill infected cells. Instead, they stimulate macrophages to be more effective in destroying intracellular microorganisms, and they help B cells and cytotoxic T cells to respond to microbial antigens.

Like the viral proteins presented to cytotoxic T cells, the proteins presented to helper T cells on dendritic cells or target cells are degraded fragments of foreign proteins. The fragments are bound to class II MHC proteins in much the same way that virus-derived peptides are bound to class I MHC proteins. But both the source of the peptide fragments presented and the route they take to find the MHC proteins generally differ.

Rather than being derived from foreign protein synthesized in the cytosol of a cell, the foreign peptides presented to helper T cells are derived from endosomes. Some come from extracellular microbes or their products that the antigen-presenting cell has endocytosed and degraded in the acidic environment of its endosomes. Others come from microbes growing within the endocytic compartment of the antigen-presenting cell. In either case, the peptides do not have to be pumped across a membrane because they are generated in a compartment that is topologically equivalent to the extracellular space. Instead of entering the lumen of the ER, where the class II MHC proteins are synthesized and assembled, they bind to preassembled class II heterodimers in an endosomal compartment. When the peptide binds, the class II MHC protein alters its conformation, trapping the

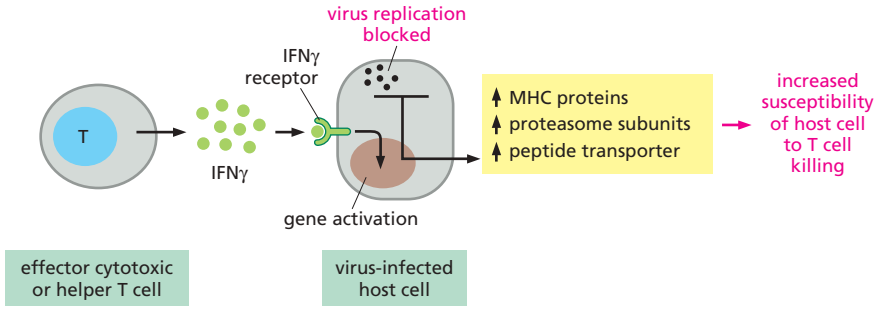


Figure 25–60 Some effects of interferon-γ (IFNγ) on virus-infected cells. The activated IFNγ receptors signal to the nucleus, altering gene transcription, which leads to the effects indicated. The effects shaded in *yellow* tend to make the infected cell a better target for killing by an effector cytotoxic T cell.

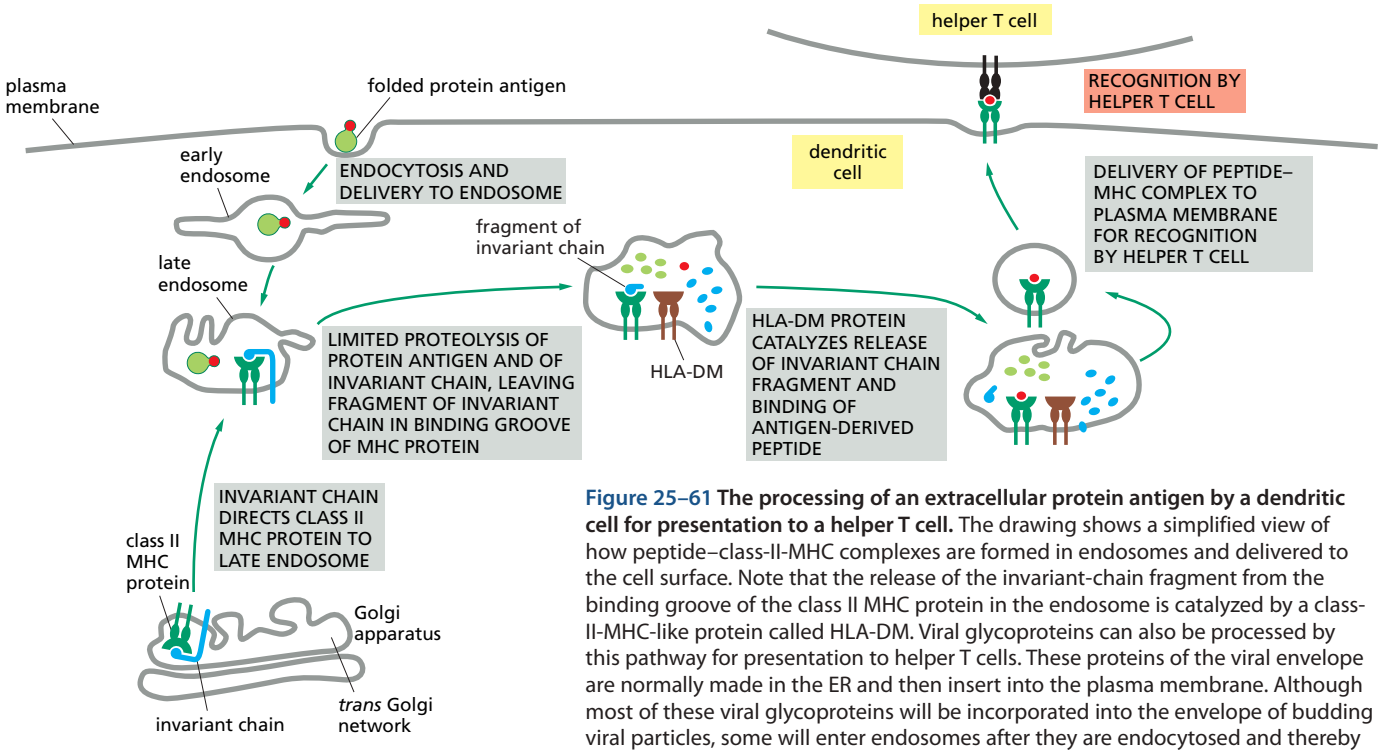


Figure 25–61 The processing of an extracellular protein antigen by a dendritic cell for presentation to a helper T cell. The drawing shows a simplified view of how peptide–class-II-MHC complexes are formed in endosomes and delivered to the cell surface. Note that the release of the invariant-chain fragment from the binding groove of the class II MHC protein in the endosome is catalyzed by a class-II-MHC-like protein called HLA-DM. Viral glycoproteins can also be processed by this pathway for presentation to helper T cells. These proteins of the viral envelope are normally made in the ER and then insert into the plasma membrane. Although most of these viral glycoproteins will be incorporated into the envelope of budding viral particles, some will enter endosomes after they are endocytosed and thereby enter the class II MHC pathway.

peptide in the binding groove and transferring it to the cell surface for presentation to helper T cells.

A newly synthesized class II MHC protein must avoid clogging its binding groove prematurely in the ER lumen with peptides pumped in from the cytosol. A special polypeptide, called the **invariant chain**, ensures this by associating with newly synthesized class II MHC heterodimers in the ER. Part of its polypeptide chain lies within the peptide-binding groove of the MHC protein, thereby blocking the groove from binding peptides in the lumen of the ER. The invariant chain also directs class II MHC proteins from the *trans* Golgi network to a late endosomal compartment. Here, proteases cleave the invariant chain, leaving only a short fragment bound in the peptide-binding groove of the MHC protein. This fragment is then released, freeing the MHC protein to bind peptides derived from endocytosed proteins (**Figure 25–61**). In this way, the functional differences between class I and class II MHC proteins are ensured—the former mainly presenting molecules that come from the cytosol, the latter mainly presenting molecules that come from the endocytic compartment.

This distinction between antigen presentation to cytotoxic T cells and to helper T cells, however, is not absolute. Dendritic cells, for example, need to be able to activate cytotoxic T cells to kill virus-infected cells even when the virus does not infect dendritic cells themselves. To do so, dendritic cells use a process called **cross-presentation**, which begins when they phagocytose fragments of virus-infected cells. They then actively transport viral proteins out of the phagosome into the cytosol, where they are degraded in proteasomes; resulting fragments of the viral proteins are then transported into the ER lumen, where they load onto assembling class I MHC proteins. Cross-presentation in dendritic cells also operates to activate cytotoxic T cells against tumor antigens of cancer cells and against the MHC proteins of foreign organ grafts.

Most of the class I and class II MHC proteins on the surface of a target cell have peptides derived from self proteins in their binding groove. For class I proteins, the fragments mainly derive from degraded cytosolic and nuclear proteins. For class II proteins, they mainly derive from degraded proteins that originate in the plasma membrane or extracellular fluid and are endocytosed. Only a small fraction of the 10^5 or so class II MHC proteins on the surface of an antigen-presenting cell will have foreign peptides bound to them. This is sufficient, however:

even a single copy of such a peptide–MHC complex on a dendritic cell can activate a helper T cell that has a TCR that binds the complex with a high enough affinity.

Potentially Useful T Cells Are Positively Selected in the Thymus

We have seen that T cells recognize antigen in association with self MHC proteins but not in association with foreign MHC proteins (see Figure 25–57): that is, T cells show *MHC restriction*. This restriction results from a process of **positive selection** during T cell development in the thymus. In this process, those immature T cells *(thymocytes)* that will be capable of recognizing foreign peptides presented by self MHC proteins are selected to survive and mature, while most of the remainder, which would be of no use to the animal, undergo apoptosis. Thus, MHC restriction is an acquired property of the immune system that emerges as T cells develop in the thymus.

The most direct way to study the selection process is to follow the fate of a set of developing T cells of known specificity. This can be done by using transgenic mice that express a specific pair of rearranged α and β TCR genes derived from a T cell clone of known antigen and MHC specificity. Such experiments show that the transgenic T cells mature in the thymus and populate the peripheral lymphoid organs only if the transgenic mouse also expresses the same allelic form of MHC protein as is recognized by the transgenic TCR. If the mouse does not express the appropriate MHC protein, the transgenic T cells die in the thymus. Thus, the survival and maturation of a developing T cell depend on a match between its TCR and the MHC proteins expressed in the thymus (which have self peptides derived from the body's own proteins bound to them). Similar experiments using transgenic mice in which MHC expression is confined to specific cell types in the thymus indicate that it is MHC proteins on epithelial cells in the cortex of the thymus that are responsible for this positive selection (**Figure 25–62**).

After positively selected T cells leave the thymus, their continued survival as naïve T cells depends on their continual stimulation by self-peptide–MHC complexes (and the cytokine *IL7*); this stimulation is enough to promote cell survival but not enough to activate the T cells to proliferate or become effector or memory cells.

