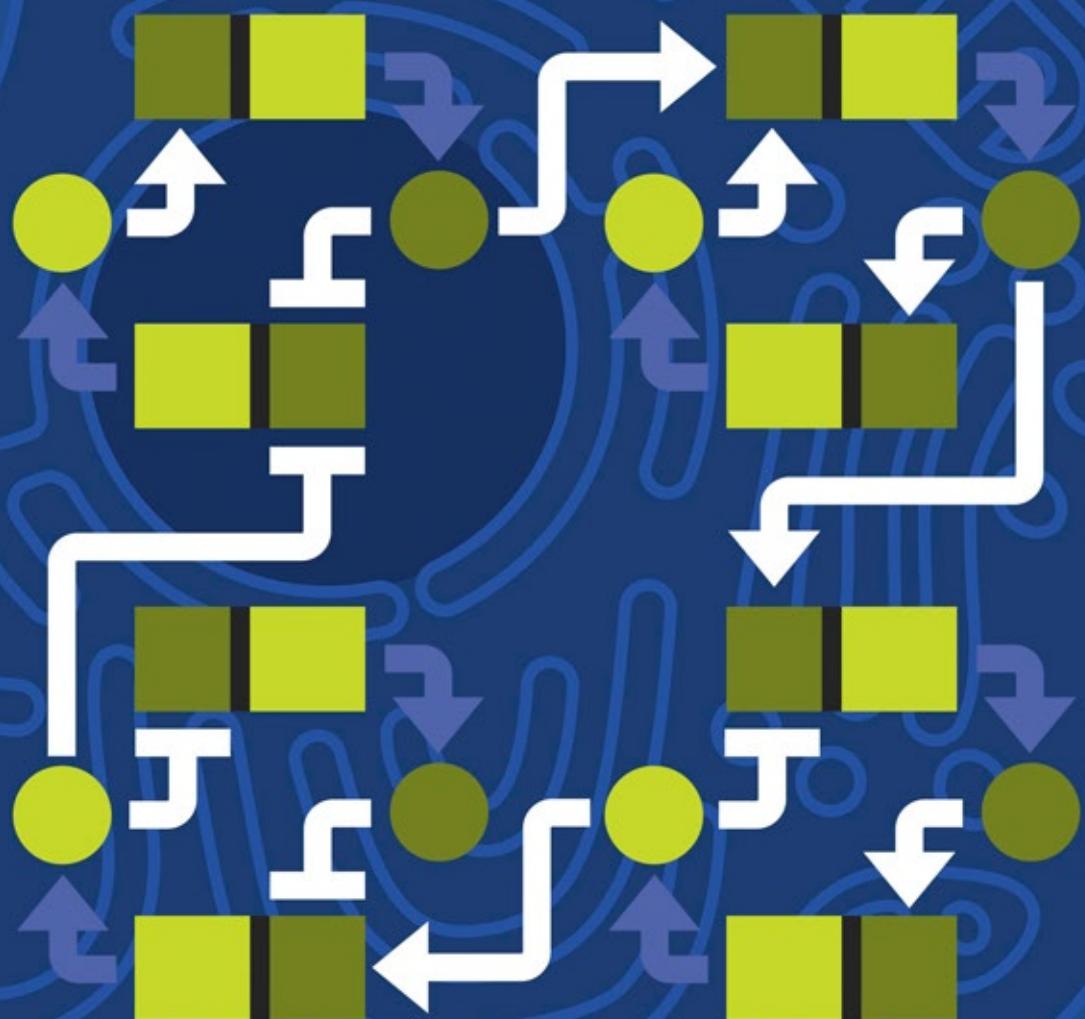


Molecular Biology of **THE CELL**

Sixth Edition



ALBERTS

JOHNSON

LEWIS

MORGAN

RAFF

ROBERTS

WALTER

CELLS IN THEIR SOCIAL CONTEXT

Cell Junctions and the Extracellular Matrix

CHAPTER

19

Of all the social interactions between cells in a multicellular organism, the most fundamental are those that hold the cells together. Cells may be linked by direct interactions, or they may be held together within the *extracellular matrix*, a complex network of proteins and polysaccharide chains that the cells secrete. By one means or another, cells must cohere if they are to form an organized multicellular structure that can withstand and respond to the various external forces that try to pull it apart.

The mechanisms of cohesion govern the architecture of the body—its shape, its strength, and the arrangement of its different cell types. The making and breaking of the attachments between cells and the modeling of the extracellular matrix govern the way cells move within the organism, guiding them as the body grows, develops, and repairs itself. Attachments to other cells and to extracellular matrix control the orientation and behavior of the cell's cytoskeleton, thereby allowing cells to sense and respond to changes in the mechanical features of their environment. Thus, the apparatus of cell junctions and the 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.

The key features of cell junctions and the extracellular matrix are best illustrated by considering two broad categories of tissues that are found in all animals (**Figure 19–1**). **Connective tissues**, such as bone or tendon, are formed from an extracellular matrix produced by cells that are distributed sparsely in the matrix. 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. These *cell-matrix junctions* link the cytoskeleton to the matrix, allowing the cells to move through the matrix and monitor changes in its mechanical properties.

In **epithelial tissues**, such as the lining of the gut or the epidermal covering of the skin, cells are tightly bound together into sheets called **epithelia**. The extracellular matrix is less pronounced, consisting mainly of a thin mat called the *basal lamina* (or *basement membrane*) underlying the sheet. Within the epithelium, cells are attached to each other directly by *cell-cell junctions*, where cytoskeletal filaments are anchored, transmitting stresses across the interiors of the cells, from

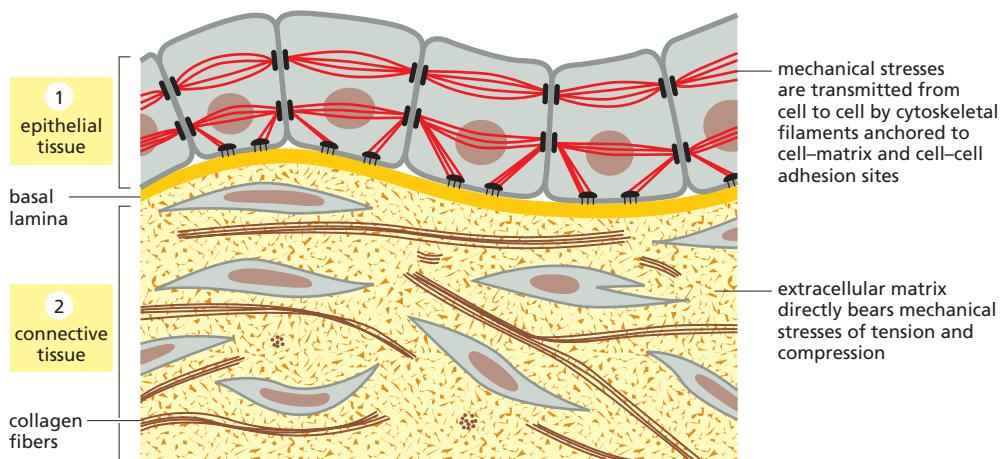
IN THIS CHAPTER

CELL-CELL JUNCTIONS

THE EXTRACELLULAR MATRIX OF ANIMALS

CELL-MATRIX JUNCTIONS

THE PLANT CELL WALL



adhesion site to adhesion site. The cytoskeleton of epithelial cells is also linked to the basal lamina through cell-matrix junctions.

Figure 19–2 provides a closer view of epithelial cells to illustrate the major types of cell-cell and cell-matrix junctions that we will discuss in this chapter. The diagram shows the typical arrangement of junctions in a *simple columnar* epithelium such as the lining of the small intestine of a vertebrate. Here, a single layer of tall cells stands on a basal lamina, with the cells' uppermost surface, or *apex*, free and exposed to the extracellular medium. On their sides, or *lateral surfaces*, the cells make junctions with one another. Two types of **anchoring junctions** link the cytoskeletons of adjacent cells: **adherens junctions** are anchorage sites for actin filaments; **desmosomes** are anchorage sites for intermediate filaments. Two additional types of anchoring junctions link the cytoskeleton of the epithelial cells to the basal lamina: **actin-linked cell-matrix junctions** anchor actin filaments to the matrix, while **hemidesmosomes** anchor intermediate filaments to it.

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 adhesive junctions. Cell-matrix attachments bond epithelial tissue to the connective tissue beneath it.

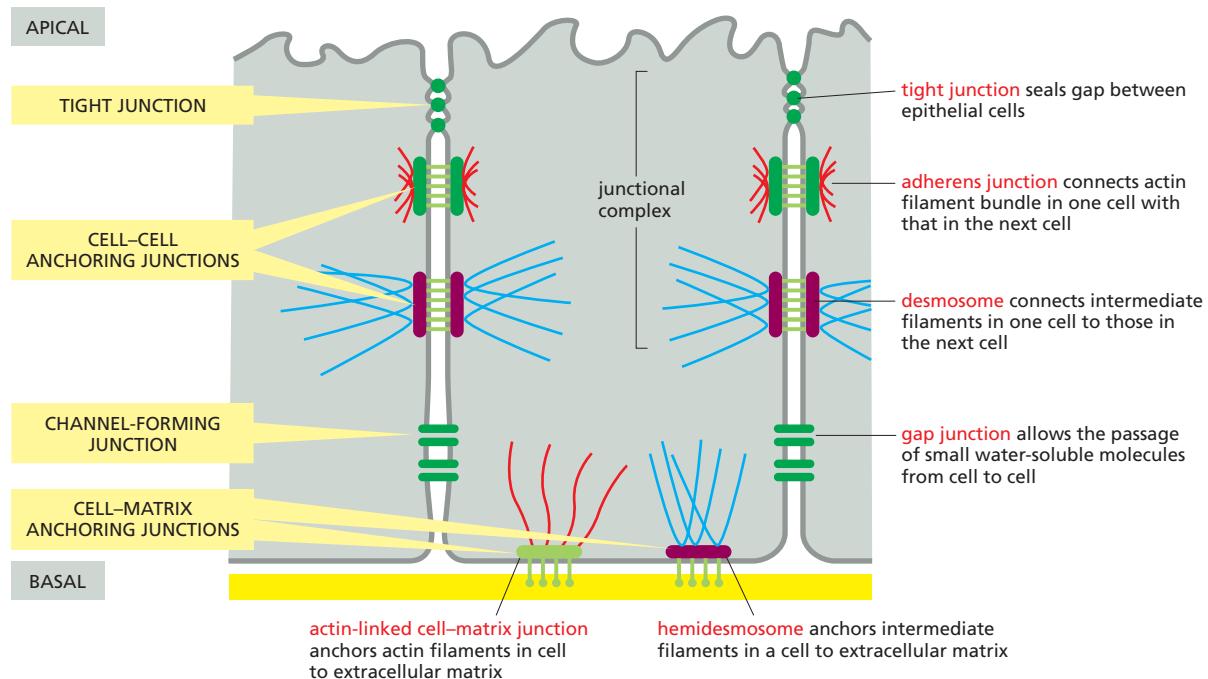


Figure 19–2 A summary of the various cell junctions found in a vertebrate epithelial cell, classified according to their primary functions. In the most apical portion of the cell, the relative positions of the junctions are the same in nearly all vertebrate epithelia. The tight junction occupies the 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. Two types of cell-matrix anchoring junctions tether the basal surface of the cell to the basal lamina. The drawing is based on epithelial cells of the small intestine.

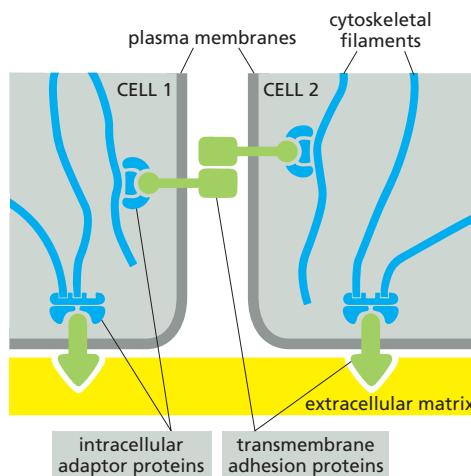


Figure 19–3 Transmembrane adhesion proteins link the cytoskeleton to extracellular structures. The external linkage may be either to other cells (cell-cell junctions, mediated typically by cadherins) or to extracellular matrix (cell-matrix junctions, mediated typically by integrins). The internal linkage to the cytoskeleton is generally indirect, via intracellular adaptor proteins, to be discussed later.

Two other types of cell-cell junction are shown in Figure 19–2. *Tight junctions* hold the cells closely together near the apex, sealing the gap between the cells and thereby preventing molecules from leaking across the epithelium. Near the basal end of the cells are channel-forming junctions, called *gap junctions*, that create passageways linking the cytoplasms of adjacent cells.

Each of the four major anchoring junction types depends on **transmembrane adhesion proteins** that span the plasma membrane, with one end linking to the cytoskeleton inside the cell and the other end linking to other structures outside it (Figure 19–3). These cytoskeleton-linked transmembrane proteins fall neatly into two superfamilies, corresponding to the two basic kinds of external attachment. Proteins of the **cadherin** superfamily chiefly mediate attachment of cell to cell (Movie 19.1). Proteins of the **integrin** superfamily chiefly mediate attachment of cells to matrix. There is specialization within each family: some cadherins link to actin and form adherens junctions, while others link to intermediate filaments and form desmosomes; likewise, some integrins link to actin and form actin-linked cell-matrix junctions, while others link to intermediate filaments and form hemidesmosomes (Table 19–1).

TABLE 19–1 Anchoring Junctions

Junction	Transmembrane adhesion protein	Extracellular ligand	Intracellular cytoskeletal attachment	Intracellular adaptor proteins
Cell–Cell				
Adherens junction	Classical cadherins	Classical cadherin on neighboring cell	Actin filaments	α -Catenin, β -catenin, plakoglobin (γ -catenin), p120-catenin, vinculin
Desmosome	Nonclassical cadherins (desmoglein, desmocollin)	Desmoglein and desmocollin on neighboring cell	Intermediate filaments	Plakoglobin (γ -catenin), plakophilin, desmoplakin
Cell–Matrix				
Actin-linked cell–matrix junction	Integrin	Extracellular matrix proteins	Actin filaments	Talin, kindlin, vinculin, paxillin, focal adhesion kinase (FAK), numerous others
Hemidesmosome	$\alpha_6\beta_4$ Integrin, type XVII collagen	Extracellular matrix proteins	Intermediate filaments	Plectin, BP230

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 transient cell–cell attachments more flimsy than anchoring junctions, but sufficient to stick cells together in special circumstances.

We begin the chapter with a discussion of the major forms of cell–cell junctions. We then consider in turn the extracellular matrix of animals, the structure and function of integrin-mediated cell–matrix junctions, and, finally, the plant cell wall, a special form of extracellular matrix.

CELL-CELL JUNCTIONS

Cell–cell junctions come in many forms and can be regulated by a variety of mechanisms. The best understood and most common are the two types of cell–cell anchoring junctions, which employ cadherins to link the cytoskeleton of one cell with that of its neighbor. Their primary function is to resist the external forces that pull cells apart. The epithelial cells of your skin, for example, must remain tightly linked when they are stretched, pinched, or poked. Cell–cell anchoring junctions must also be dynamic and adaptable, so that they can be altered or rearranged when tissues are remodeled or repaired, or when there are changes in the forces acting on them.

In this section, we focus primarily on the cadherin-based anchoring junctions. We then briefly describe tight junctions and gap junctions. Finally, we consider the more transient cell–cell adhesion mechanisms employed by some cells in the bloodstream.

Cadherins Form a Diverse Family of Adhesion Molecules

Cadherins are present in all multicellular animals whose genomes have been analyzed. They are also present in the choanoflagellates, which 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 eukaryotes, including fungi and plants, lack cadherins, and they are also absent from bacteria and archaea. 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 apart. The first three cadherins to be 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 other tissues. These and other **classical cadherins** are closely related in sequence throughout their extracellular and intracellular domains.

There are also a large number of **nonclassical cadherins** that are 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 desmosomes (see Table 19–1). Other family members are involved primarily in signaling. Together, the classical and nonclassical cadherin proteins constitute the **cadherin superfamily** (Figure 19–4), with more than 180 members in humans.