As part of the positive selection process in the thymus, developing T cells that express TCRs recognizing class I MHC proteins are selected to become

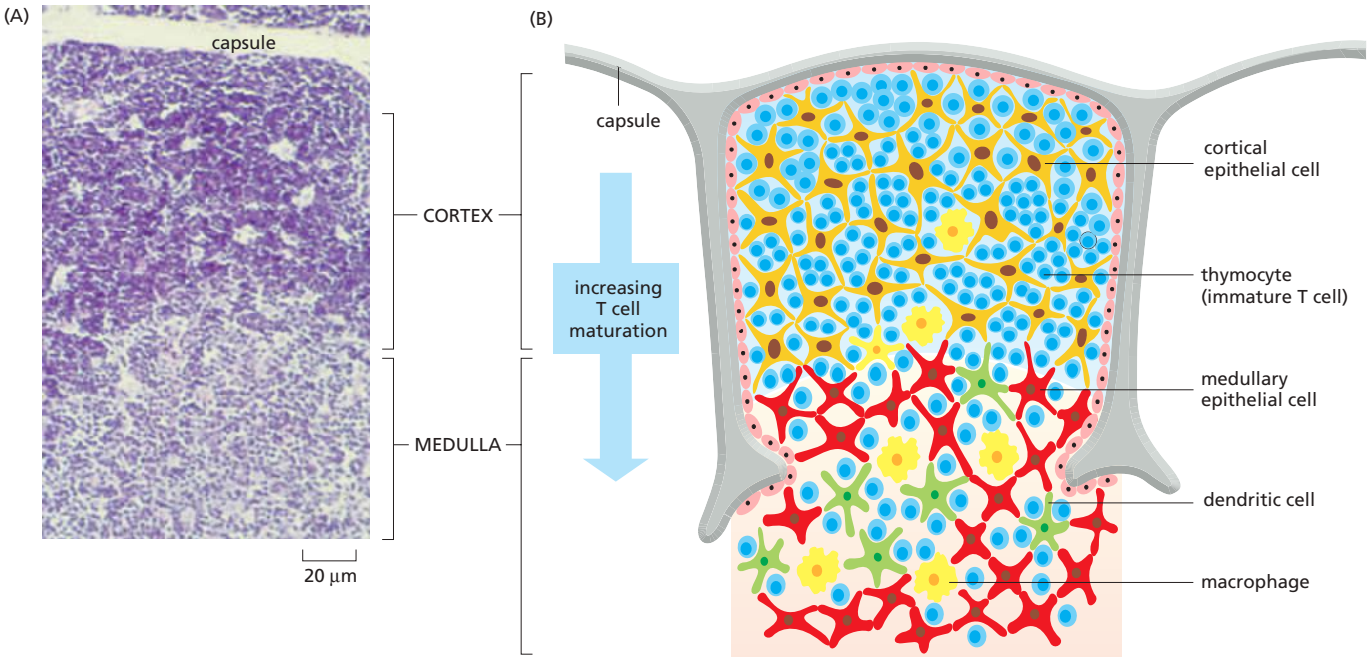


Figure 25–62 The cellular organization of the human thymus. (A) A light micrograph of a stained section of one thymus lobule showing the outer cortex and inner medulla. (B) A schematic drawing of the lobule showing the cellular composition. The cortex contains immature thymocytes, and the medulla contains mature thymocytes. The thymocytes, macrophages, and dendritic cells develop from cells that migrate in from the bone marrow. The functions of these different regions and cell types will be discussed later, when we consider how developing thymocytes are selected for survival. Because of these selection processes, more than 95% of the thymocytes produced in the thymus die by apoptosis. The dead cells are rapidly phagocytosed and digested by the macrophages. (Adapted from K. Murphy et al., Janeway's Immunobiology, 7th ed. New York: Garland Science, 2008.)

cytotoxic cells, while T cells that express TCRs recognizing class II MHC proteins are selected to become either helper cells or regulatory cells. Thus, genetically engineered mice that lack cell-surface class I MHC proteins lack cytotoxic T cells specifically, whereas mice that lack class II MHC proteins lack both helper and regulatory T cells specifically. The development of regulatory T cells depends on special groups of epithelial cells in the thymus medulla called *Hassall's corpuscles*.

The cells that are undergoing positive selection initially express both CD4 and CD8 co-receptors, and these are required for the selection process: without CD4, helper and regulatory T cells fail to develop, and without CD8, cytotoxic T cells fail to develop. Once they develop, cytotoxic T cells lose CD4, and helper and regulatory T cells lose CD8.

Positive selection still leaves a large problem to be solved. If developing cytotoxic and helper T cells with receptors that recognize self peptides associated with self MHC proteins were to mature in the thymus and migrate to peripheral lymphoid tissues, they might wreak havoc. A second, *negative selection* process in the thymus is required to help avoid this potential disaster.

Most Developing Cytotoxic and Helper T Cells That Could Be Activated by Self-Peptide–MHC Complexes Are Eliminated in the Thymus

As discussed previously, a fundamental feature of the adaptive immune system is that it can learn to distinguish self from nonself and normally does not react against self molecules. An important mechanism in achieving this state of immunological self tolerance is the deletion in the thymus of developing self-reactive cytotoxic and helper T cells—that is, T cells whose TCRs bind strongly enough to the complex of a self peptide and a self MHC protein to become activated. Because, as we discuss later, most B cells require helper T cells to respond to antigen, the elimination of self-reactive helper T cells also helps ensure that any self-reactive B cells that escape from the mechanisms responsible for B cell tolerance induction are harmless (see Figure 25–13).

Before discussing the negative selection process that removes self-reactive T cells in the thymus, it is useful to speculate on the logic behind a two-step system that ends up selecting for the small fraction of developing T cells that express a TCR that binds weakly, but not strongly, to a self MHC protein carrying a self-peptide. As illustrated in **Figure 25–63**, it is thought that the production of a large repertoire of such T cells guarantees that at least a few of the T cells will be able to bind strongly to the complex of a foreign peptide with the same MHC protein, thus triggering an adaptive immune response. However, it is of course not enough for the thymus to select *for* T cells that recognize self MHC proteins; it must also select *against* cytotoxic and helper T cells that could become activated by self MHC proteins complexed with self peptides in peripheral lymphoid

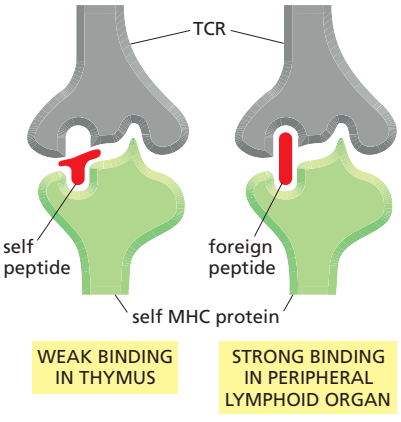


Figure 25–63 Schematic drawing showing how a TCR selected in the thymus because it binds weakly to a self MHC protein complexed with a self peptide can bind strongly to the same MHC protein complexed with a foreign peptide. Because the self peptide complex gives weak binding and the foreign peptide is not present in the thymus, a T cell expressing this TCR in the thymus would be positively selected and also avoid negative selection.

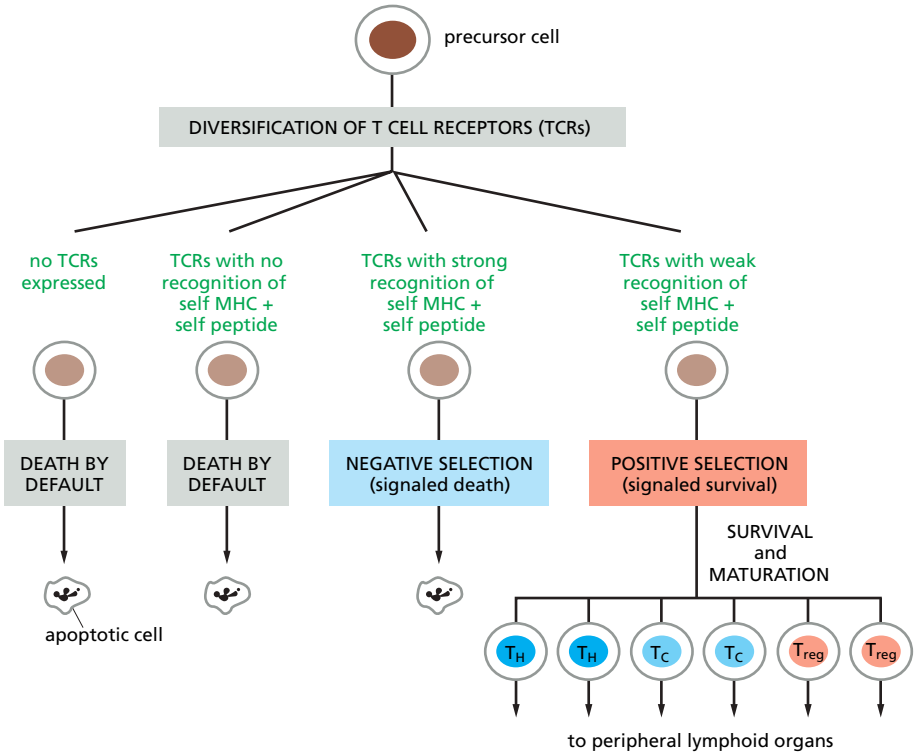


Figure 25–64 The result of positive and negative selection in the thymus. Cells with TCRs that would enable them to respond to foreign peptides in association with self MHC proteins (see Figure 25–63) are positively selected: they survive, mature, and migrate to peripheral lymphoid organs. All of the other cells undergo apoptosis—either because they do not express TCRs that recognize self MHC proteins with self peptides bound or because they recognize such complexes too well and undergo negative selection.

Although not shown, the cells undergoing positive selection initially express both CD4 and CD8 co-receptors. During the process of positive selection, helper T cells (T_H), cytotoxic T cells (T_C), and regulatory T cells (T_{reg}) diverge by a poorly understood mechanism. In this process, helper cells and regulatory cells develop that express CD4 but not CD8 and recognize peptides in association with class II MHC proteins, while cytotoxic cells develop that express CD8 but not CD4 and recognize peptides in association with class I MHC proteins.

organs. The overall goal can be achieved by (1) ensuring the death of cytotoxic and helper T cells that bind *strongly* to the self-peptide–MHC complexes in the thymus while (2) promoting the survival of those that bind weakly and (3) permitting the death of those that do not bind at all. Process 2 constitutes a major part of the positive selection we have just discussed. Process 1 is called **negative selection**, or clonal deletion in the thymus (see Figure 25–13). In both process 1 and process 3, the cells die by apoptosis (**Figure 25–64**).

The most convincing evidence for negative selection in the thymus derives once again from experiments with transgenic mice. The introduction of TCR transgenes encoding a receptor that recognizes a male-specific peptide antigen, for example, results in large numbers of mature T cells expressing the transgenic receptor in both the thymus and peripheral lymphoid organs of female mice, which do not express the peptide. Very few, however, are found in male mice, in which the cells die in the thymus before they have a chance to mature. Like positive selection, negative selection requires the interaction of a TCR and a CD4 or CD8 co-receptor with an appropriate MHC protein. Unlike positive selection of developing helper and cytotoxic T cells, however, which occurs mainly on the surface of epithelial cells in the thymus cortex, negative selection of these cells occur in the thymus medulla, mainly on the surface of dendritic cells that are descendants of cells that have migrated into the thymus from the bone marrow.

Some Organ-specific Proteins Are Ectopically Expressed in the Thymus Medulla

After the discovery of negative selection of developing T cells in the thymus, immunologists wondered how T cells avoid responses against self proteins that are not present in the thymus. One explanation is that some potentially self-reactive T cells are deleted or functionally inactivated after they leave the thymus. This occurs when the cells recognize self peptides bound to MHC proteins on the surface of dendritic cells that have not been activated by pathogens and therefore do not provide appropriate activating signals. It can also occur via regulatory T cells in the periphery that suppress the activity of some self-reactive effector T cells. These two mechanisms are examples of *peripheral tolerance*,

because, unlike T cell deletion in the thymus *(central tolerance)*, they occur after T cells leave the thymus (see Figure 25–13).

Recently, a third explanation has been discovered. A special class of epithelial cells in the thymus medulla ectopically expresses proteins previously thought to be expressed only outside the thymus in specific organs; insulin, for example, which is made by β cells in the pancreas, is also made by a small subset of medullary thymic epithelial cells. The ectopic expression of many of these proteins, including insulin, depends on a nuclear protein called *autimmune regulator (AIRE)*, which is specifically expressed in the same medullary thymic epithelial cells. Inactivation of the gene encoding AIRE in mice or humans results in a severe multi-organ autoimmune disease, indicating the importance of AIRE-dependent central tolerance to at least some "organ-specific" self proteins. It remains a mystery how AIRE promotes this ectopic expression of genes in the thymus medulla.

The Function of MHC Proteins Helps Explain Their Polymorphism

The role of MHC proteins in binding foreign peptides and presenting them to T cells helps explain the extensive polymorphism of these proteins. In the evolutionary war between pathogens and the adaptive immune system, pathogens tend to change their antigens to avoid associating with MHC proteins. When a pathogen succeeds, it is able to sweep through a population as an epidemic. In such circumstances, the few individuals that produce a new MHC protein that can associate with an antigen of the altered pathogen have a large selective advantage. In addition, individuals with two different alleles at any given MHC locus (heterozygotes) have a better chance of resisting infection than those with identical alleles at the locus, as they have a greater capacity to present peptides from a wide range of pathogens. Thus, this type of selection will tend to promote and maintain a large diversity of MHC proteins in the population. Support for the idea that infectious diseases have been a driving force for MHC polymorphism has come from studies in West Africa. In this region, individuals with a specific MHC allele have a reduced susceptibility to a severe form of malaria. Although the allele is rare elsewhere, it is found in 25% of the West African population where this form of malaria is common.

If greater MHC diversity means greater resistance to infection, why do we each have so few MHC genes encoding these molecules? Why have we not evolved strategies for increasing the diversity of MHC proteins—by alternative RNA splicing, for example, or by the genetic recombination mechanisms used to diversify antibodies and TCRs? One reason for the restricted diversity of MHC proteins in an individual may be that each time a new MHC protein is added to the repertoire, the T cells with TCRs that bind strongly to self peptides bound to the new MHC protein must be eliminated to maintain self tolerance. The elimination of these T cells would counteract the advantage of adding the new MHC protein. Thus, the number of MHC proteins we express is thought to represent a balance between the advantages of presenting a wide diversity of foreign peptides to T cells against the disadvantages of severely restricting the T cell repertoire during negative selection in the thymus. Computer modeling studies support this explanation.

Summary

There are three main functionally distinct classes of T cells. Cytotoxic T cells kill infected cells directly by inducing them to undergo apoptosis. Helper T cells help activate B cells to make antibody responses, cytotoxic T cells to kill their target cells, dendritic cells to stimulate T cell responses, and macrophages to destroy microorganisms that either invaded the macrophage or were ingested by it. Finally, regulatory T cells suppress the activity of effector T cells and dendritic cells and are crucial for self tolerance.

All T cells express cell-surface, antibodylike receptors (TCRs), which are encoded by genes that are assembled from multiple gene segments during T cell development

in the thymus. TCRs recognize fragments of foreign proteins that are displayed on the surface of host cells in association with MHC proteins. T cells are activated in peripheral lymphoid organs by antigen-presenting cells, which express peptide–MHC complexes, co-stimulatory proteins, and various cell–cell adhesion molecules on their cell surface. The most potent of these antigen-presenting cells are dendritic cells, which are specialized for antigen presentation and are required for the activation of naïve T cells.

Class I and class II MHC proteins have crucial roles in presenting foreign protein antigens to T cells: class I MHC proteins present antigens to cytotoxic T cells, while class II MHC proteins present antigens to helper and regulatory T cells. Whereas class I proteins are expressed on almost all vertebrate cells, class II proteins are normally restricted to those cell types that interact with helper T cells, such as dendritic cells, macrophages, and B lymphocytes.

Both classes of MHC proteins have a single peptide-binding groove, which binds small peptide fragments derived from proteins. Each MHC protein can bind a large set of peptides, which are constantly being produced intracellularly by normal protein degradation processes. However, class I MHC proteins mainly bind fragments produced in the cytosol, while class II MHC proteins mainly bind fragments produced in endocytic compartments. After they have formed inside the target cell, the peptide–MHC complexes are transported to the cell surface. Complexes that contain a peptide derived from a foreign protein are recognized by TCRs, which interact with both the peptide and the walls of the peptide-binding groove of the MHC protein. T cells also express CD4 or CD8 co-receptors, which simultaneously recognize non-polymorphic regions of MHC proteins on the antigen-presenting cell or target cell: helper cells and regulatory cells express CD4, which recognizes class II MHC proteins, while cytotoxic T cells express CD8, which recognizes class I MHC proteins.

A combination of positive and negative selection processes operates during T cell development in the thymus to shape the TCR repertoire. These processes help to ensure that only T cells with potentially useful receptors survive and mature, while all of the others die by apoptosis. First, T cells that can respond to peptides complexed with self MHC proteins are positively selected; subsequently, the T cells in this group that can react strongly with self peptides complexed with self MHC proteins are eliminated. Helper and cytotoxic T cells that leave the thymus with receptors that could react with self antigens are eliminated, functionally inactivated, or actively suppressed when they recognize self antigens on nonactivated dendritic cells.

HELPER T CELLS AND LYMPHOCYTE ACTIVATION

Helper T cells are arguably the most important cells in adaptive immunity, as they are required for almost all adaptive immune responses. They help activate B cells to secrete antibodies, and they help macrophages to destroy ingested pathogens. They also help activate cytotoxic T cells to kill infected target cells, at least partly by stimulating dendritic cells to activate naïve cytotoxic cells more efficiently. As dramatically demonstrated in AIDS patients, without helper T cells we cannot even defend ourselves against many microbes that are normally harmless.

Helper T cells themselves, however, can function only when activated to become effector cells. Naïve helper cells are activated on the surface of dendritic cells, which become activated during the innate immune responses triggered by an infection. The innate responses, mainly via activated dendritic cells, also dictate what kind of effector cell a helper T cell will develop into, and they thereby determine the nature of the adaptive immune responses elicited.

In this final section, we discuss the multiple signals that help activate a T cell and how a helper T cell, once activated to become an effector cell, helps activate other cells. We also consider how innate immune responses determine the nature of adaptive responses by stimulating helper T cells to differentiate into different types of effector cells. Finally, we discuss the probable evolutionary origins of the Ig superfamily of proteins, which includes the MHC proteins, antibodies, and TCRs.

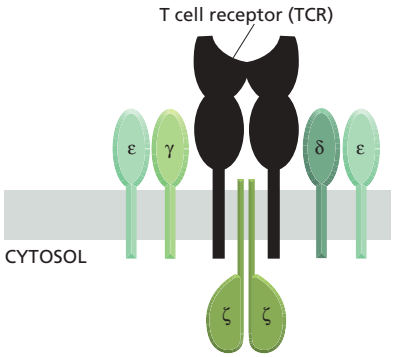


Figure 25–65 The TCR and its associated CD3 complex. All of the CD3 polypeptide chains (shown in *green*), except for the ζ (zeta) chains, have extracellular Ig-like domains and are therefore members of the Ig superfamily. All of the four types of CD3 polypeptide chains form heterodimers or homodimers (as shown) and are rapidly phosphorylated on tyrosines in their intracellular domains following TCR activation (not shown). Some of these phosphorylated tyrosines then serve as docking sites for intracellular signaling proteins, as shown in Figure 25–66.

Activated Dendritic Cells Use Multiple Mechanisms to Activate T Cells

When a dendritic cell is activated during an infection, it changes its shape and migratory behavior, increases the amount of MHC proteins displayed on its surface, activates its antigen-processing pathways, and starts producing both cell-surface-bound co-stimulatory proteins and secreted cytokines (including chemokines). The dramatic changes also enable the dendritic cell to migrate to a peripheral lymphoid organ and activate naïve T cells to become effector cells.

The dendritic cell initially signals to the T cell through the T cell receptors (TCRs), which bind to a foreign peptide complexed with a class II MHC protein on an opposing dendritic cell surface. The TCR, however, does not act on its own to transmit the signal into the T cell. It is helped by a complex of invariant transmembrane proteins called **CD3**, with which the TCR is associated (**Figure 25–65**). Moreover, the CD4 co-receptor on a helper or regulatory T cell and the CD8 co-receptor on a cytotoxic T cell bind to the same MHC protein as the TCR and also play a crucial part in transmitting the signal, as illustrated in **Figure 25–66**.

In addition to signaling through the TCR and its associated proteins and co-receptors, co-stimulatory proteins on the dendritic cell surface bind to other receptors on the T cell surface to provide further signals required to activate the T cell. Among these co-stimulatory proteins on the activated dendritic cell are the **B7 proteins**, which are recognized by the co-receptor protein **CD28** on the surface of the T cell. Once activated, a helper T cell itself expresses a co-stimulatory protein called *CD40 ligand*, which acts back on *CD40 receptors* on the dendritic cell surface to increase and sustain the activation of the dendritic cell, creating a positive feedback loop that amplifies the T cell response.

Once bound to the surface of a dendritic cell, a T cell increases the strength of the binding by activating an integrin adhesion protein, which then binds

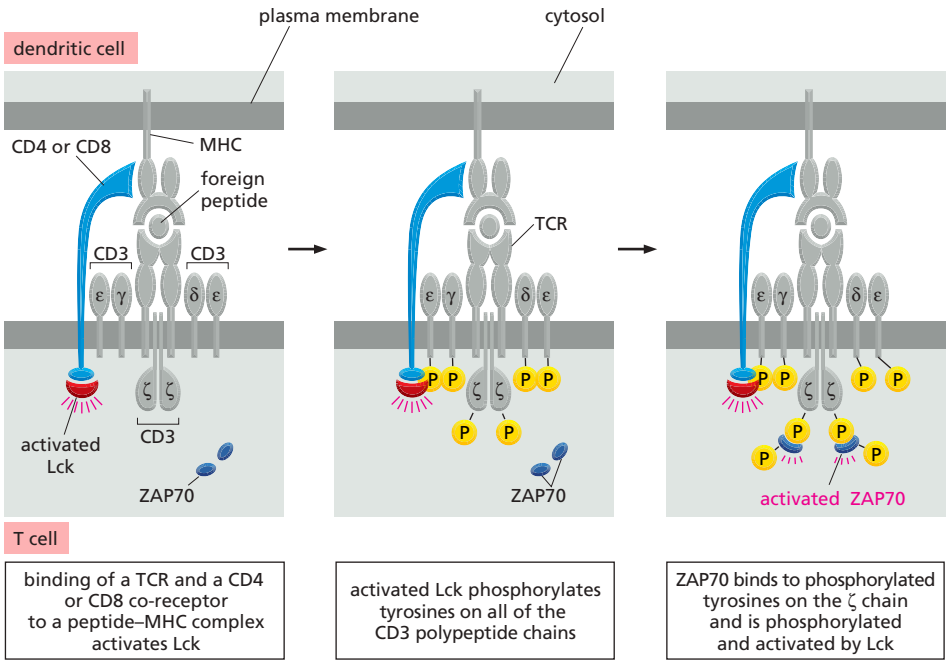


Figure 25–66 Signaling events initiated by the binding of peptide–MHC complexes to TCRs. When TCRs (and CD3) are clustered by binding to peptide–MHC complexes on an activated dendritic cell, CD4 molecules on helper cells or CD8 molecules on cytotoxic T cells are clustered with them, binding to invariant parts of the same class II or class I MHC proteins, respectively, on the dendritic cell. This brings the Src-like cytoplasmic tyrosine kinase Lck into the signaling complex and activates it. Lck activation also depends on a transmembrane protein tyrosine phosphatase on the T cell surface called CD45, which removes inhibitory phosphates from Lck (not shown).