Cadherins Mediate Homophilic Adhesion

Anchoring junctions between cells are usually symmetrical: if the linkage is to actin in the cell on one side of the junction, it will be to actin in the cell on the other side. In fact, the binding between cadherins is generally **homophilic** (like-to-like, Figure 19–5): cadherin molecules of a specific subtype on one cell bind to cadherin molecules of the same or closely related subtype on adjacent cells.

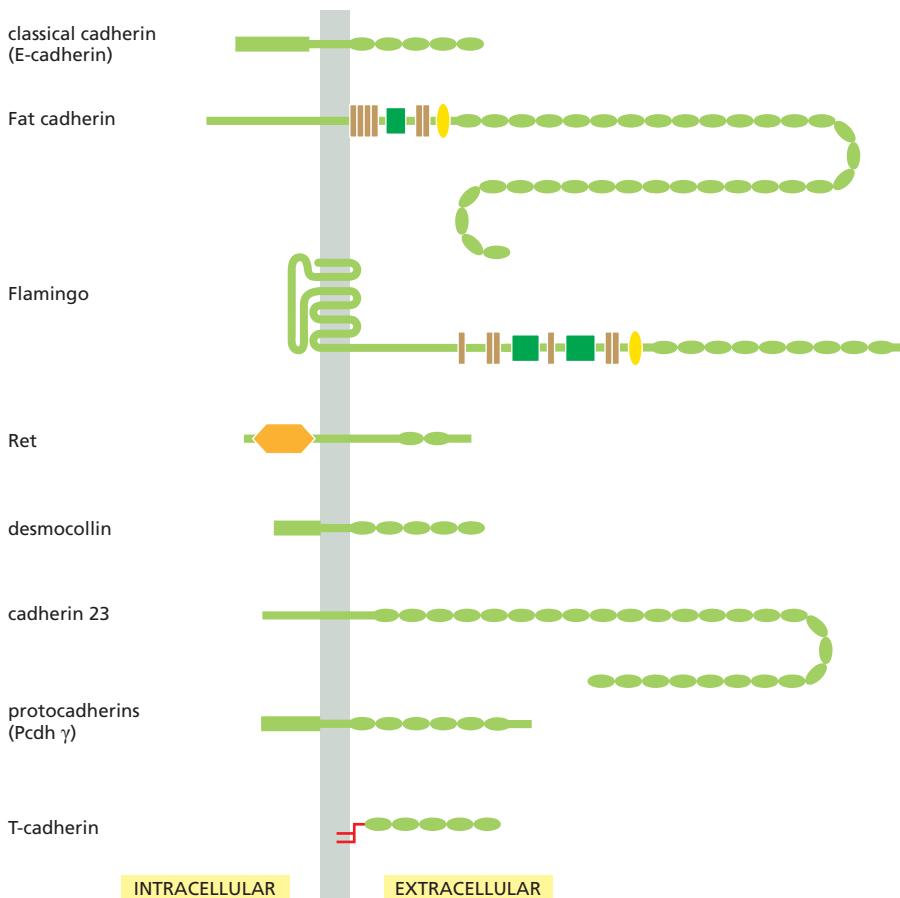


Figure 19–4 The cadherin superfamily.

The diagram shows some of the diversity among cadherin superfamily members. These proteins all have extracellular portions containing multiple copies of the extracellular cadherin domain (green ovals). In the classical cadherins of vertebrates there are 5 of these domains, and in desmogleins and desmocollins there are 4 or 5, but some nonclassical cadherins have more than 30. The intracellular portions are more varied, reflecting interactions with a wide variety of intracellular ligands, including signaling molecules and adaptor proteins that connect the cadherin to the cytoskeleton. In some cases, such as T-cadherin, a transmembrane domain is not present and the protein is attached to the plasma membrane by a glycosylphosphatidylinositol (GPI) anchor. The differently colored motifs in Fat, Flamingo, and Ret represent conserved domains that are also found in other protein families.

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 the *extracellular cadherin (EC) domain*. Homophilic binding occurs at the N-terminal tips of the cadherin molecules—the cadherin domains that lie furthest from the membrane. These terminal domains each form a knob and a nearby pocket, and the cadherin molecules protruding from opposite cell membranes bind by insertion of the knob of one domain into the pocket of the other (Figure 19–6A).

Each cadherin domain forms a more-or-less rigid unit, joined to the next cadherin domain by a hinge. Ca^{2+} ions bind to sites near each hinge and prevent it from flexing, so that the whole string of cadherin domains behaves as a rigid and slightly curved rod. When Ca^{2+} is removed, the hinges can flex, and the structure becomes floppy (Figure 19–6B). 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.

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 (Figure 19–6C). The strength of this junction is far greater than that of any individual intermolecular bond, and yet regulatory mechanisms can easily disassemble the junction by separating the molecules sequentially, just as two pieces of fabric can be joined strongly by Velcro and yet easily peeled apart from the sides. A similar “Velcro principle” also operates at cell-cell and cell-matrix adhesions formed by other types of transmembrane adhesion proteins.

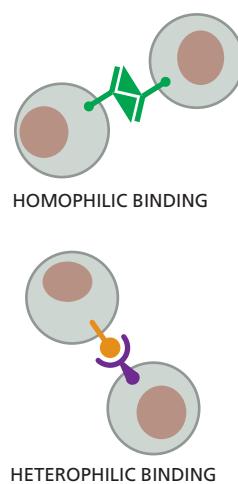


Figure 19–5 Homophilic versus heterophilic binding. Cadherins in general bind homophilically; some other cell adhesion molecules, discussed later, bind heterophilically.

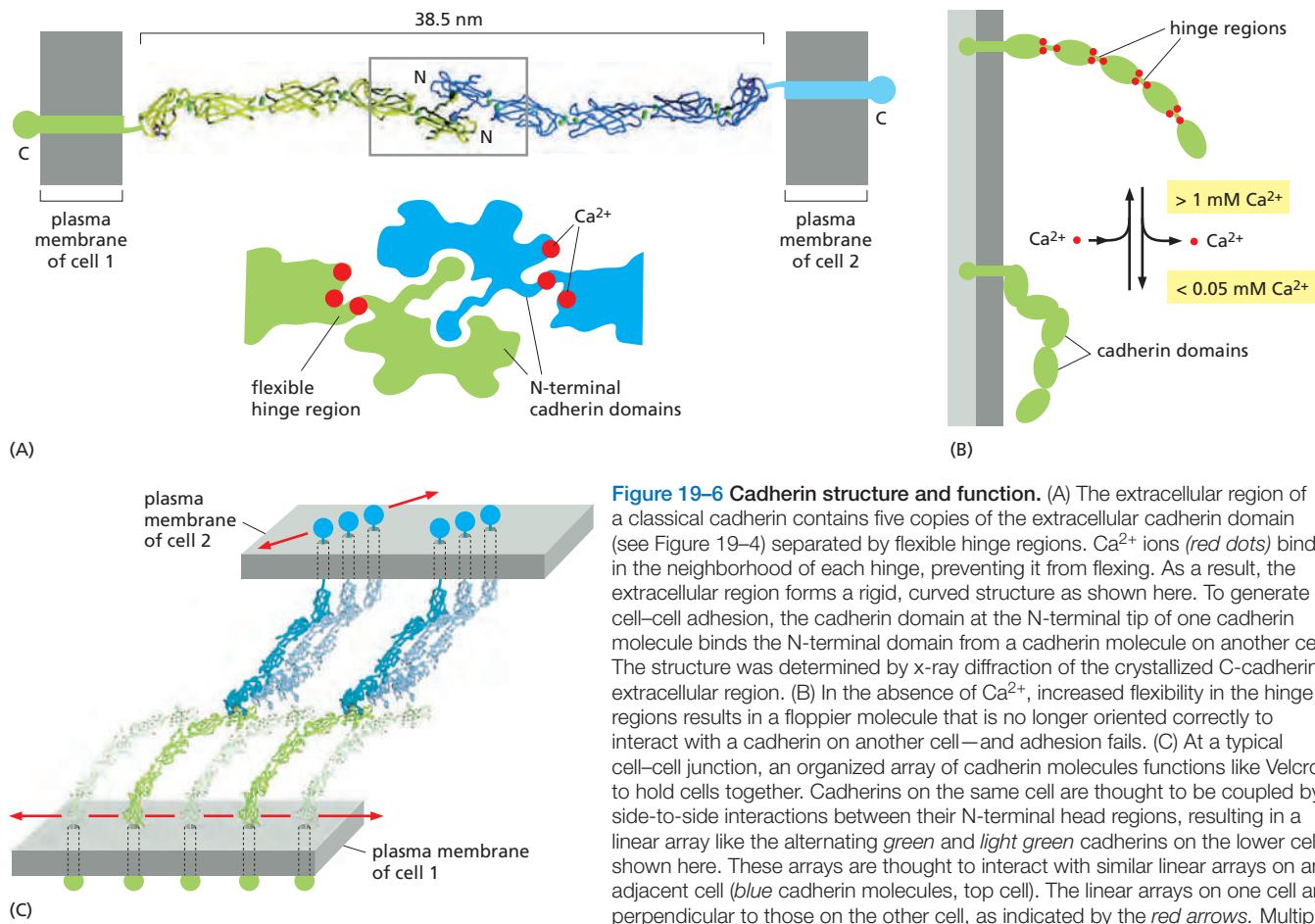


Figure 19–6 Cadherin structure and function. (A) The extracellular region of a classical cadherin contains five copies of the extracellular cadherin domain (see Figure 19–4) separated by flexible hinge regions. Ca^{2+} ions (red dots) bind in the neighborhood of each hinge, preventing it from flexing. As a result, the extracellular region forms a rigid, curved structure as shown here. To generate cell-cell adhesion, the cadherin domain at the N-terminal tip of one cadherin molecule binds the N-terminal domain from a cadherin molecule on another cell. The structure was determined by x-ray diffraction of the crystallized C-cadherin extracellular region. (B) In the absence of Ca^{2+} , increased flexibility in the hinge regions results in a floppier molecule that is no longer oriented correctly to interact with a cadherin on another cell—and adhesion fails. (C) At a typical cell-cell junction, an organized array of cadherin molecules functions 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, resulting in a linear array like the alternating green and light green cadherins on the lower cell shown here. These arrays are thought to interact with similar linear arrays on an adjacent cell (blue cadherin molecules, top cell). The linear arrays on one cell are perpendicular to those on the other cell, as indicated by the red arrows. Multiple perpendicular arrays on both cells interact to form a tight-knit mat of cadherin proteins. (A, based on T.J. Boggon et al., *Science* 296:1308–1313, 2002; C, based on O.J. Harrison et al. *Structure* 19:244–256, 2011.)

Cadherin-Dependent Cell–Cell Adhesion Guides the Organization of Developing Tissues

Cadherins form specific homophilic attachments, explaining 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.

Selectivity in the way that animal cells consort with one another was first demonstrated in the 1950s, 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 into structures resembling those of the original embryo (**Figure 19–7**). These experiments, together with numerous more recent experiments, reveal that selective cell–cell recognition systems make cells of the same differentiated tissue preferentially adhere to one another.

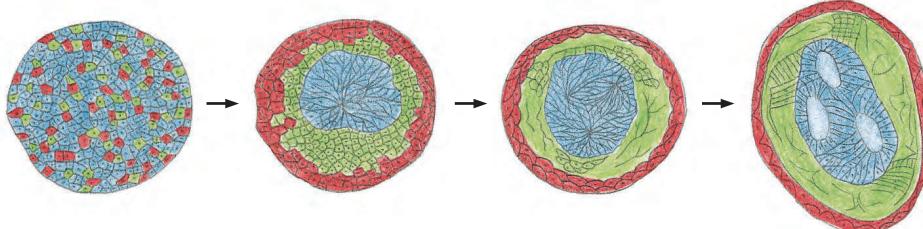


Figure 19–7 Sorting out. Cells from different layers of an early amphibian embryo will sort out according to their origins. In the classical experiment shown here, mesoderm cells (green), neural plate cells (blue), and epidermal cells (red) have been disaggregated and then reaggregated in a random mixture. 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.)

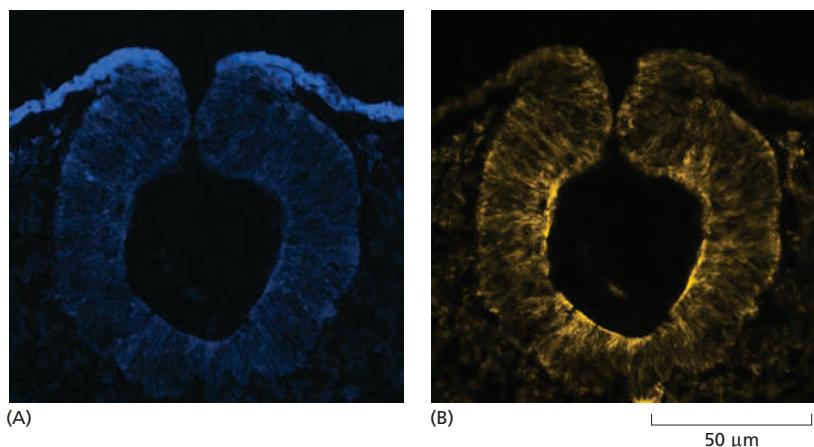
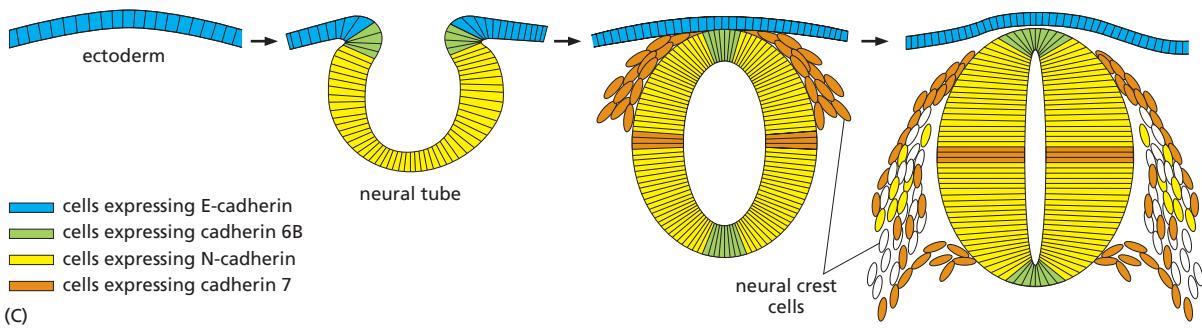


Figure 19-8 Changing patterns of cadherin expression during construction of the vertebrate nervous system. The figure shows cross sections of the early chick embryo, as the neural tube detaches from the ectoderm and then as neural crest cells detach from the neural tube. (A, B) Immunofluorescence micrographs showing the developing neural tube labeled with antibodies against (A) E-cadherin (blue) and (B) N-cadherin (yellow). (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 Miwako Nomura and Masatoshi Takeichi.)



Cadherins play a crucial part in these cell-sorting processes during development. 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. In the vertebrate embryo, for example, changes in cadherin expression are seen when the neural tube forms and pinches off from the overlying ectoderm: 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 19-8A and 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 19-8C). Finally, when the cells aggregate to form a ganglion, they switch on expression of N-cadherin again. 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 further support the idea 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, E-cadherins on one cell bind to E-cadherins on another, resulting in cell-cell adhesion. 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 19-9A), 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 19-9B). It therefore seems likely that both qualitative and quantitative differences in the expression of cadherins have a role in organizing tissues.

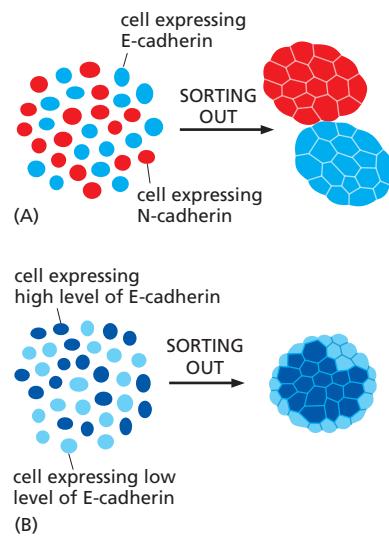


Figure 19-9 Cadherin-dependent cell sorting. Cells in culture can sort themselves out according to the type and level of cadherins they express. This can be visualized by labeling different populations of cells with dyes of different colors. (A) Cells expressing N-cadherin sort out from cells expressing E-cadherin. (B) Cells expressing high levels of E-cadherin sort out from cells expressing low levels of E-cadherin. The cells expressing high levels adhere more strongly and end up internally.