Once activated, Lck initiates a tyrosine phosphorylation cascade by phosphorylating tyrosines on all of the chains of the CD3 complex. The phosphotyrosines on the CD3 ζ chain now serve as docking sites for yet another cytoplasmic tyrosine kinase called *ZAP70*. Lck phosphorylates, and thereby activates, ZAP70. Although not shown, ZAP70 then phosphorylates tyrosines on the tail of another transmembrane protein (called LAT), which then serve as docking sites for a variety of adaptor proteins and enzymes. These proteins then help relay the signal to the nucleus and other parts of the cell by activating the inositol phospholipid and MAP kinase signaling pathways (discussed in Chapter 15), as well as a Rho family GTPase that regulates the actin cytoskeleton (discussed in Chapter 16).

more strongly to its Ig-like ligand on the surface of the dendritic cell. This increased adhesion enables the T cell to remain bound to the antigen-presenting cell long enough to become activated.

This initial signaling through the TCR and associated proteins triggers the active assembly of an **immunological synapse** at the interface between the T cell and the dendritic cell. In these bull's-eye-like structures, TCRs and their associated CD3 subunits, co-receptors, and intracellular signaling proteins are clustered at the center, with cell–cell adhesion proteins forming a peripheral ring. Similar structures form when an effector helper or cytotoxic cell interacts with its target cell. Not all of the TCRs in the synapse are bound to a foreign peptide complexed with an MHC protein; some are bound to a self peptide bound to an MHC protein, and these TCRs also contribute to the activation of the T cell (recall that all T cells are initially positively selected in the thymus for their weak recognition of such self-peptide–MHC complexes).

The combined actions of the various signals just discussed stimulate helper T cells to proliferate and to begin to differentiate into effector cells by a curiously indirect mechanism. The signals cause the T cells to help stimulate their own proliferation and differentiation by inducing the cells to secrete a cytokine called **interleukin-2 (IL2)** and simultaneously to synthesize high-affinity cell-surface receptors that bind it. The binding of IL2 to the IL2 receptors activates intracellular signaling pathways that turn on genes that help the T cells to proliferate and differentiate into effector cells (**Figure 25–67**). Although some T cells do not make IL2, as long as they have been activated by their antigen and therefore express IL2 receptors they can be helped to proliferate and differentiate by IL2 made by neighboring activated T cells. IL2 also plays an important part in the development of regulatory T cells in the thymus, because without it these cells fail to develop.

Dendritic cells are not only important for activating T cells, they are also important for inactivating or eliminating self-reactive T cells. When T cells recognize self-peptide–MHC complexes on the surface of dendritic cells that have not been activated by a pathogen, they are either inactivated, so that they do not respond to the same peptide–MHC complexes even on activated dendritic cells, or they proliferate briefly and then die by apoptosis. Both of these mechanisms—clonal inactivation and clonal deletion—contribute to peripheral self tolerance. Dendritic cells also contribute to peripheral self tolerance by activating regulatory T cells, which then suppress the activity of self-reactive effector T cells, although the details of how dendritic cells selectively activate regulatory T cells are poorly understood (see Figure 25–13).

The Activation of T Cells Is Controlled by Negative Feedback

During the multi-step activation of a T cell, the cell starts to express a cell-surface protein called *CTLA4*, which inhibits intracellular signaling. This protein resembles CD28, and, like CD28, it binds to B7 proteins on the surface of the activating dendritic cell (see Figure 25–67). CTLA4 binds B7 with much higher affinity than does CD28 and, in doing so, it blocks the activating activity of CD28,

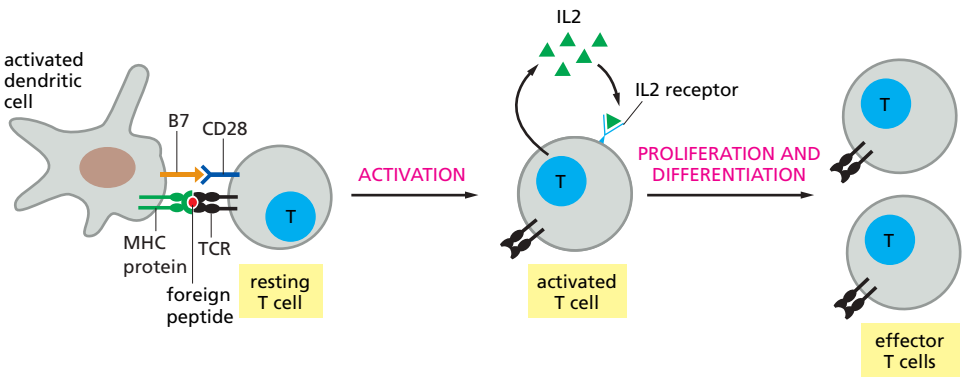


Figure 25–67 The stimulation of T cells by IL2. This model can apply to both helper and cytotoxic T cells, at least in culture. The combination of peptide–MHC complexes and a co-stimulatory B7 protein (either B7-1 or B7-2; also called CD80 and CD86, respectively) on the surface of an activated dendritic cell helps stimulate a resting T cell to make high affinity IL2 receptors and to secrete IL2. The binding of IL2 to its receptors then helps stimulate the T cells to proliferate and differentiate into effector cells. The various proteins associated with the TCRs (see Figure 25–65) are not shown.

Table 25–3 Some Accessory Proteins on the Surface of T Cells

PROTEIN*	SUPERFAMILY	EXPRESSED ON	LIGAND ON TARGET CELL	FUNCTIONS
CD3 complex	Ig (except for ζ)	all T cells	—	helps transduce signal when antigen–MHC complexes bind to TCRs; helps transport TCRs to cell surface
CD4	Ig	helper T cells, regulatory T cells	class II MHC	promotes adhesion to dendritic cells and target cells; signals T cell
CD8	Ig	cytotoxic T cells	class I MHC	promotes adhesion to dendritic cells and infected target cells; signals T cell
CD28	Ig	most T cells	B7 proteins (CD80 and CD86)	helps activate T cells
CTLA4	Ig	activated T cells	B7 proteins (CD80 and CD86)	inhibits T cell activation
CD40 ligand	Fas ligand family	effector helper T cells	CD40	co-stimulatory protein that helps activate macrophages, B cells, and dendritic cells

*CD stands for cluster of differentiation, as each of the CD proteins was originally defined as a blood cell "differentiation antigen" recognized by multiple monoclonal antibodies. Their identification depended on large-scale collaborative studies in which hundreds of such antibodies, generated in many laboratories, were compared and found to consist of relatively few groups (or "clusters"), each recognizing a single cell-surface protein. Since these initial studies, however, more than 240 CD proteins have been identified.

thereby providing negative feedback that holds the activation process in check. Thus, mice with a disrupted *Ctla4* gene die from a massive accumulation of activated T cells.

Table 25–3 summarizes some of the co-receptors and other accessory proteins found on the surface of T cells that we have discussed in this chapter.

Most of the T (and B) effector cells produced during an immune response must be eliminated after they have done their job. Although most of the cells die by apoptosis, the extracellular mechanisms responsible for their elimination are poorly understood. One possibility is that, as antigen levels fall and the response subsides, effector T cells are deprived of the antigen and cytokine stimulation that they need to survive, so that only memory cells and some long-lived effector cells survive. The death of effector T cells, however, does not occur solely through a lack of survival signals. In the case of effector cytotoxic T cells, for example, the cytokine *interferon-γ* (*IFNγ*) plays an important part in inducing the cell death; as effector cytotoxic T cells make IFNγ (see Figure 25–60), this is another form of negative feedback.

Before considering how effector helper T cells help activate macrophages and B cells, we need to discuss the two main functionally distinct subclasses of effector helper T cells, T_H1 and T_H2 cells, and how they are generated.

The Subclass of Effector Helper T Cell Determines the Nature of the Adaptive Immune Response

When an activated dendritic cell activates a naïve helper T cell in a peripheral lymphoid organ, the T cell usually differentiates into a T_H1 or T_H2 effector helper cell. The outcome depends on the affinity of the TCR for the peptide–MHC complex, on the density of the complex on the dendritic cell surface, and on the nature of the dendritic cell.

The two main subclasses of effector helper T cells can be distinguished by the cytokines they secrete. **T_H1 cells** mainly secrete *IFNγ* and *tumor necrosis factor-α (TNFα)*, which activate macrophages to kill microbes located within the macrophages' phagosomes. They also help activate cytotoxic T cells to kill infected cells. Thus, T_H1 cells mainly defend an animal against intracellular microbes. They also, however, stimulate B cells to secrete specific subclasses of IgG antibodies that can coat extracellular microbes and activate complement, thereby helping to eliminate some extracellular microbes as well.

T_H2 cells, by contrast, mainly defend an animal against extracellular pathogens, including microbes and multicellular parasites. They secrete a variety

of cytokines, including *interleukins 4 and 10 (IL4* and *IL10),* and they help stimulate B cells to make most classes of antibodies, including IgM, IgA, IgE, and some subclasses of IgG. Some of these antibodies bind to mast cells, basophils, and eosinophils; when activated by antigen binding, these cells release local mediators that cause sneezing, coughing, or diarrhea and thereby help expel extracellular microbes and larger parasites from epithelial surfaces of the body.

Thus, the decision of naïve helper T cells to differentiate into T_H1 or T_H2 effector cells influences the type of adaptive immune response that an animal mounts against the pathogen—whether it will be dominated by macrophage activation or by antibody production. The specific cytokines produced during the process of helper T cell activation have an important influence on the type of effector helper cell produced. Some intracellular bacteria, for example, stimulate dendritic cells to produce *IL12,* which induces T_H1 development, and thereby macrophage activation. As expected, mice that are deficient in either IL12 or its receptor are much more susceptible to these bacterial infections than are normal mice. Many parasitic protozoa and worms, by contrast, stimulate a dendritic cell to express the *Jagged protein* on its surface. Jagged is an activating ligand for *Notch receptors* (discussed in Chapters 15 and 22) on the T cell surface, and the resulting Notch signaling helps induce T_H2 development and the production of IL4. The T_H2 cells and IL4 help stimulate antibody production and eosinophil activation, leading to parasite expulsion (**Figure 25–68**). The IL4 also generates a positive feedback loop, as it is a potent inducer of T_H2 development.

Once a T_H1 or T_H2 effector cell develops, it inhibits the differentiation of the other type of helper T cell. IFNγ produced by T_H1 cells inhibits the development of T_H2 cells, while IL4 and IL10 produced by T_H2 cells inhibit the development of T_H1 cells. Thus, as the response proceeds, it reinforces the initial choice through its effect on the response of other T cells nearby.

Individuals infected with *Mycobacterium leprae,* the bacterium that causes leprosy, illustrate the importance of the T_H1/T_H2 decision. This bacterium replicates mainly within macrophages and causes either of two forms of disease, depending mainly on the genetic make-up of the infected individual. In some patients, the *tuberculoid* form of the disease occurs. Here, T_H1 cells develop and stimulate the infected macrophages to kill the bacteria. This produces a local inflammatory response, which damages skin and nerves. The result is a chronic

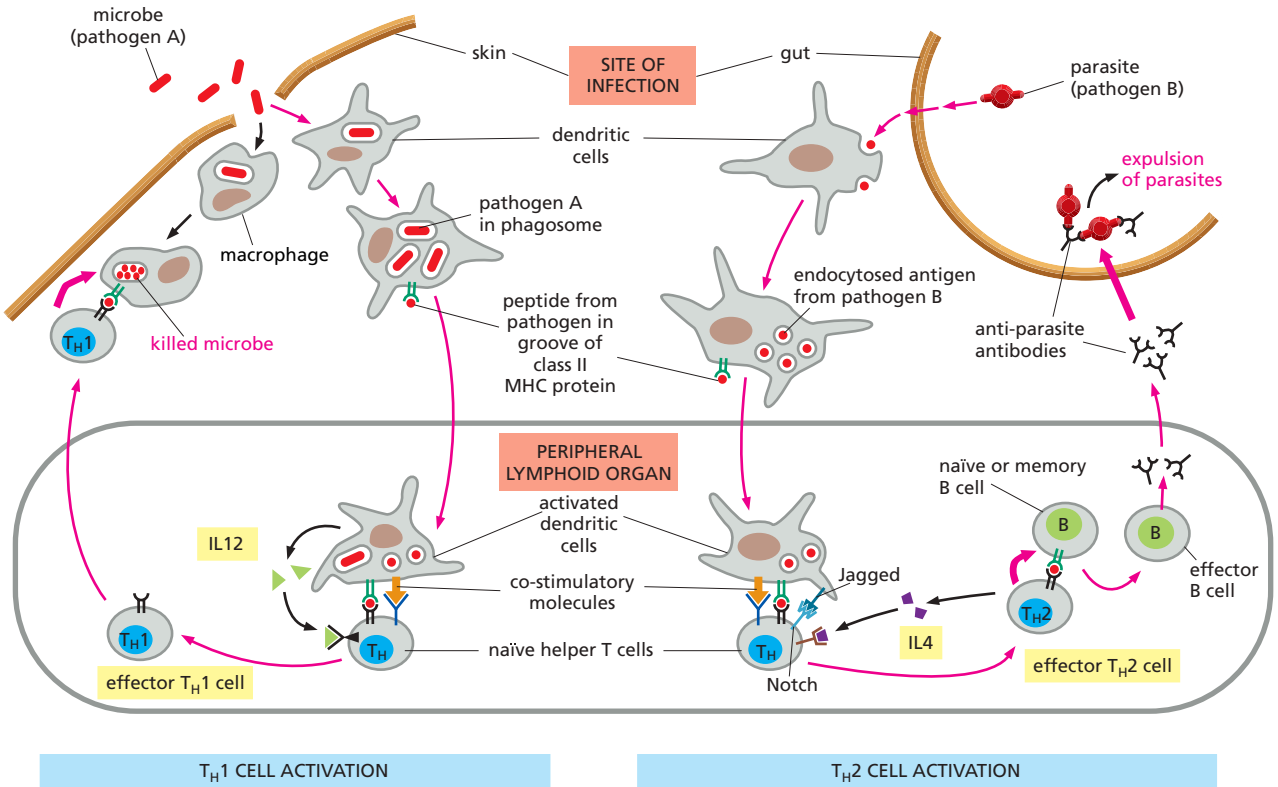


Figure 25–68 The activation of T_H1 and T_H2 cells. The differentiation of naïve helper T cells into either T_H1 or T_H2 effector cells determines the nature of the subsequent adaptive immune responses. Whether a naïve helper T cell becomes a T_H1 or T_H2 cell depends mainly on the signal proteins present when an activated dendritic cell in a peripheral lymphoid organ stimulates the helper T cell. The types of signal proteins produced depend on the local environment and the nature of the pathogen that activated the dendritic cell at the site of infection. IL12 secreted by activated dendritic cells promotes T_H1 cell development. By contrast, both the transmembrane Notch ligand Jagged on the surface of activated dendritic cells and the IL4 made by basophils, mast cells, and T_H2 cells promote T_H2 cell development.

In this figure, the effector T_H1 cell produced in the peripheral lymphoid organ migrates to the site of infection and helps a macrophage kill the microbes it has phagocytosed. The effector T_H2 cell remains in the lymphoid organ and helps activate a B cell to produce antibodies against the parasite. In addition to binding to the parasites, the antibodies bind to Fc receptors on mast cells, basophils, and eosinophils (see Figure 25–27), which then can help expel the parasite from the gut. Although not shown here, T_H1 cells also help activate B cells to make antibodies.

disease that progresses slowly but does not kill the host. In other patients, by contrast, the *lepromatous* form of the disease occurs. Here, T_H2 cells develop and stimulate the production of antibodies. As the antibodies cannot get through the plasma membrane to attack the intracellular bacteria, the bacteria proliferate unchecked and eventually kill the host. For unknown reasons, there is also a general depression of T-cell-mediated immunity to most antigens in the lepromatous form of the disease.

Naïve helper T cells can also develop into a recently described third type of effector cell called *T_H17 cells* because they secrete the pro-inflammatory interleukin *IL17*. T_H17 cells help defend against extracellular pathogens, but they also play an important part in many autoimmune diseases. They develop when some naïve helper T cells are activated by antigen in the presence of TGFβ and the interleukin IL6.

T_H1 Cells Activate Infected Macrophages and Stimulate An Inflammatory Response

Macrophages and dendritic cells ingest pathogens and their products at sites of infection. The dendritic cells become activated and carry microbial antigens to a peripheral lymphoid organ, where they preferentially induce the development of T_H1 cells. The T_H1 cells then migrate to the site of infection to help activate infected macrophages (see Figure 25–68).

T_H1 effector cells use two signals to activate the specific macrophages they recognize. They secrete IFNγ, which binds to IFNγ receptors on the macrophage surface, and they display the co-stimulatory protein CD40 ligand, which binds to CD40 on the macrophage (**Figure 25–69**). (We see later that CD40 ligand is also used by helper T cells to activate B cells.) Once activated, the macrophage can kill the microbes in their phagosomes: lysosomes can now fuse more readily with the phagosomes, unleashing a hydrolytic attack, and the activated macrophage makes oxygen radicals and nitric oxide, both of which are highly toxic to the microbes (discussed in Chapter 24). Because dendritic cells also express CD40, the T_H1 cells at sites of infection can also help activate them. As a result, the dendritic cells increase their production of class II MHC proteins, co-stimulatory proteins, and various cytokines, especially IL12. This makes them more effective at stimulating naïve helper T cells to differentiate into T_H1 effector cells in peripheral lymphoid organs, providing a positive feedback loop that increases the production of T_H1 cells and, thereby, the activation of macrophages.

T_H1 effector cells also stimulate an *inflammatory response* (discussed in Chapter 24) by recruiting more phagocytic cells into the infected site. They do so in three ways:

1. They secrete cytokines that act on the bone marrow to increase the production of monocytes (macrophage precursors that circulate in the blood) and neutrophils.
2. They secrete other cytokines that activate endothelial cells lining local blood vessels to express cell adhesion molecules that cause monocytes and neutrophils in the blood to adhere there.
3. They secrete chemokines that direct the migration of the adherent monocytes and neutrophils out of the bloodstream into the site of infection.

A T_H1 cell can also help activate a cytotoxic T cell in a peripheral lymphoid organ by producing chemokines that attract the cytotoxic cells to the site of the T_H1-cell–dendritic-cell interaction, while at the same time stimulating the dendritic cell to produce more co-stimulatory proteins. In addition, T_H1 cells can help effector cytotoxic T cells kill virus-infected target cells by secreting IFNγ, which increases the efficiency with which target cells process viral antigens for presentation to cytotoxic T cells (see Figure 25–60). An effector T_H1 cell can also directly kill some cells itself, including effector lymphocytes: by expressing *Fas ligand* on its surface, it can induce effector T or B cells that express cell-surface *Fas* to undergo apoptosis (see 25–47B).

Both T_H1 and T_H2 cells can also help stimulate B cells to proliferate and differentiate and to switch the class of antibody they make, from IgM and IgD to

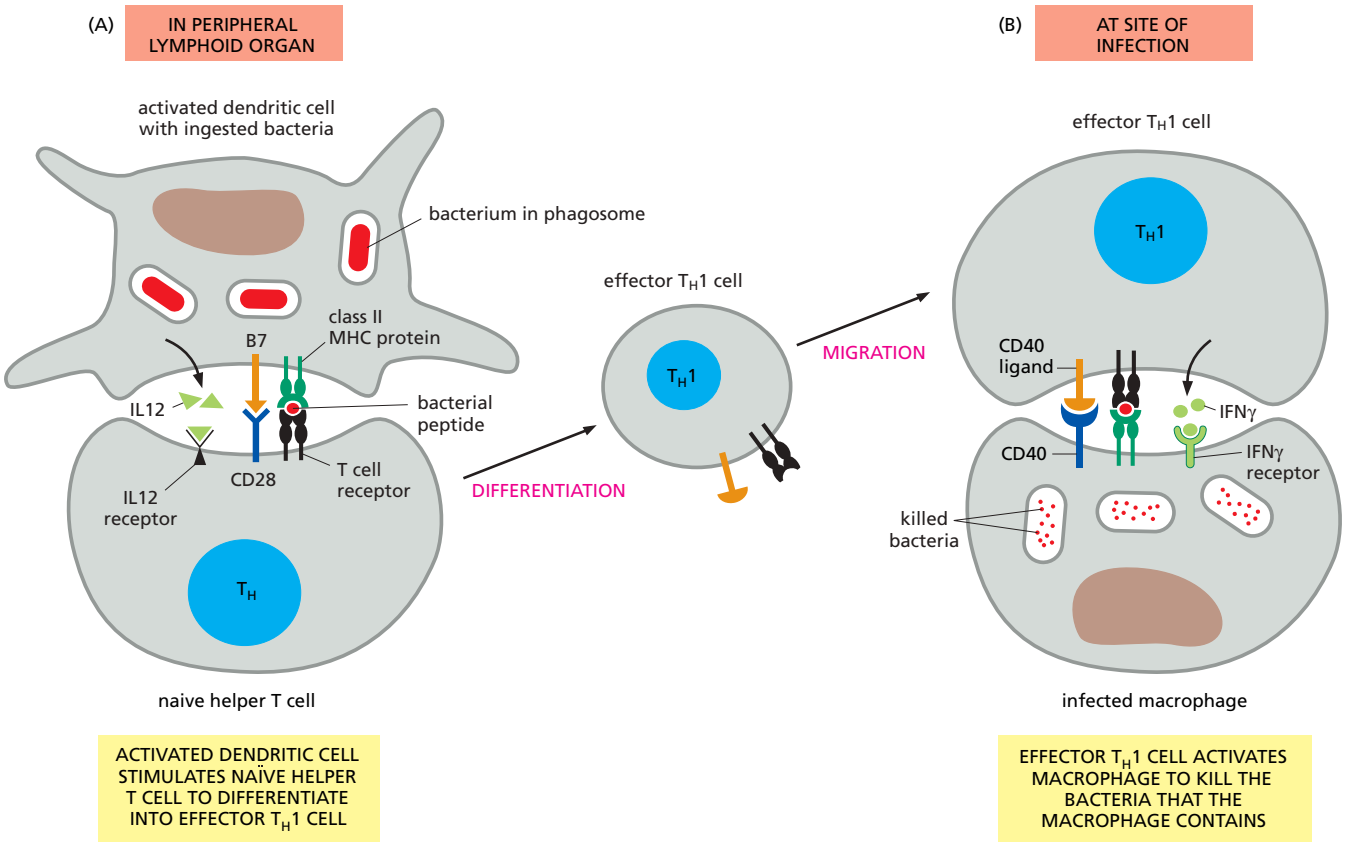


Figure 25–69 The differentiation of T_H1 cells and their activation of macrophages. (A) An activated dendritic cell that has ingested bacteria at a site of infection and migrated to a peripheral lymphoid organ activates a naïve helper T cell to differentiate into a T_H1 effector cell. The dendritic cell uses both cell-surface co-stimulatory proteins such as the B7 proteins and secreted IL12 to induce T_H1 cell differentiation. (B) The activated T_H1 effector cell then migrates from the peripheral lymphoid organ to the infected site, where it helps activate macrophages to kill the bacteria harbored within the macrophages' phagosomes. As indicated, it does this by means of secreted IFNγ and membrane-bound CD40 ligand, which binds to CD40 on the macrophage.

one of the secondary classes of antibody. Before considering how helper T cells do this, we need to discuss the role of the B cell antigen receptor in the activation of B cells.