Epithelial–Mesenchymal Transitions Depend on Control of Cadherins

The assembly of cells into an epithelium is a reversible process. By switching on expression of adhesion molecules, dispersed unattached *mesenchymal cells*, such as fibroblasts, can come together to form an epithelium. Conversely, epithelial cells can change their character, disassemble, and migrate away from their parent epithelium as separate cells. Such *epithelial–mesenchymal transitions* play an important part in normal embryonic development; the origin of the neural crest is one example. These transitions depend in part on transcription regulatory proteins called Slug, Snail, and Twist. Increased expression of Twist, for example, converts epithelial cells to a mesenchymal character, and switching it off does the opposite. Twist exerts its effects, in part, by inhibiting expression of cadherins, including E-cadherin, that hold epithelial cells together.

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 the cells 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. Mutations that disrupt the production or function of E-cadherin are often found in cancer cells and are thought to help make them malignant.

Catenins Link Classical Cadherins to the Actin Cytoskeleton

The extracellular domains of cadherins mediate homophilic binding at adherens junctions. The intracellular domains of typical cadherins, including all classical and some nonclassical ones, interact with filaments of the cytoskeleton: actin at adherens junctions and intermediate filaments at desmosomes (see Table 19–1). These cytoskeletal linkages are essential for efficient cell–cell adhesion, as cadherins that lack their cytoplasmic domains cannot stably hold cells together.

The linkage of cadherins to the cytoskeleton is indirect and depends on adaptor proteins that assemble on the cytoplasmic tail of the cadherin. At adherens junctions, the cadherin tail binds two such proteins: β -catenin and a distant relative called *p120-catenin*; a third protein called α -catenin interacts with β -catenin and recruits a variety of other proteins to provide a dynamic linkage to actin filaments (Figure 19–10). At desmosomes, cadherins are linked to intermediate filaments through other adaptor proteins, including a β -catenin-related protein called *plakoglobin*, as we discuss later.

In their mature form, adherens junctions are enormous protein complexes containing hundreds to thousands of cadherin molecules, packed into dense, regular arrays that are linked on the extracellular side by lateral interactions between cadherin domains, as we discussed earlier (see Figure 19–6C). On the cytoplasmic side, a complex network of catenins, actin regulators, and contractile actin bundles holds the cluster of cadherins together and links it to the actin cytoskeleton. Assembling a structure of this complexity is not a simple task, and it involves a complex sequence of events controlled by the actin-regulatory proteins discussed in Chapter 16. The general features of the assembly process are summarized in Figure 19–11.

Adherens Junctions Respond to Forces Generated by the Actin Cytoskeleton

Most adherens junctions are linked to contractile bundles of actin filaments and non-muscle myosin II. These junctions are therefore subjected to pulling forces generated by the attached actin. The pulling forces are important for junction assembly and maintenance: disruption of myosin activity, for example, results in

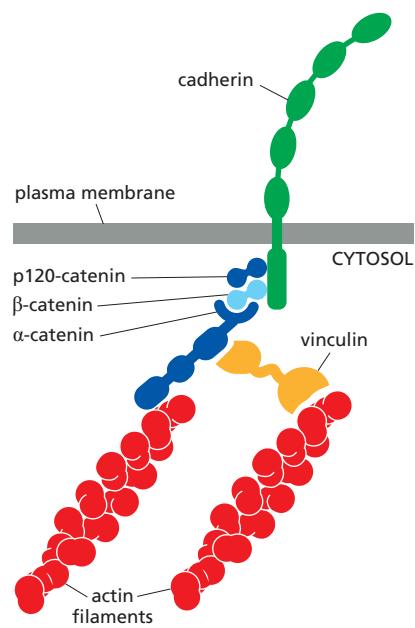


Figure 19–10 The linkage of classical cadherins to actin filaments. The cadherins are coupled indirectly to actin filaments through an adaptor protein complex containing p120-catenin, β -catenin, and α -catenin. Other proteins, including vinculin, associate with α -catenin and help provide the linkage to actin. β -Catenin has a second, and very important, function in intracellular signaling, as we discuss in Chapter 15 (see Figure 15–60). For clarity, this diagram does not show the cadherin of the adjacent cell in the junction.

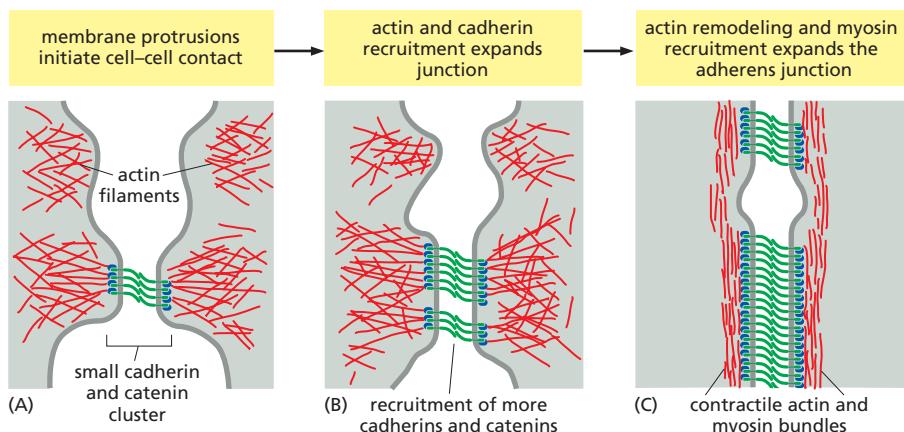


Figure 19–11 Assembly of an adherens junction. (A) Assembly begins when two unattached epithelial cell precursors explore their surroundings with membrane protrusions, generated by local nucleation of actin networks. When the cells make contact, small cadherin and catenin clusters take shape at the contact sites and associate with actin, leading to activation of the small monomeric GTPase Rac (not shown), an important actin regulator (see Figure 16–85). (B) Rac promotes additional actin protrusions in the vicinity, expanding the size of the contact zone and thereby promoting further recruitment of cadherins and their associated catenin proteins. (C) Eventually, Rac is inactivated and replaced by the related GTPase Rho (not shown), which shifts actin remodeling toward the assembly of linear, contractile filament bundles. Rho also promotes the assembly of myosin II filaments that associate with bundles of actin filaments to generate contractile activity. This contractile activity generates tension that stimulates further actin recruitment and expansion of the junction, in part through the mechanisms illustrated in Figure 19–12.

the disassembly of many adherens junctions. Furthermore, the contractile forces acting on a junction in one cell are balanced by contractile forces at the junction of the opposite cell, so that no cell pulls others toward it and thereby disrupts the uniform distribution of cells in the tissue.

We do not understand the mechanisms responsible for maintaining this balance. Adherens junctions seem to sense the forces acting on them and modify local actin and myosin behavior to balance the forces on both sides of the junction. Evidence for these mechanisms comes from studies of pairs of cultured mammalian cells connected by adherens junctions. If contractile activity in one cell is increased experimentally, the adherens junctions linking the two cells increase in size, and the contractile activity of the second cell increases to match that of the first—resulting in a balance of forces across the junction. These and other experiments suggest that adherens junctions are not simply passive sites of protein–protein binding but are dynamic tension sensors that regulate their behavior in response to changing mechanical conditions. This ability to transduce a mechanical signal into a change in junctional behavior is an example of *mechanotransduction*. We will see later that it is also important at cell–matrix junctions.

The mechanotransduction at cell–cell junctions is thought to depend, at least in part, on proteins in the cadherin complex that alter their shape when stretched by tension. The protein α -catenin, for example, is stretched from a folded to an extended conformation when contractile activity increases at the junction. The unfolding exposes a cryptic binding site for another protein, vinculin, which promotes the recruitment of more actin to the junction (Figure 19–12). By mechanisms such as this, pulling on a junction makes it stronger. Furthermore, as noted above, pulling on a junction in one cell will increase the contractile force generated in the attached cell.

In some cell types, actin contractility reduces cell–cell adhesion, particularly if large forces are involved. Large actin-based contractile forces might, in some tissues, pull sufficiently hard on the edges of cell–cell adhesions to peel them apart, particularly if contraction is coupled to additional regulatory mechanisms that weaken the adhesion. This mechanism might be important in certain forms of tissue remodeling during development, as we describe next.

Tissue Remodeling Depends on the Coordination of Actin-Mediated Contraction With Cell–Cell Adhesion

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 linear attachments that connect the cortical

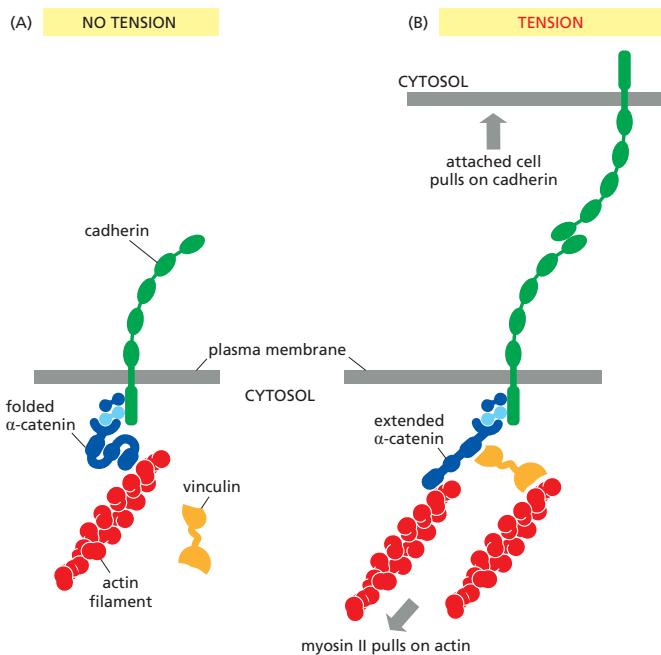


Figure 19–12 Mechanotransduction in an adherens junction. (A) Cell-cell junctions are able to sense increased tension and respond by strengthening their actin linkages. Tension sensing is thought to depend in part on α -catenin (see Figure 19–10). (B) When actin filaments are pulled from within the cell by non-muscle myosin II, the resulting force unfolds a domain in α -catenin, thereby exposing an otherwise hidden binding site for the adaptor protein vinculin. Vinculin then promotes additional actin recruitment, strengthening the linkages between the junction and the cytoskeleton.

actin filaments beneath the plasma membranes of two interacting cells. In heart muscle, they anchor the actin bundles of the contractile apparatus and act in parallel with desmosomes to link the contractile cells end-to-end. But the prototypical examples of adherens junctions occur in epithelia, where they often form a continuous **adhesion belt** (or *zonula adherens*) just beneath the apical face of the epithelium, encircling each of the interacting cells in the sheet (Figure 19–13). Within each cell, a contractile bundle of actin filaments and myosin II lies adjacent to the adhesion belt, oriented parallel to the plasma membrane and tethered to

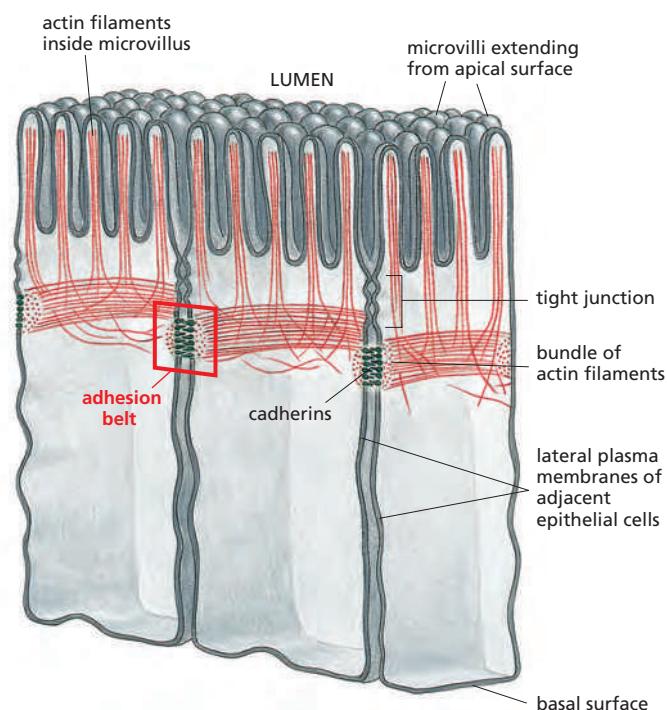


Figure 19–13 Adherens junctions between epithelial cells in the small intestine. These cells are specialized for absorption of nutrients; at their apex, facing the lumen of the gut, they have many microvilli (protrusions that increase the absorptive surface area). The adherens junction takes the form 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 proteins to cadherins, which bind to cadherins on the adjacent cell. In this way, the actin filament bundles in adjacent cells are tied together. For clarity, this drawing does not show most of the other cell-cell and cell–matrix junctions of epithelial cells (see Figure 19–2).

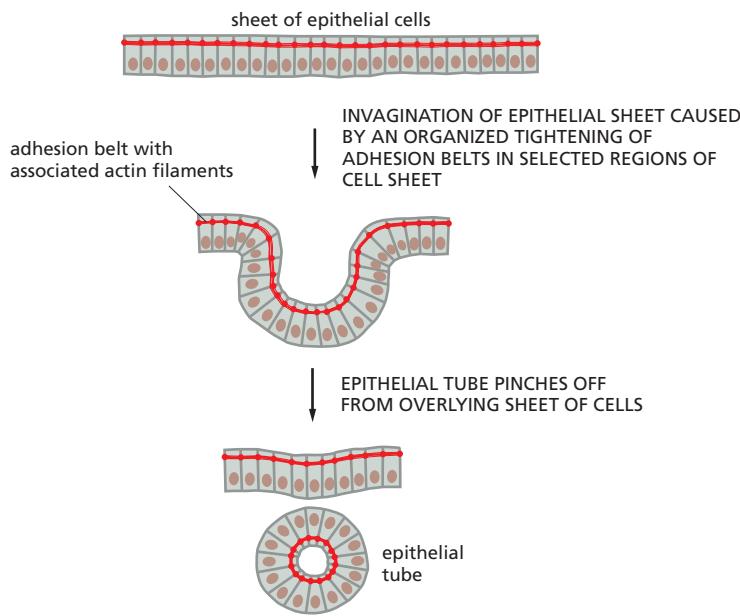


Figure 19–14 The folding of an epithelial sheet to form an epithelial tube. The oriented contraction of the bundles of actin and myosin filaments running along adhesion belts causes the epithelial cells to narrow at their apex and helps the epithelial sheet to roll up into a tube. An example is the formation of the neural tube in early vertebrate development (see Figure 19–8).

it by the cadherins and their associated intracellular adaptor proteins. The actin-myosin bundles are thus linked, via the cadherins, into an extensive transcellular network. Coordinated contraction of this network provides the motile force for a fundamental process in animal morphogenesis—the folding of epithelial cell sheets into tubes, vesicles, and other related structures (**Figure 19–14**).

The coordination of cell-cell adhesion and actin contractility is beautifully illustrated by cellular rearrangements that occur early in the development of the fruit fly *Drosophila melanogaster*. Soon after gastrulation, the outer epithelium of the embryo is elongated by a process called *germ-band extension*, in which the cells converge inward toward the dorsal-ventral axis and extend along the anterior-posterior axis (**Figure 19–15**). Actin-dependent contraction along specific cell boundaries is coordinated with a loss of specific adherens junctions to allow cells to insert themselves between other cells (a process called *intercalation*), resulting in a longer and narrower epithelium. The mechanisms underlying the loss of adhesion along specific cell boundaries are not clear, but they depend in part on increased degradation of β -catenin, due to its phosphorylation by a protein kinase that is localized specifically at those boundaries.

Desmosomes Give Epithelia Mechanical Strength

Desmosomes are structurally similar to adherens junctions but contain specialized cadherins that link to intermediate filaments instead of actin filaments. Their main function is to provide mechanical strength. Desmosomes are important

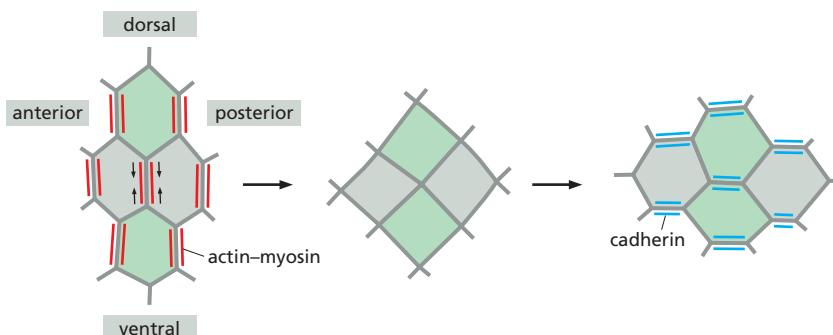


Figure 19–15 Remodeling of cell-cell adhesions in embryonic *Drosophila* epithelium. Depicted at left is a group of cells in the outer epithelium of a *Drosophila* embryo. During germ-band extension, cells converge toward each other (middle) on the dorsal-ventral axis and then extend (right) along the anterior-posterior axis. The result is intercalation: cells that were originally far apart along the dorsal-ventral axis (dark green) are inserted between the cells (light green) that separated them. These rearrangements depend on the spatial regulation of actin-myosin contractile bundles, which are localized primarily at the vertical cell boundaries (red, left). Contraction of these bundles is accompanied by removal of E-cadherin (not shown) at the same cell boundaries, resulting in shrinkage and loss of adhesion along the vertical axis (middle). New cadherin-based adhesions (blue, right) then form and expand along horizontal boundaries, resulting in extension of the cells in the anterior-posterior dimension.

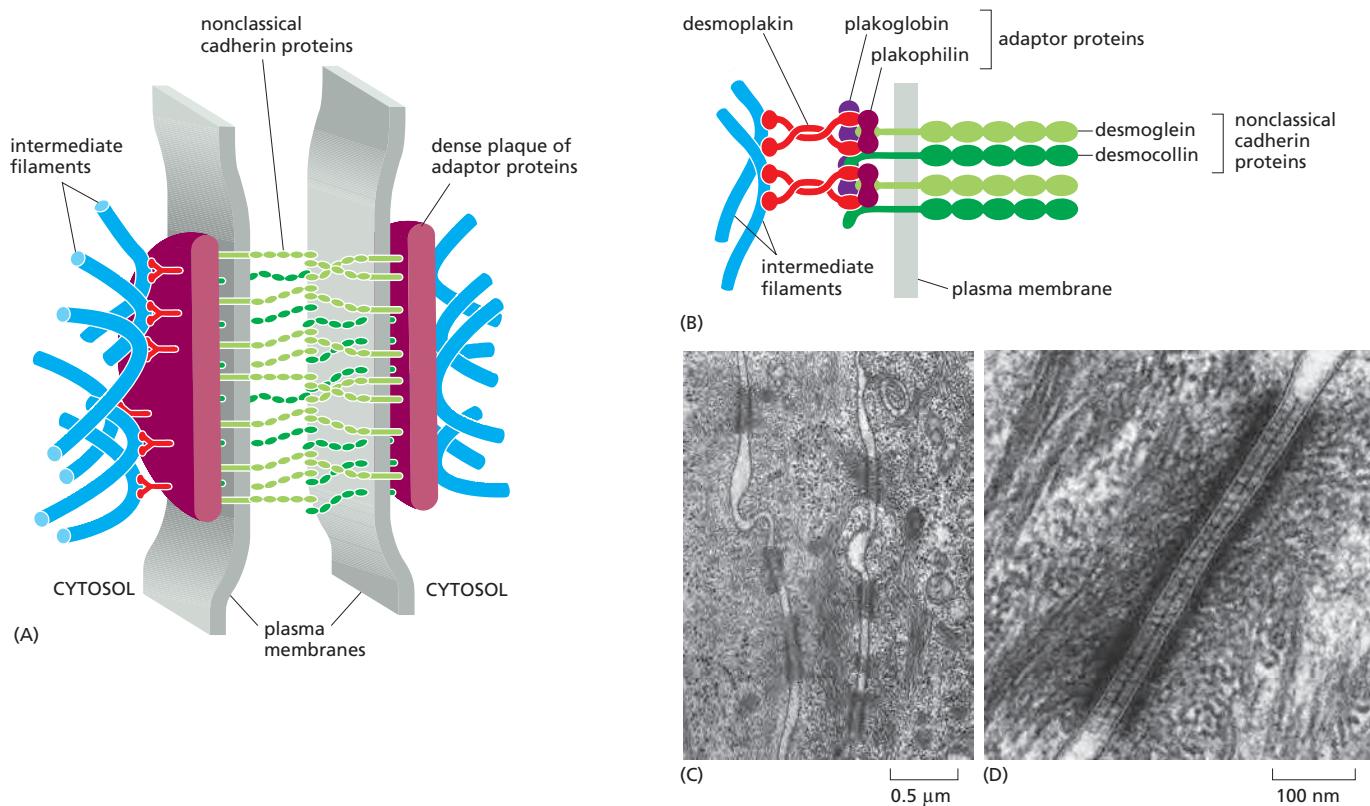


Figure 19–16 Desmosomes. (A) The structural components of a desmosome. On the cytoplasmic surface of each interacting plasma membrane is a dense plaque composed of a mixture of intracellular adaptor proteins. A bundle of keratin intermediate filaments is attached to the surface of each plaque. Transmembrane nonclassical cadherins bind to the plaques and interact through their extracellular domains to hold the adjacent membranes together. (B) Some of the molecular components of a desmosome. Desmoglein and desmocollin are nonclassical cadherins. 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 three 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.)

in vertebrates but are not found, for example, in *Drosophila*. They are present in most mature vertebrate epithelia and are particularly plentiful in tissues that are subject to high levels of mechanical stress, such as heart muscle and the epidermis, the epithelium that forms the outer layer of the skin.

Figure 19–16A shows the general structure of a desmosome, and Figure 19–16B shows some of the proteins that form it. Desmosomes typically appear as buttonlike spots of adhesion, riveting the cells together (Figure 19–16C). Inside the cell, the bundles of ropelike intermediate filaments that are anchored to the desmosomes form a structural framework of great tensile strength (Figure 19–16D), with linkage to similar bundles in adjacent cells, creating a network that extends throughout the tissue (Figure 19–17). 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 desmosomes 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.

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

Sheets of epithelial cells 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 but retaining an organization based on a set of conserved molecular mechanisms.

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. 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 **tight junctions**, so that molecules cannot leak freely across the cell sheet.

The epithelium of the small intestine provides a good illustration of tight-junction structure and function (see Figure 19–2). 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 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 passive transport 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 19–18). Moreover, the transport proteins must be correctly distributed in the plasma membranes: the apical transporters must be delivered to the cell apex and must not be allowed to drift to the basolateral membrane, and the basolateral transporters must be delivered to and remain in the basolateral membrane. Tight junctions, besides sealing the gaps between the cells, also function as “fences” that help prevent apical or basolateral proteins from diffusing into the wrong region.

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 19–19). This seal is not absolute, however. Although all tight junctions are impermeable to macromolecules, their permeability to ions and other small molecules varies. Tight junctions in the epithelium lining the small intestine, for example, are 10,000 times more permeable to inorganic ions, such as Na^+ , than the tight junctions in the epithelium lining the urinary bladder. The movement of ions and other molecules between epithelial cells is called *paracellular transport*, and tissue-specific differences in transport rates generally result from differences in the proteins that form tight junctions.

Tight Junctions Contain Strands of Transmembrane Adhesion Proteins

When tight junctions are visualized by freeze-fracture electron microscopy, they are seen as a branching network of *sealing strands* that completely encircles the

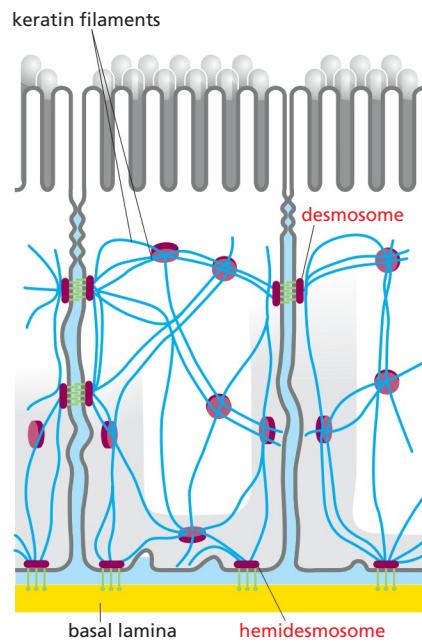


Figure 19–17 Desmosomes, hemidesmosomes, and the intermediate filament network. The keratin intermediate filament networks of adjacent cells—in this example, epithelial cells of the small intestine—are indirectly connected to one another through desmosomes, and to the basal lamina through hemidesmosomes.

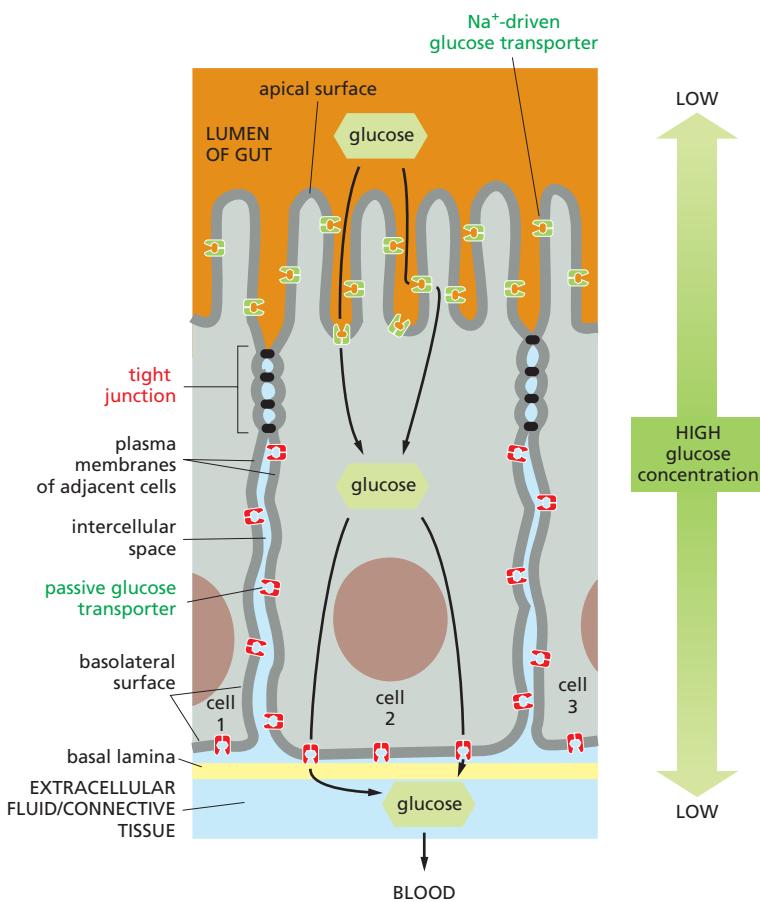


Figure 19–18 The role of tight junctions in transcellular transport. For clarity, only the tight junctions are shown. Transport proteins are confined to different regions of the plasma membrane in epithelial cells of the small intestine. This segregation permits a vectorial transfer of nutrients across the epithelium from the gut lumen to the blood. In the example shown, glucose is actively transported into the cell by Na⁺-driven glucose transporters at its apical surface, and it leaves the cell through passive glucose transporters 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 (see Movie 11.2).

apical end of each cell in the epithelial sheet (Figure 19–20A and B). In conventional electron micrographs, the outer leaflets of the two interacting plasma membranes are tightly apposed where sealing strands are present (Figure 19–20C). Each sealing strand is composed of a long row of transmembrane homophilic 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 19–21).

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

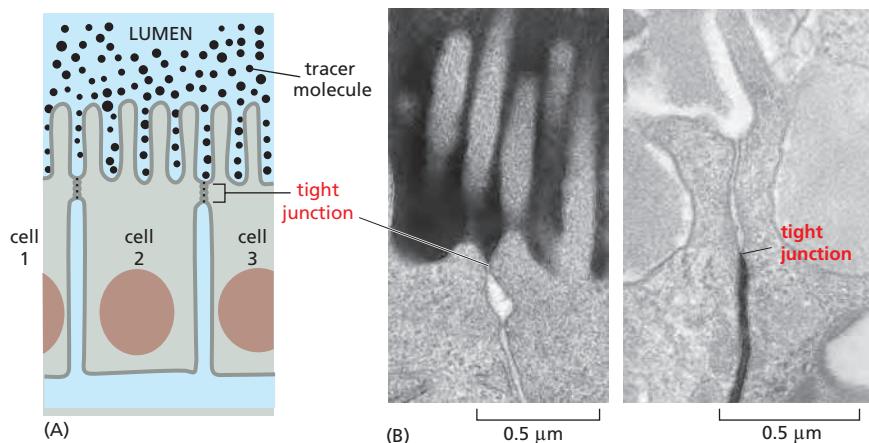


Figure 19–19 The role of tight junctions in allowing epithelia to serve as barriers to solute diffusion. (A) The drawing shows how a small extracellular tracer molecule added on one side of an epithelium is prevented from crossing the epithelium by the tight junctions that seal adjacent cells together. Adherens junctions and other cell junctions are not shown for clarity. (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). The tight junction blocks passage of the tracer in both directions. (B, courtesy of Daniel Friend.)

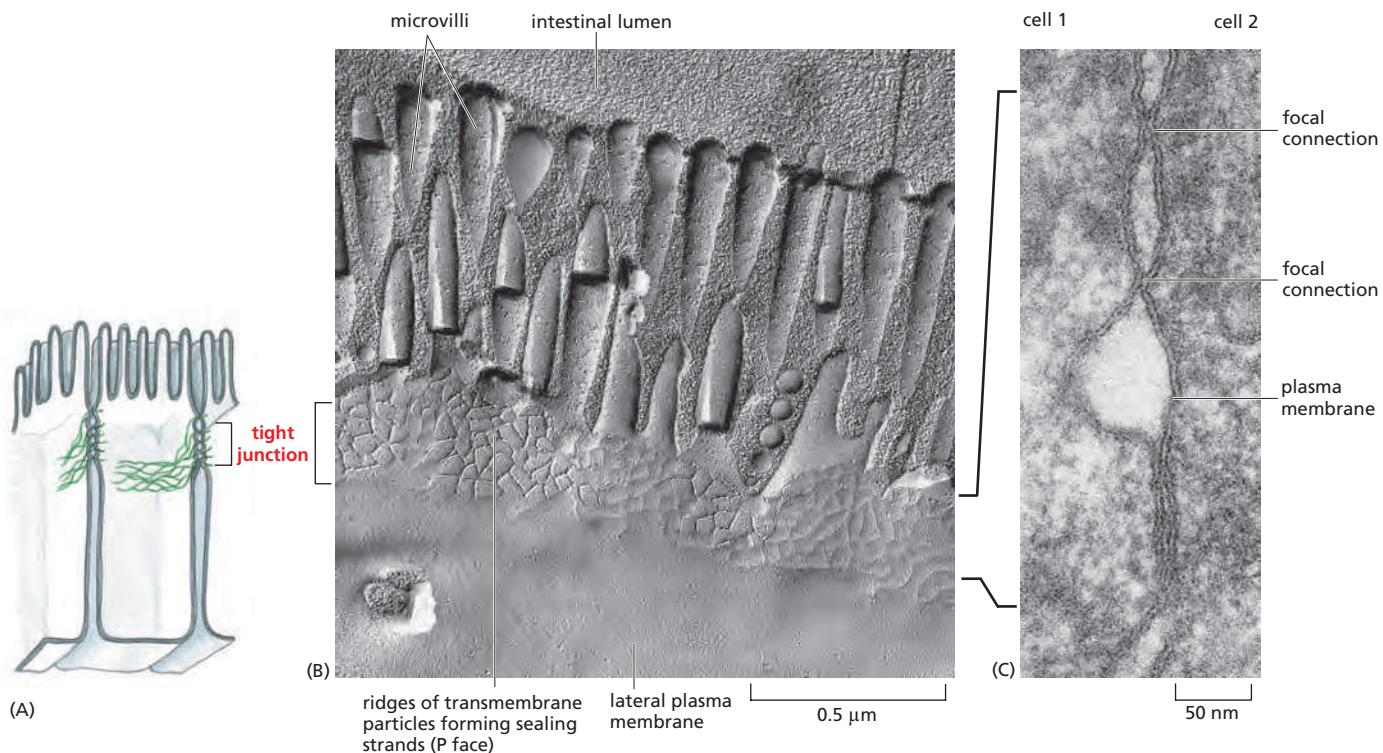


Figure 19–20 The structure of a tight junction between epithelial cells of the small intestine. The junctions are shown (A) schematically, (B) in a freeze-fracture electron micrograph, and (C) in a conventional electron micrograph. In (B), the plane of the micrograph is parallel to the plane of the membrane, and the tight junction appears as a band of branching sealing strands that encircle each cell in the epithelium (see Figure 19–21A). 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.)