Antigen Binding to B Cell Receptors (BCRs) Is Only One Step in B Cell Activation

Like T cells, most B cells require multiple extracellular signals to become activated. One signal is provided by antigen binding to the **B cell receptor** (**BCR**) for antigen, which, as discussed previously, is a membrane-bound antibody molecule. A helper T cell usually provides the other required signals. If a B cell receives the first signal only, it may be eliminated or functionally inactivated, which is one way in which B cells become tolerant to self antigens.

Signaling through the BCR works in much the same way as signaling through the TCR (see Figure 25–66). The receptor is associated with two invariant protein chains, Igα and Igβ, which help convert antigen binding to BCRs into intracellular signals. When antigen cross-links BCRs on the surface of a B cell, it causes them and their associated invariant chains to cluster into small aggregates. This aggregation leads to the assembly of an intracellular signaling complex at the site of the clustered receptors and to the initiation of a tyrosine phosphorylation cascade (**Figure 25–70**).

A co-receptor complex that binds complement proteins greatly enhances the efficiency of signaling through the BCR and its associated invariant chains.

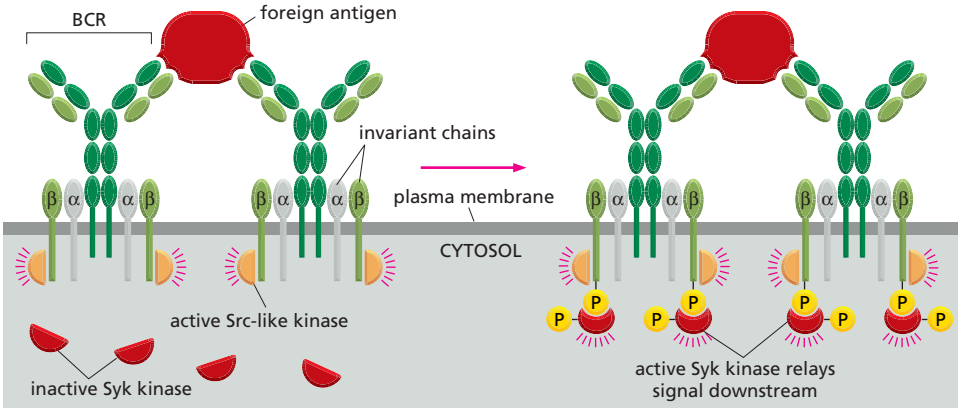


Figure 25–70 Early signaling events in B cells activated by the binding of antigen to BCRs. The antigen cross-links adjacent BCRs, which are transmembrane antibody molecules. The cross-linking causes the BCRs and their associated invariant chains (heterodimers of Igα and Igβ) to cluster. A Src-like cytoplasmic tyrosine kinase, which can be *Fyn*, *Blk*, or *Lyn*, is associated with the cytosolic tail of Igβ. It joins the cluster and phosphorylates Igα and Igβ (for simplicity, only the phosphorylation on Igβ is shown). As in the case of TCR activation, the protein tyrosine phosphatase CD45 is also required to activate these Src-like kinases (not shown). The resulting phosphotyrosines on Igα and Igβ serve as docking sites for another Src-like tyrosine kinase called *Syk*, which is homologous to ZAP70 in T cells (see Figure 25–66). Like ZAP70, Syk becomes phosphorylated and thereby activated, and it then relays the signal downstream.

If a microbe directly activates the complement system (discussed in Chapter 24), complement proteins are often deposited on the microbe surface, greatly increasing the B cell response to the microbe. Now, when the microbe clusters BCRs on a B cell, the *complement-binding co-receptor complexes* are brought into the cluster, increasing the strength of signaling by activating PI 3-kinase (discussed in Chapter 15) (**Figure 25–71**A). As expected, antibody responses are greatly reduced in mice lacking either one of the required complement components or one of the complement-binding co-receptor subunits on B cells.

Later in the immune response, by contrast, when IgG antibodies are present on the surface of the microbe, a different co-receptor dampens down the B cell response. These are *Fc receptors*, which bind the tails of the IgG antibodies. They recruit both lipid and protein phosphatase enzymes into the signaling complex that decrease the strength of signaling (Figure 25–71B). In this way, the Fc receptors on B cells act as inhibitory co-receptors, just as the CTLA4 proteins do on T cells. Thus, the co-receptors on a T cell or B cell allow the cell to gain additional information about the antigen bound to its receptors and thereby make a more informed decision as to how to respond.

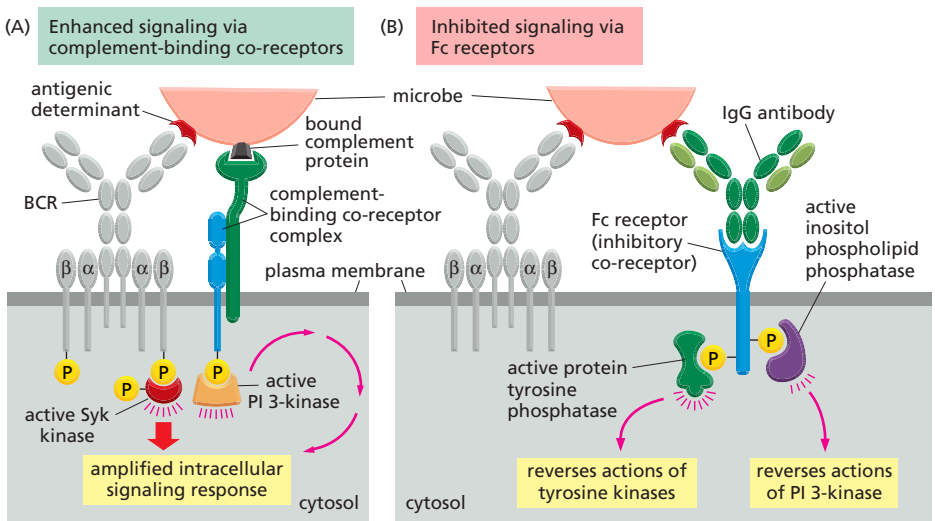


Figure 25–71 The influence of B cell co-receptors on the effectiveness of signaling through BCRs. (A) The binding of microbe–complement complexes to BCRs cross-links the BCRs to complement-binding, co-receptor complexes. The cytosolic tail of one component of the co-receptor complex becomes phosphorylated on tyrosines, which then serve as docking sites for PI 3-kinase. As discussed in Chapter 15, PI 3-kinase is activated to phosphorylate specific inositol phospholipids in the plasma membrane, which then act as docking sites to recruit intracellular signaling proteins (not shown). These signaling proteins act together with the signals generated by the Syk kinase to amplify the response. (B) When IgG antibodies bind to foreign antigen, usually late in a response, the Fc regions of the antibodies bind to Fc receptors on the B cell surface and are thus recruited into the signaling complex. The Fc receptors become phosphorylated on tyrosines; these then serve as docking sites for two types of phosphatase enzymes: (1) an inositol phospholipid phosphatase, which dephosphorylates the inositol phospholipid docking sites in the plasma membrane generated by PI 3-kinase, thereby reversing the activating effects of PI 3-kinase; (2) protein tyrosine phosphatases, which inhibit signaling by the activated tyrosine kinases.

Antigen-Specific Helper T Cells Are Essential for Activating Most B Cells

Whereas antigen-presenting cells such as dendritic cells and macrophages are omnivorous and ingest and present antigens on their MHC proteins nonspecifically, a B cell generally presents only peptides derived from an antigen that it specifically recognizes using its BCRs. Thus, BCRs do more than just bind antigen to begin the process of B cell activation; they also play a crucial part in recruiting T cell help. Through endocytosis, they deliver their bound protein antigen to an endosomal compartment, where the antigen is degraded into peptides; many of these peptides are returned to the B cell surface bound to class II MHC proteins (see Figure 25–61). These peptide–class-II-MHC complexes are recognized by antigen-specific helper T cells, which then deliver further signals to the B cell that are required for its proliferation and antibody secretion (**Figure 25–72**).

How do the antigen-specific T cells required for B cell activation originate? As discussed previously, during primary antibody responses, naïve helper T cells are activated in a peripheral lymphoid organ by binding to a foreign peptide bound to a class II MHC protein on the surface of an activated dendritic cell. The effector helper T cells that results from this activation can then activate a B cell that displays the same complex of foreign peptide and class II MHC protein on its surface. Thus, the helper T cell activates only those B cells with BCRs that specifically recognize the antigen that initially activated the T cell, even though the TCRs and BCRs usually recognize distinct antigenic determinants on the antigen (see Figure 25–72). This requirement for *linked recognition* of antigen by a T cell and a B cell helps avoid autoimmune B cell responses, which would require the simultaneous presence of both helper T cells and B cells that recognize the same self antigen.

In secondary antibody responses, memory B cells themselves can act as antigen-presenting cells and activate helper T cells, as well as being the subsequent targets of the effector helper T cells. The mutually reinforcing actions of helper T cells and B cells lead to an antibody response that is both intense and highly specific.

Once a helper T cell has been activated to become an effector cell and contacts a B cell, the contact initiates an internal rearrangement of the helper cell's cytoplasm. The T cell orients its centrosome and Golgi apparatus toward the B cell, as described previously for an effector cytotoxic T cell contacting its target cell (see Figure 25–46). In this case, however, the orientation is thought to enable the effector helper T cell to direct both membrane-bound and secreted cytokines onto the B cell surface (see Figure 25–72). One crucial membrane-bound signal molecule is CD40 ligand, which we encountered earlier. It is expressed on the surface of effector helper T cells, but not on nonactivated naïve or memory helper T cells, and it is recognized by the CD40 protein on the B cell surface. This interaction between CD40 ligand and CD40 is required for helper T cells to activate B cells to proliferate and differentiate into memory or antibody-

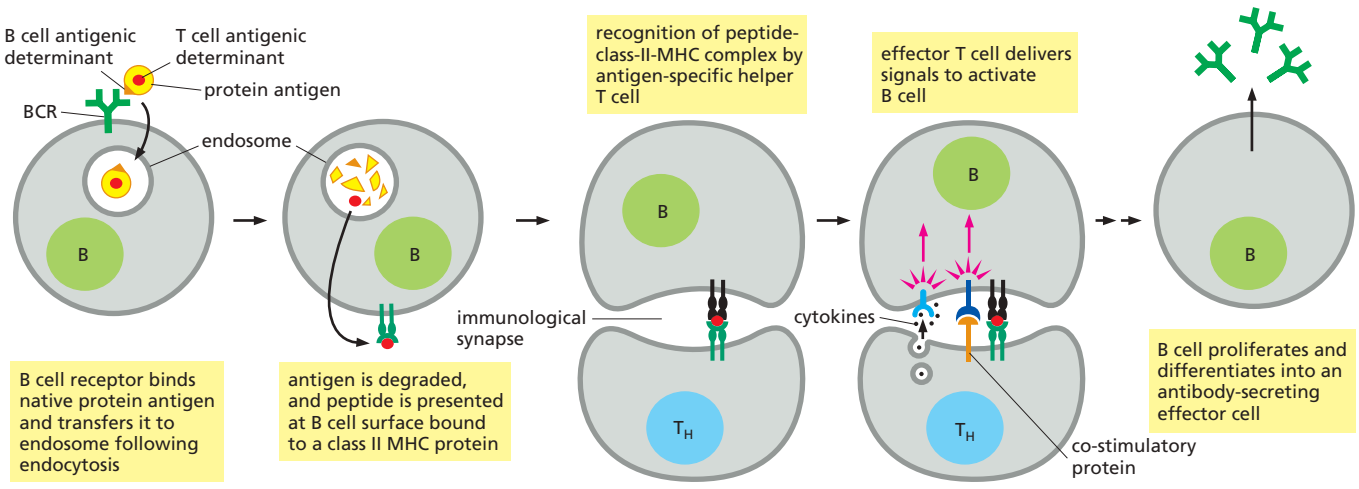


Figure 25–72 The activation of a B cell by a protein antigen and an effector helper T cell. Note that the B cell and T cells recognize different antigenic determinants on the antigen and that the effector helper T cell uses both secreted and membrane-bound co-stimulatory molecules to help activate the B cell.