will form tight-junctional connections with one another. Normal tight junctions also contain a second major transmembrane protein called *occludin*, which is not essential for the assembly or structure of the tight junction but is important for limiting junctional permeability. A third transmembrane protein, *tricellulin*, 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 kidney tubules 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 Organize Junctional Protein Complexes

Like the cadherin molecules of an adherens junction, the claudins and occludins of a tight junction interact with each other on their extracellular sides to promote junction assembly. Also as in adherens junctions, the organization of adhesion proteins in a tight junction depends on additional proteins that bind the cytoplasmic side of the adhesion proteins. The key organizational proteins at tight junctions are the *zonula occludens (ZO)* proteins. The three major members of the ZO family—ZO-1, ZO-2, and ZO-3—are large **scaffold proteins** that provide a structural support on which the tight junction is built. These intracellular molecules

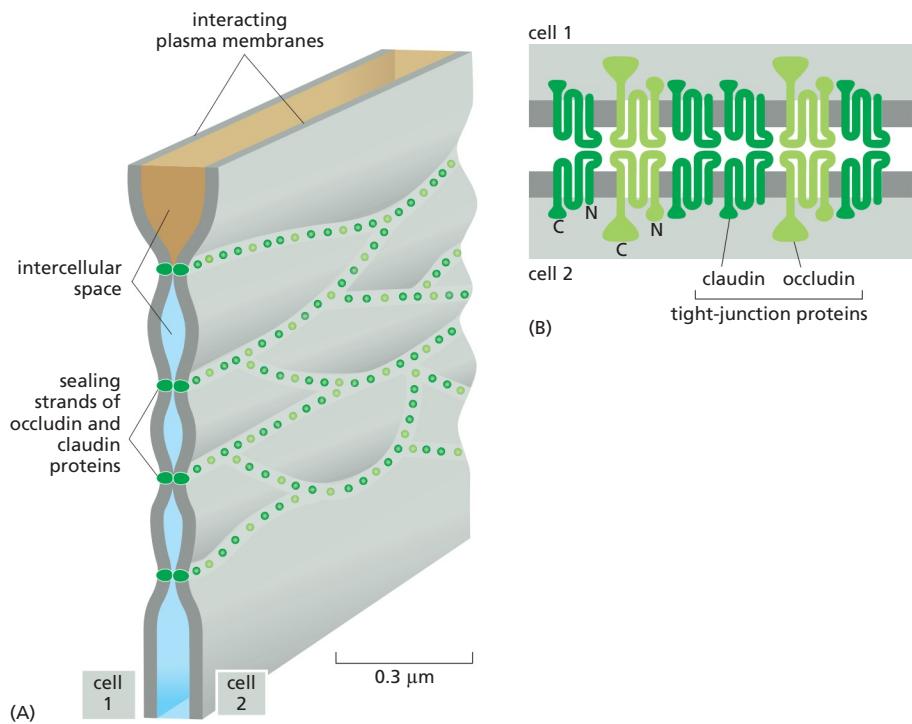


Figure 19-21 A model of a tight junction. (A) The sealing strands hold adjacent plasma membranes together. The strands are composed of transmembrane proteins that make contact across the intercellular space and create a seal. (B) The molecular composition of a sealing strand. The major extracellular components of the tight junction are members of a family of proteins with four transmembrane domains. One of these proteins, claudin, is the most important for the assembly and structure of the sealing strands, whereas the related protein occludin has the less critical role of determining junction permeability. The two termini of these proteins are both on the cytoplasmic side of the membrane, where they interact with large scaffolding proteins that organize the sealing strands and link the tight junction to the actin cytoskeleton (not shown here, but see Figure 19-22).

consist of strings of protein-binding domains, typically including several **PDZ domains**—segments about 80 amino acids long that can recognize and bind the C-terminal tails of specific partner proteins (Figure 19-22). One domain of these scaffold proteins can attach to a claudin protein, while others can attach to occludin or the actin cytoskeleton. Moreover, one molecule of scaffold protein can bind to another. In this way, the cell can assemble a mat of intracellular proteins that organizes and positions the sealing strands of the tight junction.

The tight-junctional network of sealing strands usually lies just apical to adherens and desmosome junctions that bond the cells together mechanically; the whole assembly is called a *junctional complex* (see Figure 19-2). The parts of this junctional complex depend on each other for their formation. For example, anti-cadherin antibodies that block the formation of adherens junctions also block the formation of tight junctions.

Gap Junctions Couple Cells Both Electrically and Metabolically

Tight junctions block the passageways through the gaps between epithelial 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 channels from the

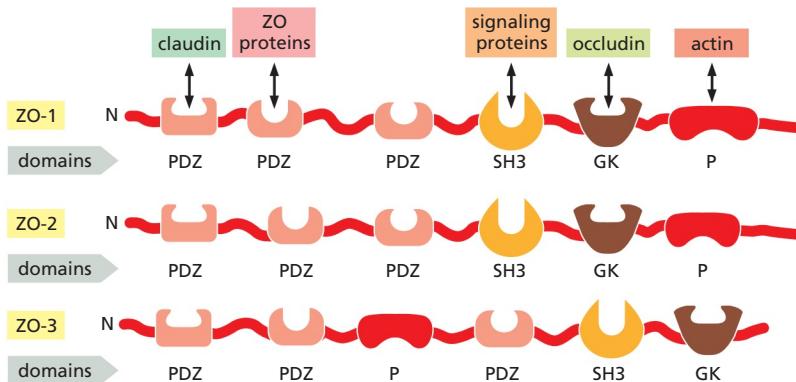


Figure 19-22 Scaffold proteins at the tight junction. The scaffold proteins ZO-1, ZO-2, and ZO-3 are concentrated beneath the plasma membrane at tight junctions. Each of the proteins contains multiple protein-binding domains, including three PDZ domains, an SH3 domain, and a GK domain, linked together like beads on a flexible string. These domains enable the proteins to interact with each other and with numerous other partners, as indicated here, to generate a tightly woven protein network that organizes the sealing strands of the tight junction and links them to the actin cytoskeleton. Scaffold proteins with similar structure help organize other junctional complexes, including those at neural synapses.

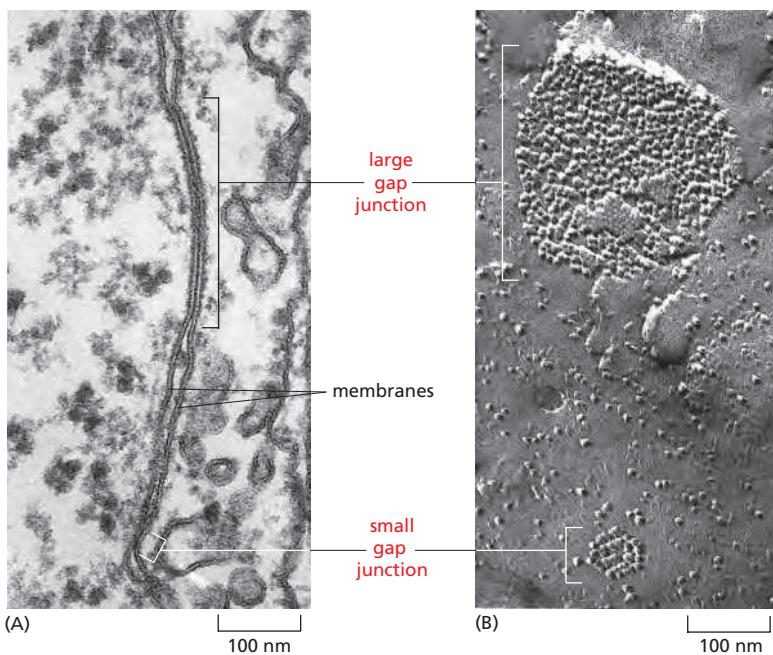


Figure 19-23 Gap junctions as seen in the electron microscope. (A) Thin-section and (B) freeze-fracture electron micrographs of a large and a small gap-junction plaque between fibroblasts in culture. In (B), each gap junction is seen as a cluster of homogeneous intramembrane particles. Each intramembrane particle corresponds to a connexon (see Figure 19-25). (From N.B. Gilula, in *Cell Communication* [R.P. Cox, ed.], pp. 1–29. New York: Wiley, 1974.)

cytoplasm of one to that of the other. These channels are called **gap junctions**.

Gap junctions are present in most animal tissues, including connective tissues as well as epithelia and heart muscle. 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 (Figure 19-23). The gap is spanned by channel-forming proteins, of which there are two distinct families, called the *connexins* and the *innexins*. Connexins are the predominant gap-junction proteins in vertebrates, with 21 isoforms in humans. Innexins are found in the gap junctions of invertebrates.

Gap junctions have a pore size of about 1.4 nm, which allows the exchange of inorganic ions and other small water-soluble molecules, but not of macromolecules such as proteins or nucleic acids (Figure 19-24). An electric current injected into one cell through a microelectrode causes an electrical disturbance in the neighboring cell, due to the flow of ions carrying electric charge through gap junctions. This electrical coupling via gap junctions serves an obvious purpose in tissues containing electrically excitable cells: action potentials can spread rapidly from cell to cell, without the delay that occurs at chemical synapses. In vertebrates, for example, 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.

A Gap-Junction Connexon Is Made 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 19-25). A gap junction consists of many such connexon pairs in parallel, forming a sort of molecular sieve. Not only does this sieve provide a communication channel between cells, but it also provides a form of cell-cell adhesion that supplements the cadherin- and claudin-mediated adhesions we discussed earlier.

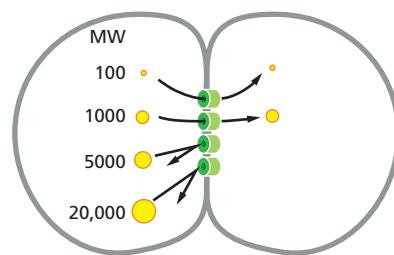


Figure 19-24 Determining the size of a gap-junction channel. When fluorescent molecules of various sizes are injected into one of two cells coupled by gap junctions, molecules with a molecular weight (MW) of less than about 1000 daltons can pass into the other cell, but larger molecules cannot. Thus, the coupled cells share their small molecules (such as inorganic ions, sugars, amino acids, nucleotides, vitamins, and the intracellular signaling molecules cyclic AMP and inositol trisphosphate) but not their macromolecules (proteins, nucleic acids, and polysaccharides).

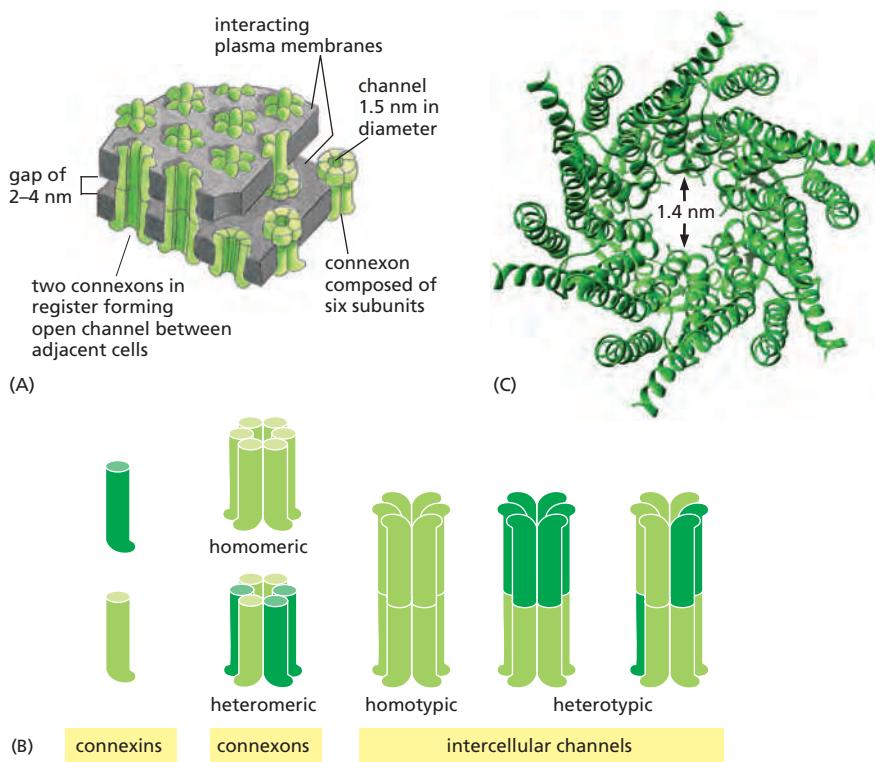


Figure 19–25 Gap junctions. (A) A drawing of the interacting plasma membranes of two adjacent cells connected by gap junctions. Each lipid bilayer is shown as a pair of red sheets. Protein assemblies called connexons (green), each of which is formed by six connexin subunits, penetrate the apposed lipid bilayers. 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. (C) The high-resolution structure of a homomeric gap-junction channel, determined by x-ray crystallography of human connexin 26. In this view, we are looking down on the pore, formed from six connexin subunits. The structure illustrates the general features of the channel and suggests a pore size of about 1.4 nm, as predicted from studies of gap-junction permeability with molecules of various sizes (see Figure 19–24). (PDB code: 2ZW3.)

Gap junctions in different tissues can have different properties because they are formed from different combinations of connexins, creating channels that differ in permeability and regulation. 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 connexins can form intercellular channels in which the two aligned half-channels are different (see Figure 19–25B).

Like conventional ion channels (discussed in Chapter 11), individual gap-junction channels do not remain open all the time; instead, they flip between open and closed states. These changes are triggered by a variety of stimuli, including the voltage difference between the two connected cells, the membrane potential of each cell, and various chemical properties of the cytoplasm, including the pH and concentration of free Ca^{2+} . Some subtypes of gap junctions can also be regulated by extracellular signals such as neurotransmitters. We are only just beginning to understand the physiological functions and structural basis of these various gating mechanisms.

Each gap-junctional plaque is a dynamic structure that can readily assemble, disassemble, or be remodeled, and it can contain a cluster of a few to many thousands of connexons (see Figure 19–23B). 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 19–26). This turnover is rapid: the connexin molecules have a half-life of only 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 connexon 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 in a closed conformation, preventing the cell from losing its small molecules by

Figure 19–26 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 to certain small fluorescent dye molecules, which 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 to label all the connexin molecules in the cells, 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 gap junctions between pairs of cells treated 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 previous 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 the old ones. (From G. Gaietta et al., *Science* 296:503–507, 2002. With permission from AAAS.)

leakage through them. But there is also evidence that in some circumstances they can open and serve as channels for the release of small signal molecules.

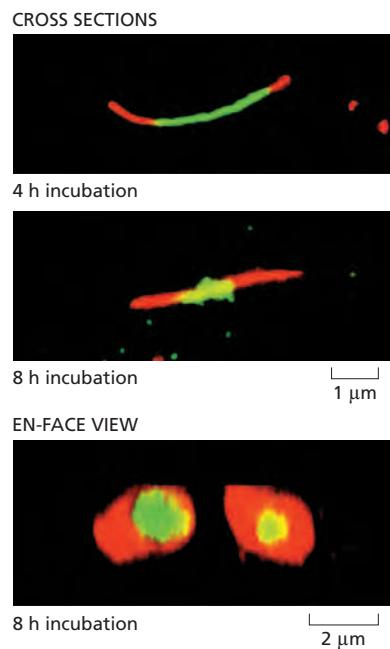
In Plants, Plasmodesmata Perform Many of the Same Functions as Gap Junctions

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 polysaccharides, 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 19–27A**, 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 (ER) in each of the connected cells (Figure 19–27B–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 removed.

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



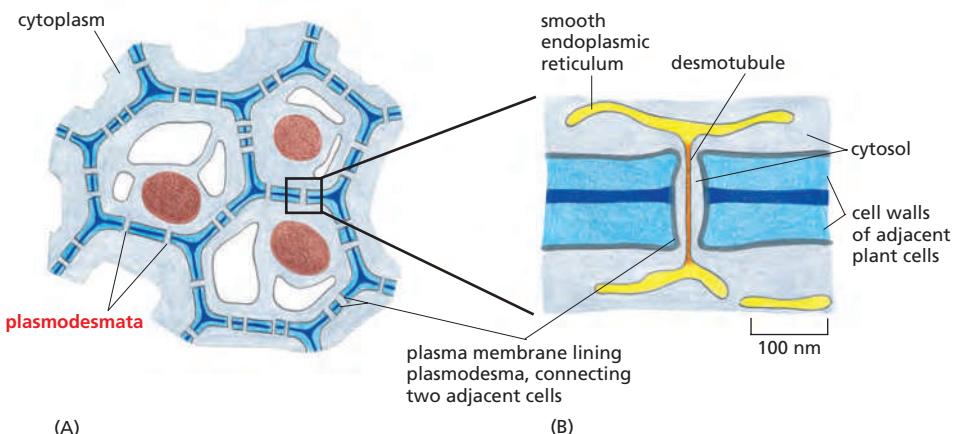
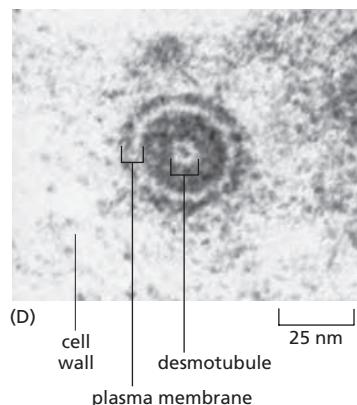
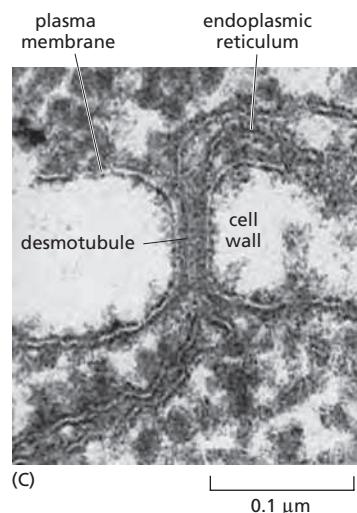


Figure 19–27 Plasmodesmata. (A) The cytoplasmic channels of plasmodesmata pierce the plant cell wall and connect cells in a plant together. (B) Each plasmodesma is lined with plasma membrane that is common to the two connected cells. It usually also contains a fine tubular structure, the desmotubule, 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* 111, pp. 134–150. Heidelberg: Springer-Verlag, 1982.)



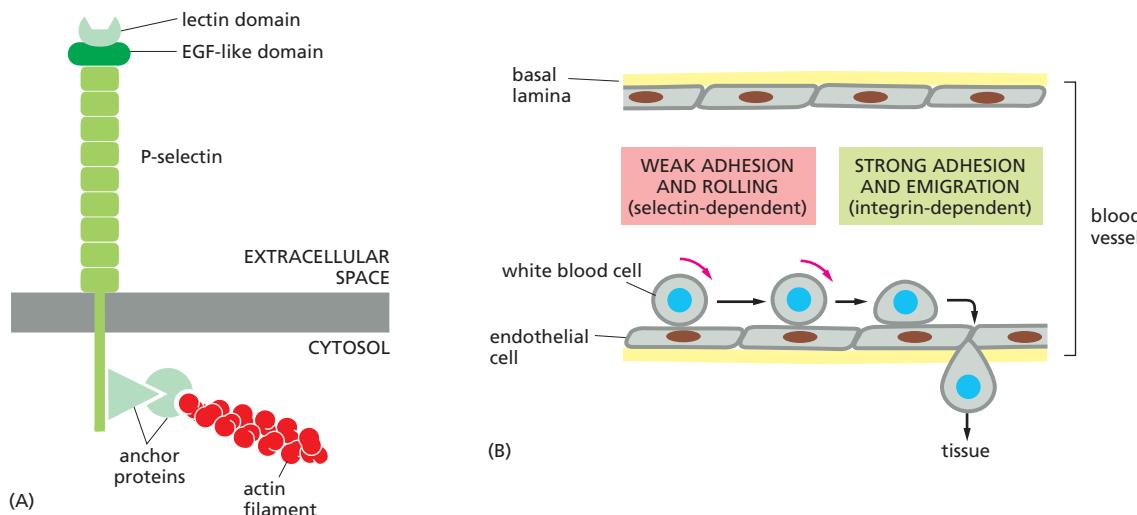
is similar to the molecular-weight cutoff for gap junctions. As with gap junctions, transport through plasmodesmata is regulated. Dye-injection experiments, for example, show that there can be barriers to the movement of even low-molecular-weight molecules between certain cells, or groups of cells, that are connected by apparently normal plasmodesmata; the mechanisms that restrict communication in these cases are not understood.

Selectins Mediate Transient Cell–Cell Adhesions in the Bloodstream

We now complete our overview of cell-cell junctions and adhesion by briefly describing some of the more specialized adhesion mechanisms used in some tissues. In addition to those we have already discussed, at least three other superfamilies of cell-cell adhesion proteins are 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 adhesion 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 governing the traffic of white blood cells into normal lymphoid organs and any inflamed tissues. 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 19–28A). 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 response, and *E-selectin* on activated endothelial cells. In a lymphoid organ, such as a lymph



node or the spleen, 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 specific Ig-family proteins.

Selectins and integrins act in sequence to let white blood cells leave the bloodstream and enter tissues (Figure 19–28B). 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 specific macromolecules 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 bloodstream into the tissue by crawling out of the blood vessel between adjacent endothelial cells.

Members of the Immunoglobulin Superfamily 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 extracellular Ig-like domains that are characteristic of antibody molecules. 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 19–29). Some forms of NCAM carry an unusually large

Figure 19–28 The structure and function of selectins. (A) The structure of P-selectin. The selectin attaches to the actin cytoskeleton through adaptor 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 and rolls along the vessel wall. Then the white blood cell activates a cell-surface integrin called LFA1, which binds to a protein called ICAM1 (belonging to the Ig superfamily) on the membrane of the endothelial cell. The white blood cell adheres to the vessel wall and then crawls out of the vessel by a process that requires another immunoglobulin superfamily member called PECAM1 (or CD31), not shown (Movie 19.2). EGF, epidermal growth factor.

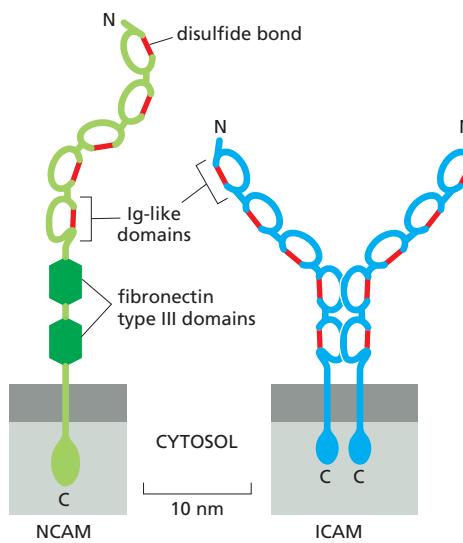


Figure 19–29 Two members of the Ig superfamily of cell-cell adhesion molecules. NCAM is expressed on neurons and many other cell types, and mediates homophilic binding. ICAM is expressed on endothelial cells and some other cell types and binds heterophilically to an integrin on white blood cells. Both NCAM and ICAM are glycoproteins, but their attached carbohydrate chains are not shown.

quantity of sialic acid (with chains containing hundreds of repeating sialic acid units). By virtue of their negative charge, the long polysialic acid chains can interfere with cell adhesion (because like charges repel one another); thus, these forms of NCAM can 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 in its environment. Although cadherins and Ig superfamily 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 cells 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.

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 called cadherins, which are anchored intracellularly to the cytoskeleton. 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.

The “classical” cadherins at adherens junctions are linked to the actin cytoskeleton by intracellular adaptor 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 detection of and response to tension and other regulatory signals at the junction.

Tight junctions seal the gaps between cells in epithelia, creating a barrier to the diffusion of molecules across the cell sheet and also helping to separate the populations of proteins in the apical and basolateral plasma membrane domains of the epithelial cell. Claudins are the major transmembrane proteins forming gap junctions. Intracellular scaffold proteins organize the claudins and other junctional proteins into a complex protein network that is linked to the actin cytoskeleton.

The cells of many animal tissues are coupled by gap junctions, which take the form of plaques of clustered connexons, which usually allow molecules smaller than about 1000 daltons to pass directly from the inside of one cell to the inside of the next. Cells connected by gap junctions share many of their inorganic ions and other small molecules and are therefore chemically and electrically coupled.

Three additional classes of transmembrane adhesion proteins mediate more transient cell-cell adhesion: selectins, immunoglobulin (Ig) superfamily members, and integrins. Selectins are expressed on white blood cells, blood platelets, and endothelial cells; they bind heterophilically to carbohydrate groups on cell surfaces, helping to mediate the adhesive interactions between these cells. Ig superfamily proteins also play a part in these interactions, 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 specific Ig superfamily proteins.

THE EXTRACELLULAR MATRIX OF ANIMALS

Tissues are not made up solely of cells. They also contain a remarkably complex and intricate network of macromolecules constituting the *extracellular matrix*. This matrix is composed of many different proteins and polysaccharides that are secreted locally and assembled into an organized meshwork in close association with the surfaces of the cells that produce them.

The classes of macromolecules constituting the extracellular matrix in different animal tissues are broadly similar, 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 describe the major features of the extracellular matrix in animal tissues, with an emphasis on vertebrates. We begin with an overview of the major classes of macromolecules in the matrix, after which we turn to the structure and function of the *basal lamina*, the thin layer of specialized extracellular matrix that lies beneath all epithelial cells. In the sections that follow, we then describe the varied types of cell-matrix junctions through which cells are connected to the matrix.

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 by cells called **fibroblasts** (Figure 19–30). 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 extracellular matrix is constructed from three major classes of macromolecules: (1) glycosaminoglycans (GAGs), which are large and highly charged polysaccharides that are usually covalently linked to protein in the form of *proteoglycans*; (2) fibrous proteins, which are primarily members of the *collagen* family; and (3) a large class of noncollagen *glycoproteins*, which carry conventional asparagine-linked oligosaccharides (described in Chapter 12). All three classes of macromolecule have many members and come in a great variety of shapes and

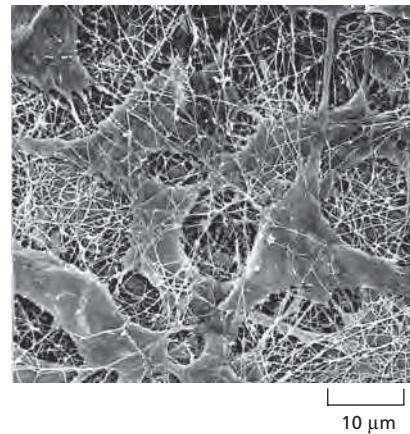


Figure 19–30 Fibroblasts in connective tissue. This scanning electron micrograph shows tissue from the cornea of a rat. The extracellular matrix surrounding the fibroblasts is here composed largely of collagen fibrils. The glycoproteins, hyaluronan, and proteoglycans, which normally form a hydrated gel filling the interstices of the fibrous network, have been removed by enzyme and acid treatment. (Courtesy of T. Nishida.)

sizes (Figure 19–31). Mammals are thought to have almost 300 matrix proteins, including about 36 proteoglycans, about 40 collagens, and over 200 glycoproteins, which usually contain multiple subdomains and self-associate to form multimers. Add to this the large number of matrix-associated proteins and enzymes that can modify matrix behavior by cross-linking, degradation, or other mechanisms, and one begins to see that the matrix is an almost infinitely variable material. Each tissue contains its own unique blend of matrix components, resulting in an extracellular matrix that is specialized for the needs of that tissue.

The proteoglycan molecules in connective tissue typically form a highly hydrated, gel-like “ground substance” in which collagens and glycoproteins 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, the many matrix glycoproteins help cells migrate, settle, and differentiate in the appropriate locations.

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

Glycosaminoglycans (GAGs) are unbranched polysaccharide chains composed of repeating disaccharide units. 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 19–32). 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 into compact globular structures, and they are strongly hydrophilic. Thus, GAGs tend to adopt highly extended conformations that occupy a huge volume relative to their mass (Figure 19–33), and they form hydrated gels even at very low concentrations. The weight of GAGs in connective tissue is usually less than 10% of the weight of proteins, but 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, a prematurely aged appearance, and generalized defects in their skin, joints, muscles, and bones.

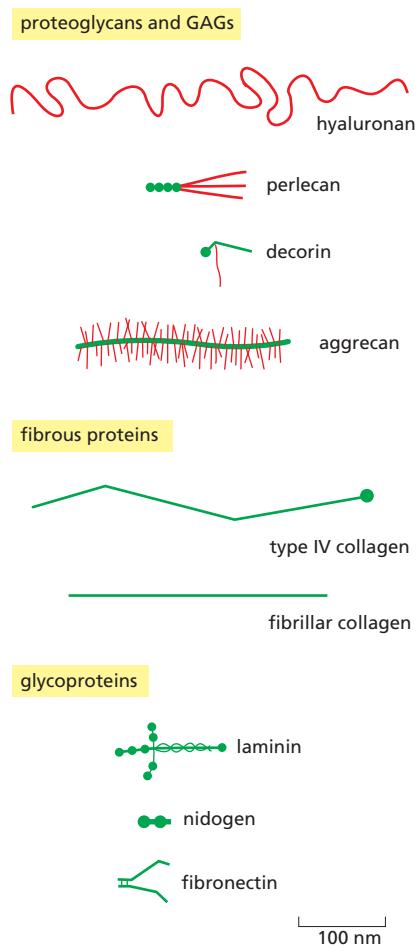


Figure 19–31 The comparative shapes and sizes of some of the major extracellular matrix macromolecules. Protein is shown in green, and glycosaminoglycan (GAG) in red.

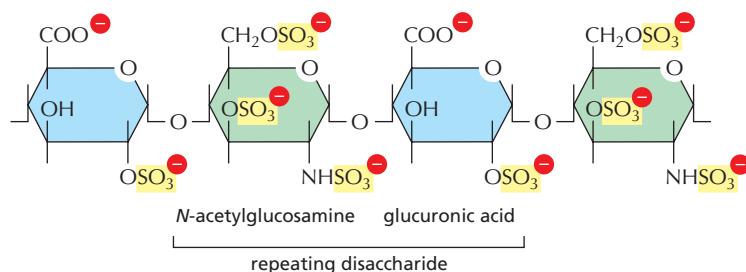


Figure 19–32 The repeating disaccharide sequence of a heparan sulfate glycosaminoglycan (GAG) chain. These chains can consist of as many as 200 disaccharide units, but are typically less than half that size. There is a high density of negative charges along the chain due to the presence of both carboxyl and sulfate groups. The 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%.

Hyaluronan Acts as a Space Filler During Tissue Morphogenesis and Repair

Hyaluronan (also called *hyaluronic acid* or *hyaluronate*) is the simplest of the GAGs (Figure 19–34). 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 a typical GAG because it contains no sulfated sugars, all its disaccharide units are identical, its chain length is enormous, 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. 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 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.