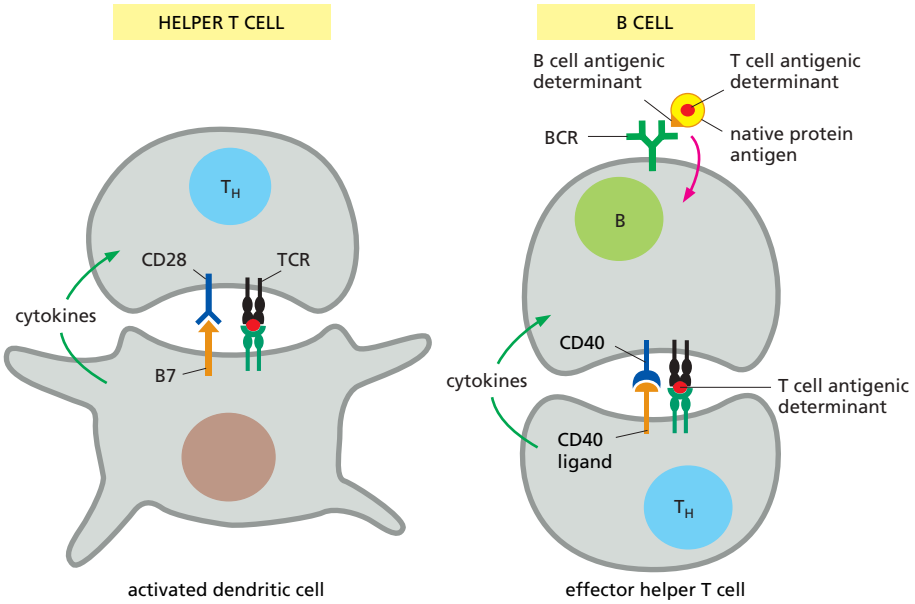


Figure 25–73 Comparison of the signals required to activate a helper T cell and a B cell to the same protein antigen. Note that in both cases secreted and membrane-bound molecules cooperate in the activation process. The *red arrow* indicates the endocytosis of the protein antigen. Although not shown, CD40 ligand is also used by effector helper T cells to increase and maintain the activation of dendritic cells, which express CD40, thereby creating a positive feedback loop.

The antigenic determinant recognized by the helper T cell is presented on the surface of both the dendritic cell and the B cell as a peptide fragment of the protein antigen bound to a class II MHC protein. By contrast, the B cell recognizes a different antigenic determinant on the surface of the folded (native) protein.

secreting effector cells. Individuals that lack CD40 ligand are severely immunodeficient. They are susceptible to the same infections that affect AIDS patients whose helper T cells have been destroyed.

Helper T cells also secrete cytokines to help B cells proliferate and differentiate and, in some cases, to switch the class of antibody they produce. The cytokines include the interleukins IL2 and IL4. IL4, for example, is produced by T_H2 cells and collaborates with CD40 ligand in stimulating B cell proliferation and differentiation; it also promotes switching to IgE antibody production. Mice deficient in IL4 production are severely impaired in their ability to make IgE.

Figure 25–73 compares the signals required for T and B cell activation; **Table 25–4** lists some of the cytokines discussed in this chapter.

A Special Class of B Cells Recognize T-cell-independent Antigens

Some antigens can stimulate B cells to proliferate and differentiate into antibody-secreting effector cells without help from T cells. Most of these *T-cell-independent antigens* are microbial polysaccharides that do not activate helper T cells. Some activate B cells directly by providing both the antigen signal and the accessory signals normally provided by helper T cells. Others are large polymers with repeating, identical antigenic determinants (see Figure 25–29B); their multipoint binding to BCRs can generate a strong enough signal on its own to activate the B cell directly, without additional signals.

Table 25–4 Properties of Some Cytokines

CYTOKINE	SOME SOURCES	SOME ACTIONS
IL2	all helper T cells; some cytotoxic T cells	stimulates proliferation and differentiation of activated T cells; required for regulatory T cell development in thymus
IL4	T_H2 cells, basophils, and mast cells	stimulates B cell proliferation, differentiation, and class switching to IgE and IgG1; promotes T_H2 and inhibits T_H1 cell development
IL7	many non-T cells	promotes memory T cell survival
IL10	T_H2 cells, macrophages, and dendritic cells	inhibits macrophages and T_H1 cell development
IL12	B cells, macrophages, dendritic cells, and granulocytes	induces T_H2 cell development and inhibits T_H1 cell development
IL15	many non-T cells	promotes memory T cell survival
IL17	some effector helper T cells	stimulates inflammatory responses
IFNγ	T_H1 cells and cytotoxic T cells	activates macrophages; increases MHC expression in many cell types
TGFβ	regulatory T cells	suppresses effector T cell activity, dendritic cells, and macrophages
TNFα	T_H1 cells and macrophages	activates endothelial cells and macrophages

Because T-cell-independent antigens do not activate helper T cells, they fail to induce B cell memory, affinity maturation, or class switching, all of which require help from T cells. They therefore mainly stimulate the production of low-affinity (but high-avidity) IgM antibodies. Most B cells that make antibodies without T cell help belong to a distinct B cell lineage. They are called *B1 cells* to distinguish them from *B2 cells*, which require T cell help. B1 cells seem to be especially important in defense against intestinal pathogens.

Immune Recognition Molecules Belong to the Ancient Ig Superfamily

Most of the proteins that mediate cell–cell recognition or antigen recognition in the immune system contain Ig or Ig-like domains, suggesting that they have a common evolutionary history. Included in this **Ig superfamily** are antibodies, TCRs, MHC proteins, the CD4, CD8, and CD28 co-receptors, the B7 co-stimulatory proteins, and most of the invariant polypeptide chains associated with TCRs and BCRs, as well as the various Fc receptors on lymphocytes and other white blood cells. All of these proteins contain one or more Ig or Ig-like domains. In fact, about 15% of the 250 or so proteins that have been characterized on the surface of white blood cells belong to this superfamily. Many of these molecules are dimers or higher oligomers in which Ig or Ig-like domains of one chain interact with those in another (**Figure 25–74**).

A separate exon usually encodes the amino acids in each Ig-like domain. It seems likely that the entire gene superfamily evolved from a gene coding for a single Ig-like domain—similar to that encoding β_2-microglobulin (see Figures 25–50A and 25–52) or the Thy-1 protein (see Figure 25–74)—that may have mediated cell–cell interactions. There is evidence that such a primordial gene arose before vertebrates diverged from their invertebrate ancestors about 400 million years ago. New family members presumably arose by exon and gene duplications.

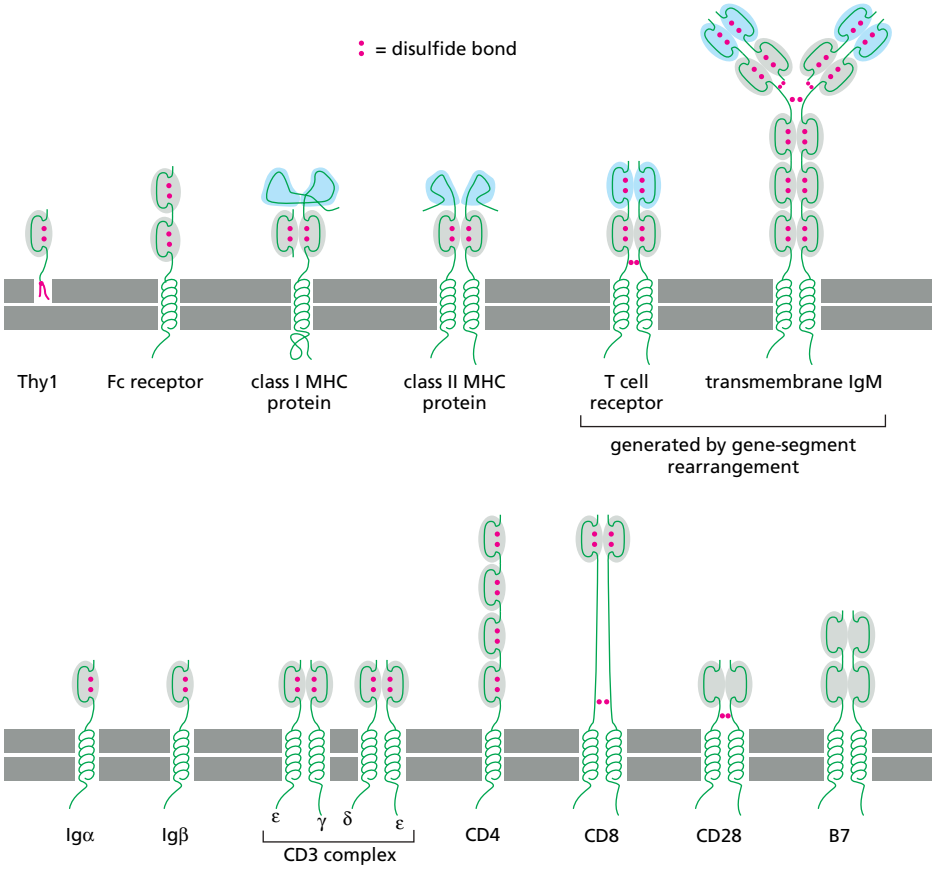


Figure 25–74 Some of the cell-surface proteins discussed in this chapter that belong to the Ig superfamily. The Ig and Ig-like domains are shaded in *gray,* except for the antigen-binding domains (not all of which are Ig domains), which are shaded in *blue*. The function of Thy1 is unknown, but it is held in the plasma membrane by a glycosylphosphatidylinositol (GPI) anchor and is widely used to identify T cells in mice. The Ig superfamily also includes many cell-surface proteins involved in cell–cell interactions outside the immune system, such as the neural cell adhesion molecule (NCAM) discussed in Chapter 19 and the receptors for various protein growth factors discussed in Chapter 15 (not shown). There are more than 750 members of the Ig superfamily in humans.

The multiple gene segments that encode antibodies and TCRs may have arisen when a transposable element, or transposon (discussed in Chapter 5), inserted into an exon of a gene encoding an Ig family member in an ancestral lymphocyte-like cell. The transposon may have contained the ancestors of the *Rag* genes, which, as discussed earlier, encode the proteins that initiate V(D)J recombination; the finding that the RAG proteins can act as transposons in a test tube strongly supports this view. Once the transposon had inserted into the exon, the gene could be expressed only if the transposon was excised by the RAG proteins and the two ends of the exon were rejoined, much as occurs when the *V* and *J* gene segments of an Ig light chain gene are assembled (see Figure 25–38). A second insertion of the transposon into the same exon may then have divided the gene into three segments, equivalent to the present-day *V, D,* and *J* gene segments. Subsequent duplication of either the individual gene segments or the entire split gene may then have generated the arrangements of gene segments that characterize the adaptive immune systems of present-day vertebrates.