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 19–35). 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 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 generally contain relatively short, branched oligosaccharide chains that contribute only a small fraction of their weight, proteoglycans can contain as much as

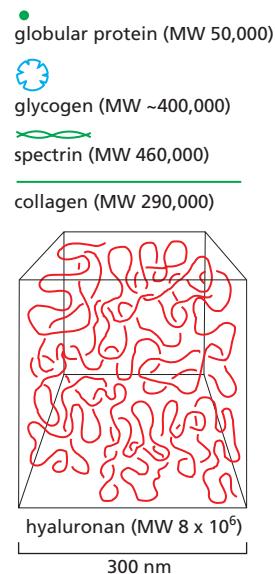


Figure 19–33 The relative dimensions and volumes occupied by various macromolecules. Several proteins, a glycogen granule, and a single hydrated molecule of hyaluronan are shown.

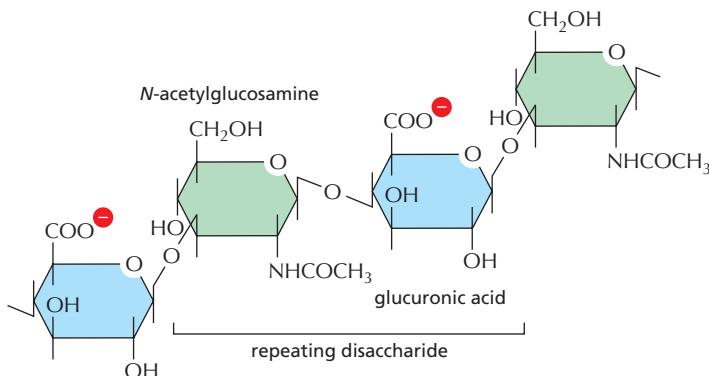
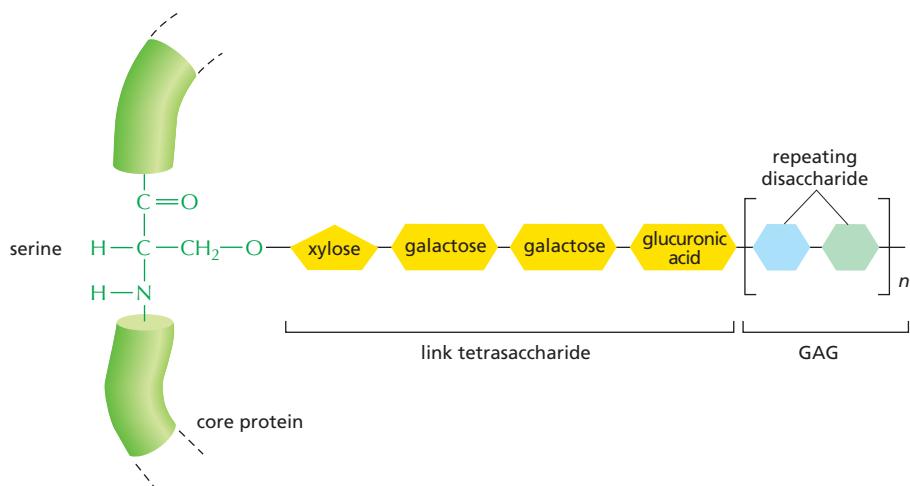


Figure 19–34 The repeating disaccharide sequence in hyaluronan, a relatively simple GAG. This ubiquitous molecule in vertebrates consists of a single long chain of up to 25,000 sugar monomers. Note the absence of sulfate groups.



95% carbohydrate by weight, mostly in the form of long, unbranched GAG chains, each typically about 80 sugars long.

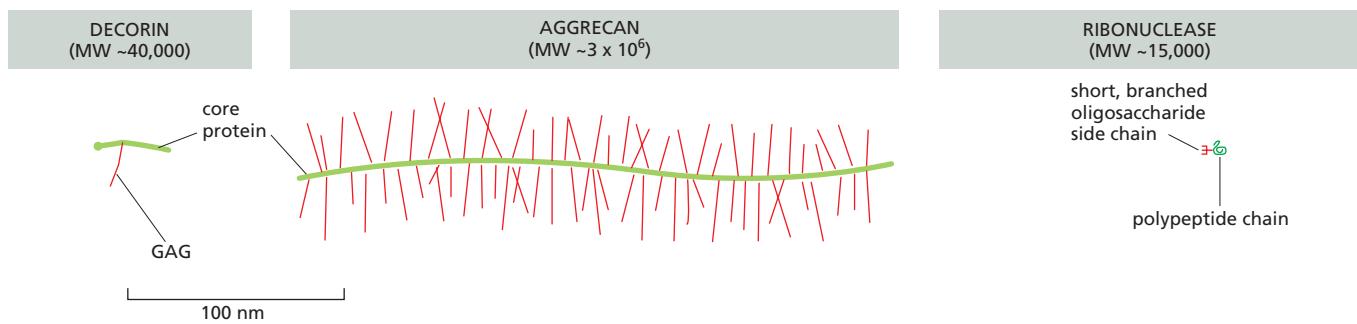
In principle, proteoglycans have the potential for almost limitless heterogeneity. Even a single type of core protein can carry highly variable numbers and types of attached GAG chains. Moreover, the underlying repeating sequence of disaccharides in each GAG can be modified by a complex pattern of sulfate 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 GAG chain (Figure 19–36). 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 19–37). 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 19–38).

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. 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

Figure 19–35 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. 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 either D-glucuronic acid or L-iduronic acid and N-acetyl-D-glucosamine; in keratan sulfate, it is D-galactose and N-acetyl-D-glucosamine.

Figure 19–36 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.



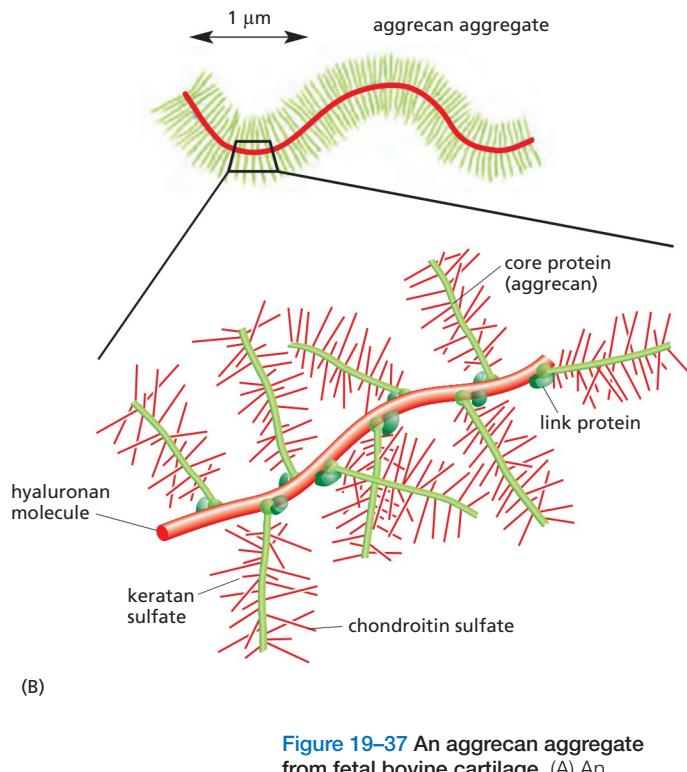
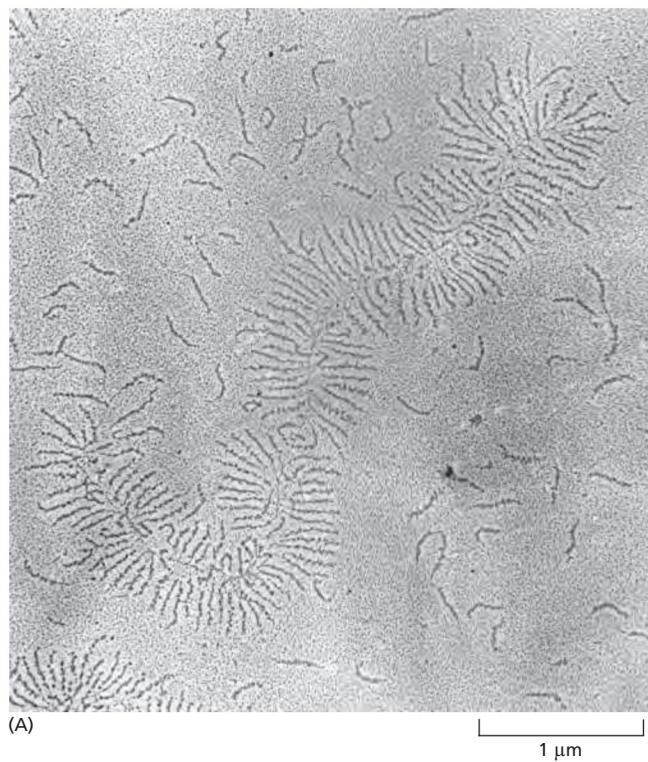


Figure 19–37 An aggrecan aggregate from fetal bovine cartilage. (A) An electron micrograph of an aggrecan aggregate shadowed with platinum. Many free aggrecan molecules are also visible. (B) A drawing of the giant aggrecan aggregate shown in (A). It consists of about 100 aggrecan monomers (each like the one shown in Figure 19–36) 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 mass of such a complex can be 10^8 daltons or more, and it occupies a volume equivalent to that of a bacterium, which is about $2 \times 10^{-12} \text{ cm}^3$. (A, courtesy of Lawrence Rosenberg.)

fibroblasts, syndecans can be found in cell-matrix adhesions, where they modulate integrin function by interacting with fibronectin on the cell surface and with cytoskeletal and signaling proteins inside the cell. As we discuss later, syndecan and other proteoglycans also interact with soluble peptide growth factors, influencing their effects on cell growth and proliferation.

Collagens Are the Major Proteins of the Extracellular Matrix

The **collagens** are a family of fibrous proteins found in all multicellular animals. They 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, collagens 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 19–39).

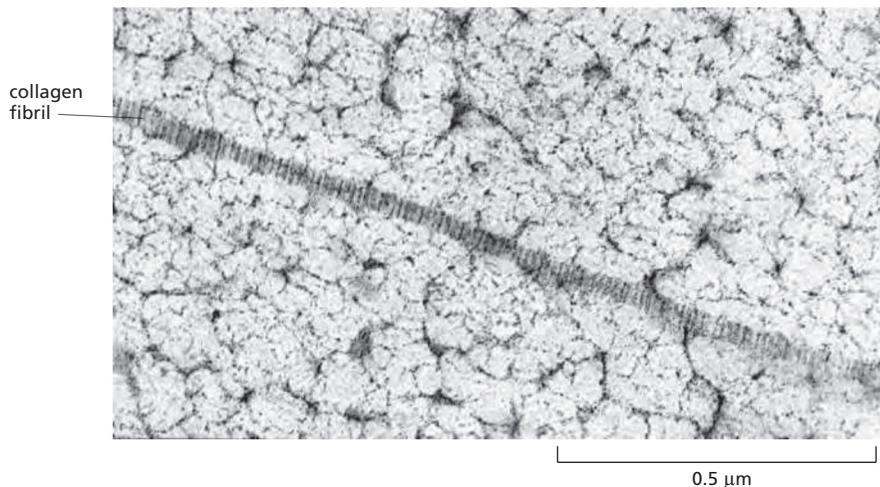


Figure 19–38 Proteoglycans in the extracellular matrix of rat cartilage. The tissue was rapidly frozen at -196°C , and fixed and stained while still frozen (a process called freeze substitution) to prevent the GAG chains from collapsing. In this electron micrograph, the proteoglycan molecules are seen 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.)

Figure 19–39 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, a form of proline that is chemically modified during collagen synthesis in the cell). (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.)

Collagens are extremely rich in proline and glycine, both of which are important in the formation of the triple-stranded helix.

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 molecules could be assembled from various combinations of the 42 α chains, only a limited number of triple-helical combinations are possible, and roughly 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: after being secreted into the extracellular space, they 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, where they are clearly visible in electron micrographs (Figure 19–40; see also Figure 19–38). 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 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 containing short collagen-like segments. These include collagen type XVII, which has a transmembrane domain and is found in hemidesmosomes, and type XVIII, 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

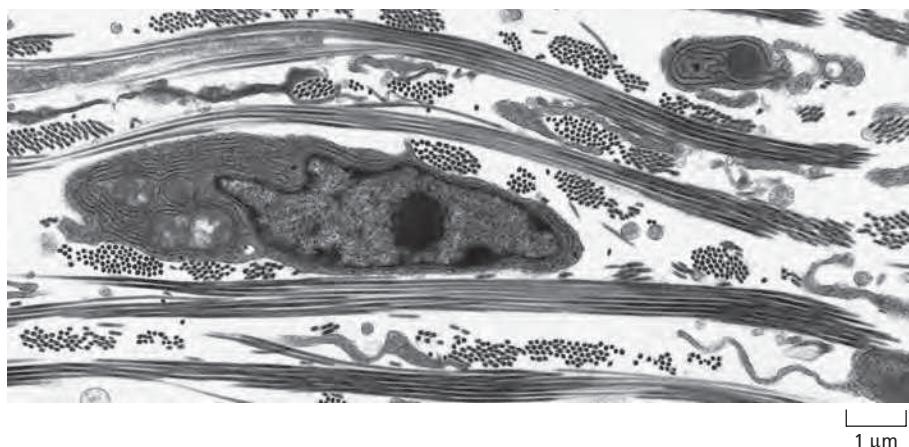
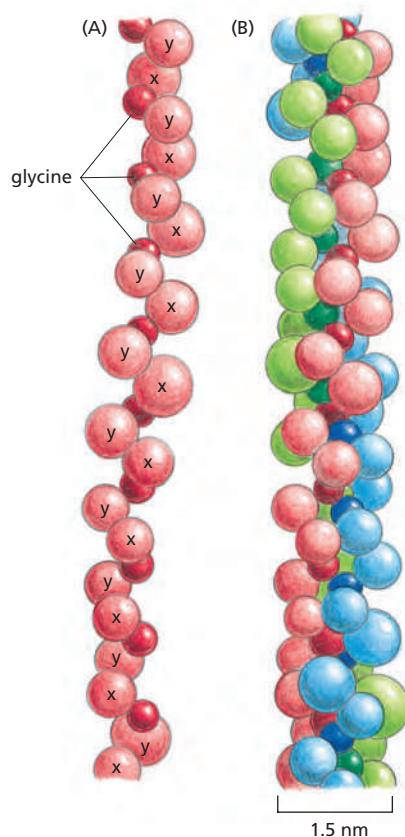


Figure 19–40 A fibroblast surrounded by collagen fibrils in the connective tissue of embryonic chick skin. In this electron micrograph, the fibrils are organized into bundles that run approximately at right angles to one another. Therefore, some bundles are oriented longitudinally, whereas others are seen in cross section. The collagen fibrils are produced by fibroblasts. (From C. Ploetz, E.I. Zychlinsky and D.E. Birk, *J. Struct. Biol.* 106:73–81, 1991. With permission from Elsevier.)

TABLE 19–2 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 (<i>osteogenesis imperfecta</i>)
	II	Fibril	Cartilage, intervertebral disc, notochord, vitreous humor of the eye	Cartilage deficiency, dwarfism (<i>chondrodysplasia</i>)
	III	Fibril	Skin, blood vessels, internal organs	Fragile skin, loose joints, blood vessels prone to rupture (<i>Ehlers-Danlos syndrome</i>)
	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	Nonfibrillar	Hemidesmosomes	Skin blistering
Proteoglycan core protein	XVIII	Nonfibrillar	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, XVII, and XVIII are composed of only one type of α chain each.

the exons are 54, or multiples of 54, nucleotides long, suggesting that these collagens originated through multiple duplications of a primordial gene containing 54 nucleotides and encoding exactly six Gly-X-Y repeats (see Figure 19–39).