Summary

The production of an effector helper T cell from of a naïve helper T cell requires multiple signals from an activated dendritic cell. MHC–peptide complexes on the dendritic cell surface provide one signal, by binding to both TCRs and a CD4 co-receptor on the T cell. Co-stimulatory proteins on the dendritic cell surface, including CD28, and secreted cytokines are the other signals. When naïve helper T cells are initially activated on a dendritic cell, most differentiate into either T_H1 or T_H2 effector cells, depending mainly on the signal proteins in their environment. T_H1 cells activate macrophages, cytotoxic T cells, and B cells, while T_H2 cells activate mainly B cells. In both cases, the effector helper T cells recognize the same complex of foreign peptide and class II MHC protein on the target cell surface as they initially recognized on the dendritic cell that activated them. They activate their target cells by a combination of membrane-bound and secreted co-stimulatory proteins. One membrane-bound signal protein used by both T_H1 and T_H2 cells is CD40 ligand.

Like T cells, B cells require multiple signals for activation. Antigen binding to the B cell antigen receptors (BCRs) provides one signal, while antigen-specific effector helper T cells provide the other signals. The requirement for multiple signals to activate either a T cell or a B cell helps to prevent inappropriate and dangerous activation of lymphocytes, including self-reactive lymphocytes.

Most of the proteins involved in cell–cell recognition and antigen recognition in the immune system, including antibodies, TCRs, and MHC proteins, as well as the various co-receptors discussed in this chapter, belong to the ancient Ig superfamily. This superfamily is thought to have evolved from a primordial gene encoding a single Ig-like domain. The mechanisms for diversifying antibodies and T cell receptors by recombining gene segments may have arisen when a transposon inserted into an exon of a gene encoding an Ig family member.

REFERENCES

General

Abbas AK, Lichtman AH & Pober JS (2005) Cellular and Molecular Immunology 3rd ed. Philadelphia: WB Saunders.

Janeway CA Jr, Travers P, Walport M & Shlomchik M (2005) Immunobiology: The Immune System in Health and Disease 6th ed. London: Garland Science.

Lymphocytes and the Cellular Basis of Adaptive Immunity

Ahmed R & Rousse BT (eds) (2006) Immunological memory. *Immunol Rev* 211:5–337.

Ebert LM, Schaerli P & Moser B (2005) Chemokine-mediated control of T cell traffic in lymphoid and peripheral tissues. *Mol Immunol* 42:799–809.

Goodnow CC, Sprent J, Fazekas de St Groth B & Vinuesa CG (2005) Cellular and genetic mechanisms of self tolerance and autoimmunity. *Nature* 435:590–597.

Gourley TS, Wherry EJ, Masopust D & Ahmed R (2004) Generation and maintenance of immunological memory. *Semin Immunol* 16:323–333.

Janeway CA Jr & Medzhitov R (2002) Innate immune recognition. *Annu Rev Immunol* 20:197–216.

Kyewski B & Klein L (2006) A central role for central tolerance. *Annu Rev Immunol* 24:571–606.

Singh NJ & Schwartz RH (2006) Primer: mechanisms of immunologic tolerance. *Nature Clin Pract Rheumatol* J2:44–52.

Steinman RM & Hemmi H (2006) Dendritic cells: translating innate to adaptive immunity. *Curr Top Microbiol Immunol* 311:17–58.

Wiedle G, Dunon D & Imhof BA (2001) Current concepts in lymphocyte homing and recirculation. *Crit Rev Clin Lab Sci* 38:1–31.

B Cells and Antibodies

Burton DR & Woof JM (1992) Human antibody effector function. *Adv Immunol* 51:1–84.
Davies DR & Cohen GH (1996) Interactions of protein antigens with antibodies. *Proc Natl Acad Sci USA* 93:7–12.
Davies DR, Padlan EA & Sheriff S (1990) Antibody-antigen complexes. *Annu Rev Biochem* 59:439–473.
Jimenez R, Salazar G, Baldridge KK & Romesberg FE (2003) Flexibility and molecular recognition in the immune system. *Proc Natl Acad Sci USA* 100:92–97.
Padlan EA (1994) Anatomy of the antibody molecule. *Mol. Immunol.* 31:169–217.
Zhang M, Srivastava G & Lu L (2004) The pre-B cell receptor and its function during B cell development. *Cell Mol Immunol* 1:89–94.

The Generation of Antibody Diversity

Besmer E, Gourzi P & Papavasiliou FN (2004) The regulation of somatic hypermutation. *Curr Opin Immunol* 16:241–245.
Corcoran AE (2005) Immunoglobulin locus silencing and allelic exclusion. *Semin Immunol* 17:141–154.
Dudley DD, Chaudhuri J, Bassing CH & Alt FW (2005) Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences. *Adv Immunol* 86:43–112.
Honjo T, Nagaoka H, Shinkura R & Muramatsu M (2005) AID to overcome the limitations of genomic information. *Nature Immunol* 6:655–61.
Jung D & Alt FW (2004) Unraveling V(D)J recombination; insights into gene regulation. *Cell* 116:299–311.
Jung D, Giallourakis C, Mostoslavsky R & Alt FW (2006) Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annu Rev Immunol* 24:541–570.
Longerich S, Basu U, Alt F & Storb U (2006) AID in somatic hypermutation and class switch recombination. *Curr Opin Immunol* 18:164–174.
Maizels N (2005) Immunoglobulin gene diversification. *Annu Rev Genet* 39:23–46.
Neuberger M & Di Noia JM (2007) Molecular mechanisms of antibody somatic hypermutation. *Annu Rev Biochem* 76:in press.
Neuberger MS, Harris RS, Di Noia JM & Petersen-Mahrt SK (2003) Immunity through DNA deamination. *Trends Biochem Sci* 28:305–312.
Selsing E (2006) Ig class switching: targeting the recombinational mechanism. *Curr Opin Immunol* 18:249–254.

T cells and MHC Proteins

Coutinho A, Caramalho I, Seixas E & Demengeot J (2005) Thymic commitment of regulatory T cells is a pathway of TCR-dependent selection that isolates repertoires undergoing positive or negative selection. *Curr Top Microbiol Immunol* 293:43–71.
Krogsgaard M & Davis MM (2005) How T cells 'see' antigen. *Nature Immunol* 6:239–245.
Loureiro J & Ploegh HL (2006) Antigen presentation and the ubiquitin-proteasome system in host-pathogen interactions. *Adv Immunol* 92:225–305.
McDevitt HO (2000) Discovering the role of the major histocompatibility complex in the immune response. *Annu. Rev. Immunol.* 18:1–17.
Mellman I & Steinman RM (2001) Dendritic cells: specialized and regulated antigen processing machines. *Cell* 106:255–258.
Pulendran B (2004) Immune activation: death, danger and dendritic cells. *Curr Biol* 14:R30–32.
Reis e Sousa C (2006) Dendritic cells in a mature age. *Nature Rev Immunol* 6:476–483.
Rudolph MG, Stanfield RL & Wilson IA (2006) How TCRs bind MHCs, peptides, and coreceptors. *Annu Rev Immunol* 24:419–466.
Russell JH & Ley TJ (2002) Lymphocyte-mediated cytotoxicity. *Annu Rev Immunol* 20:323–370.
Sallusto F, Geginat J & Lanzavecchia A (2004) Central memory and effector memory T cell subsets: function, generation, and maintenance. *Annu Rev Immunol* 22:745–763.
Sakaguchi S, Ono M, Setoguchi R et al (2006) Foxp3$^+$ CD25$^+$ CD4$^+$ natural regulatory T cells in dominant self-tolerance and autoimmune disease. *Immunol Rev* 212:8–27.
Starr TK, Jameson SC & Hogquist KA (2003) Positive and negative selection of T cells. *Annu Rev Immunol* 21:139–176.
Steinman RM, Bonifaz L, Fujii S et al (2005) The innate functions of dendritic cells in peripheral lymphoid tissues. *Adv Exp Med Biol* 560:83–97.
Trombetta ES & Mellman I (2005) Cell biology of antigen processing *in vitro* and *in vivo*. *Annu Rev Immunol* 23:975–1028.
Villasenor J, Benoist C & Mathis D (2005) AIRE and APECED: molecular insights into an autoimmune disease. *Immunol Rev* 204:156–164.
von Boehmer H (2004) Selection of the T-cell repertoire: receptor-controlled checkpoints in T-cell development. *Adv Immunol* 84:201–238.
Watanabe N, Wang YH, Lee HK et al (2005) Hassall's corpuscles instruct dendritic cells to induce CD4$^+$CD25$^+$ regulatory T cells in human thymus. *Nature* 436:1181–1185.

Helper T Cells and Lymphocyte Activation

Abbas AK (2003) The control of T cell activation vs. tolerance. *Autoimmun Rev* 2:115–118.
Alegre ML, Frauwirth KA & Thompson CB (2001) T-cell regulation by CD28 and CTLA-4. *Nature Rev Immunol* 1:220–228.
Barclay AN (2003) Membrane proteins with immunoglobulin-like domains—a master superfamily of interaction molecules. *Semin Immunol* 15:215–223.
Castellino F & Germain RN (2006) Cooperation between CD4$^+$ and CD8$^+$ T cells: when, where, and how. *Annu Rev Immunol* 24:519–540.
Deane JA & Fruman DA (2004) Phosphoinositide 3-kinase: diverse roles in immune cell activation. *Annu Rev Immunol* 22:563–598.
Dustin ML, Tseng SY, Varma R & Campi G (2006) T cell-dendritic cell immunological synapses. *Curr Opin Immunol* 18:512–516.
Eason DD, Cannon JP, Haire RN et al (2004) Mechanisms of antigen receptor evolution. *Semin Immunol* 16:215–226.
Matthews SA & Cantrell DA (2006) The role of serine/threonine kinases in T-cell activation. *Curr Opin Immunol* 18:314–320.
McHeyzer-Williams LJ, Malherbe LP & McHeyzer-Williams MG (2006) Helper T cell-regulated B cell immunity. *Curr Top Microbiol Immunol* 311:59–83.
Mills DM & Cambier JC (2003) B lymphocyte activation during cognate interactions with CD4$^+$ T lymphocytes: molecular dynamics and immunologic consequences. *Semin Immunol* 15:325–329.
Murphy KM & Reiner SL (2002) The lineage decisions of helper T cells. *Nature Rev Immunol* 2:933–944.
Peggs KS & Allison JP (2005) Co-stimulatory pathways in lymphocyte regulation: the immunoglobulin superfamily. *B J Haematol* 130:809–824.
Reinhardt RL, Kang SJ, Liang HE & Locksley RM (2006) T helper cell effector fates—who, how and where? *Curr Opin Immunol* 18:271–277.
Reth M & Brummer T (2004) Feedback regulation of lymphocyte signalling. *Nature Rev Immunol* 4:269–277.
Samelson LE (2002) Signal transduction mediated by the T cell antigen receptor: the role of adapter proteins. *Annu Rev Immunol* 20:371–394.
Steinman RM, Hawiger D & Nussenzweig MC (2003) Tolerogenic dendritic cells. *Annu Rev Immunol* 21:685–711.
Villiers MB, Perrin-Cocon L, Marche PN & Villiers C (2004) Complement receptors and B lymphocytes. *Crit Rev Immunol* 24:465–478.