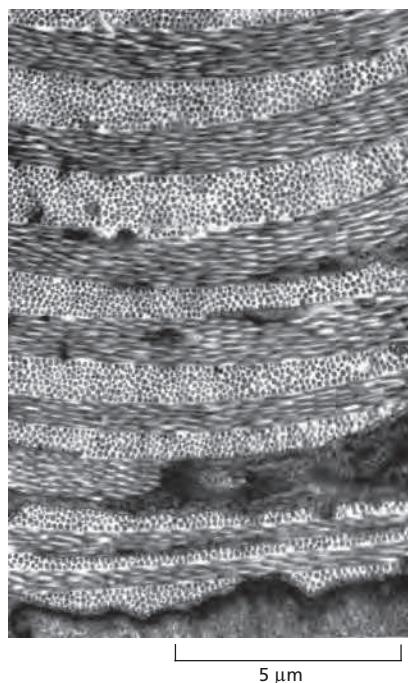
Table 19–2 provides additional details for some of the collagen types discussed in this chapter.

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 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 19–41).

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 collagen molecules. But even fibrils composed of the same mixture of collagens have different arrangements in different tissues. How is this achieved? Part of the answer is that cells can regulate the disposition of the collagen

Figure 19–41 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.)



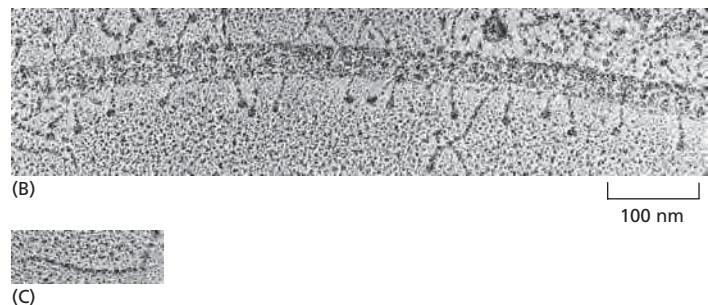
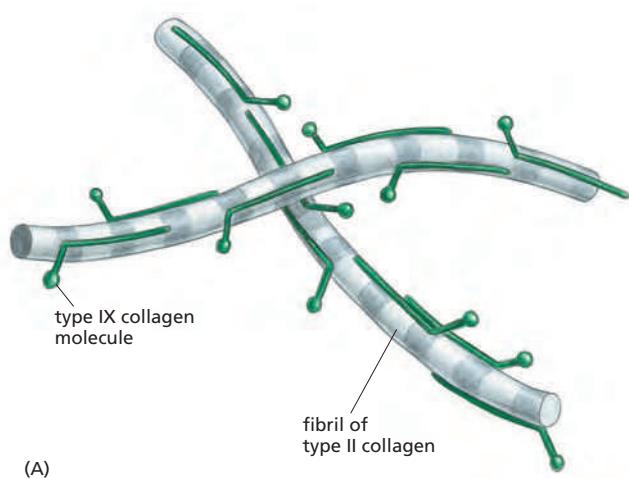


Figure 19–42 Type IX collagen. (A) Type IX collagen molecules binding in a periodic pattern to the surface of a fibril containing type II collagen. (B) Electron micrograph of a rotary-shadowed type-II-collagen-containing fibril in cartilage, decorated by type IX collagen molecules. (C) An individual 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.)

molecules after secretion by guiding collagen fibril formation near the plasma membrane. 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 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 organizing collagen fibrils. They differ from fibrillar collagens in the following ways. First, their triple-stranded helical structure is interrupted by one or two short nonhelical domains, which makes the molecules more flexible than fibrillar collagen molecules. Second, 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 19–42), 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.

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 the 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 19–43). 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 they synthesize the collagen fibrils and deposit them in the correct orientation. Then they work 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 surround and bind together most organs.

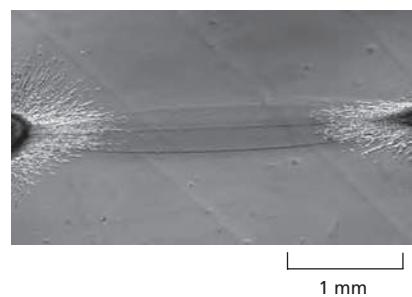


Figure 19–43 The shaping of the extracellular matrix by cells. This micrograph shows a region between two pieces of embryonic chick heart (rich in fibroblasts as well as heart muscle cells) that were cultured on a collagen gel for 4 days. A dense tract of aligned collagen fibers has formed between the 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.)

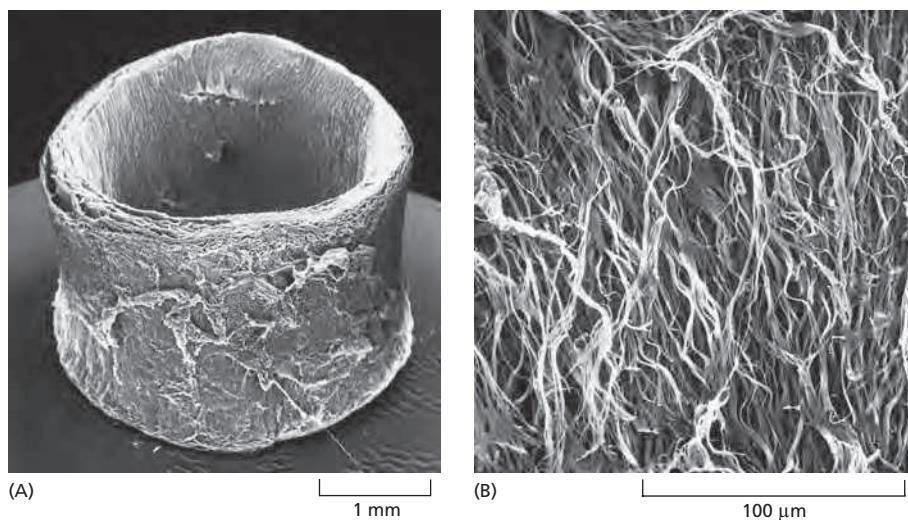


Figure 19–44 Elastic fibers. These scanning electron micrographs show (A) a low-power view of a segment of a dog's aorta and (B) a high-power view of the dense network of longitudinally oriented elastic fibers in the outer layer of the same blood vessel. All the other components have been digested away with enzymes and formic acid. (From K.S. Haas et al., *Anat. Rec.* 230:86–96, 1991. With permission from Wiley-Liss.)

Elastin Gives Tissues Their Elasticity

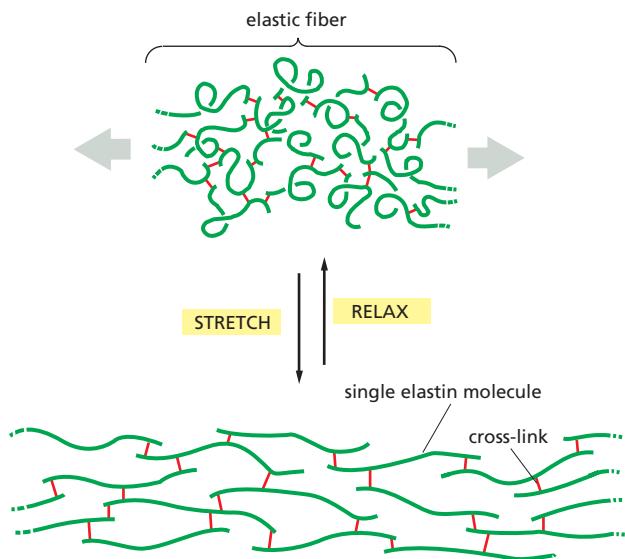
Many vertebrate tissues, such as skin, blood vessels, and lungs, need to be both strong and elastic in order to function. A network of **elastic fibers** in the extracellular matrix of these tissues gives them the resilience to recoil after transient stretch (Figure 19–44). Elastic fibers are at least 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. 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.

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 are cross-linked to adjacent molecules by covalent attachment of lysine residues. 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 19–45).

Elastin is the dominant extracellular matrix protein in arteries, comprising 50% of the dry weight of the largest artery—the aorta (see Figure 19–44). Mutations in the elastin gene causing a deficiency of the protein in mice or humans result in narrowing of the aorta and 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 disorder. In the most severely affected individuals, the aorta is 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 and Other Multidomain Glycoproteins Help Organize the Matrix

In addition to proteoglycans, collagens, and elastic fibers, the extracellular matrix contains a large and varied assortment of glycoproteins that typically have multiple domains, each with specific binding sites for other matrix macromolecules and for receptors on the surface of cells (Figure 19–46). 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

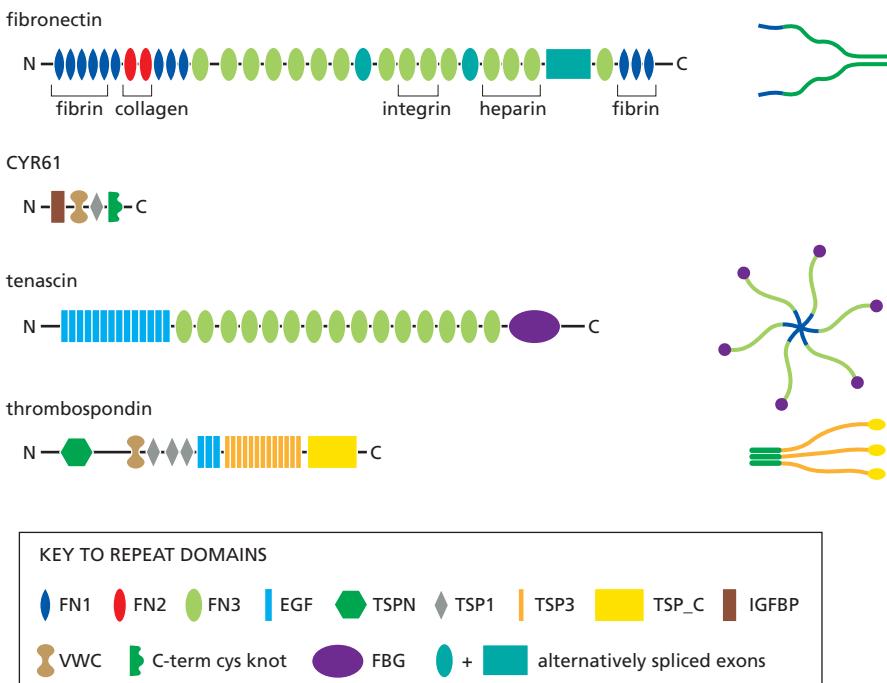
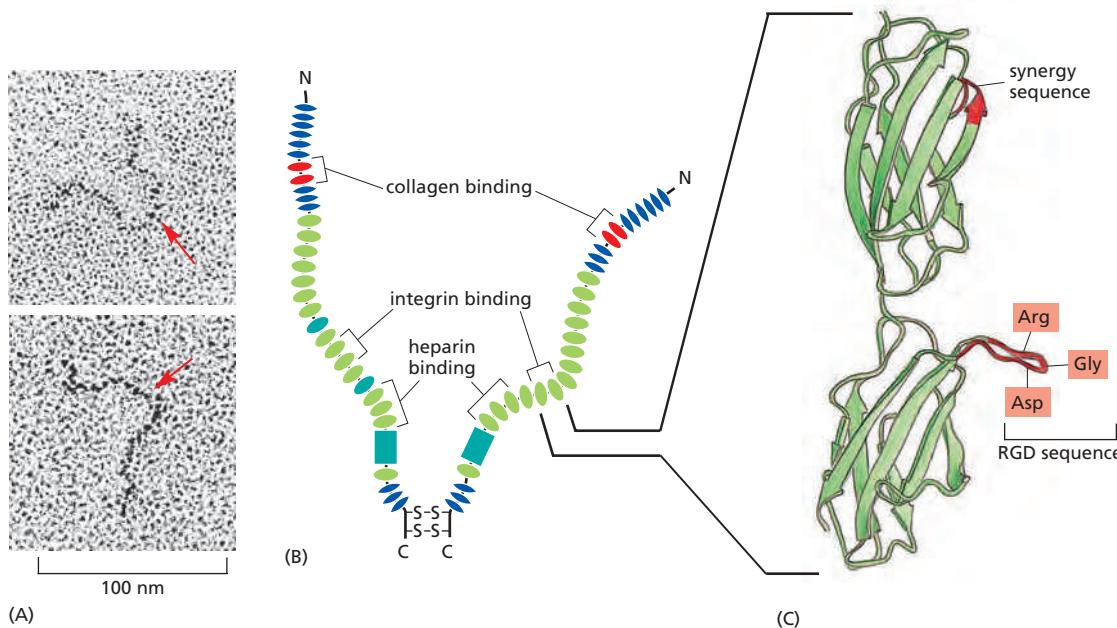


Figure 19–45 Stretching a network of elastin molecules. The molecules are joined together by covalent bonds (red) to generate a cross-linked network. In this model, each elastin molecule in the network can extend and contract in a manner resembling a random coil, so that the entire assembly can stretch and recoil like a rubber band.

Figure 19–46 Complex glycoproteins of the extracellular matrix. Many matrix glycoproteins are large scaffold proteins containing multiple copies of specific protein-interaction domains. Each domain is folded into a discrete globular structure, and many such domains are arrayed along the protein like beads on a string. This diagram shows four representative proteins among the roughly 200 matrix glycoproteins that are found in mammals. Each protein contains multiple repeat domains, with the names listed in the key at the bottom. Fibronectin, for example, contains numerous copies of three different *fibronectin repeats* (types I–III, labeled here as FN1, FN2, and FN3). Two type III repeats near the C-terminus contain important binding sites for cell-surface integrins, whereas other FN repeats are involved in binding fibrin, collagen, and heparin, as indicated (see Figure 19–47). Other matrix proteins contain repeated sequences resembling those of epidermal growth factor (EGF), a major regulator of cell growth and proliferation; these repeats might serve a similar signaling function in matrix proteins. Other proteins contain domains, such as the insulin-like growth factor-binding protein (IGFBP) repeat, that bind and regulate the function of soluble growth factors. To add more structural diversity, many of these proteins are encoded by RNA transcripts that can be spliced in different ways, adding or removing exons, such as those in fibronectin. Finally, the scaffolding and regulatory functions of many matrix proteins are further expanded by assembly into multimeric forms, as shown at the right: fibronectin forms dimers linked at the C-termini, whereas tenascin and thrombospondin form N-terminally linked hexamers and trimers, respectively. Other domains include four repeats from thrombospondin (TSPN, TSP1, TSP3, TSP_C), VWC, von Willebrand type C; FBG, fibrinogen-like. (Adapted from R.O. Hynes and A. Naba, *Cold Spring Harb. Perspect. Biol.* 4:a004903, 2012.)



as tracks along which cells can migrate or as repellents that keep cells out of forbidden areas. They can also bind and thereby influence the function of peptide growth factors and other small molecules produced by nearby cells.

The best-understood member of this class of matrix proteins is **fibronectin**, a large glycoprotein found in all vertebrates and important for many cell-matrix interactions. 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 their C-terminal ends. Each subunit contains a series of small repeated domains, or modules, separated by short stretches of flexible polypeptide chain (**Figure 19-47**). Each domain is usually encoded by a separate exon, suggesting that the fibronectin gene, like the genes encoding many matrix proteins, 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 multiple fibronectin isoforms (see Figure 19-46). The major repeat domain in fibronectin is called the **type III fibronectin repeat**, which is about 90 amino acids long and occurs at least 15 times in each subunit. This repeat is among the most common of all protein domains in vertebrates.

Fibronectin Binds to Integrins

One way to analyze a complex multifunctional protein molecule such as fibronectin is to synthesize individual regions of the protein and test their ability to bind other proteins. By these and other methods, it was possible to show that one region of fibronectin binds to collagen, another to proteoglycans, and another to specific integrins on the surface of various types of cells (see Figure 19-47B). Synthetic peptides corresponding to different segments of the integrin-binding domain were then used to show that binding depends on a specific tripeptide sequence (*Arg-Gly-Asp*, or **RGD**) that is found in one of the type III repeats (see Figure 19-47C). 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.

Several extracellular proteins besides fibronectin also have an RGD sequence that mediates cell-surface binding. Many of these proteins are components of the extracellular matrix, while others are involved in blood clotting. Peptides

Figure 19-47 The structure of a fibronectin dimer. (A) Electron micrographs of individual fibronectin dimer molecules shadowed with platinum; red arrows mark the joined 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 multiple domains (see Figure 19-46). As indicated, some domains are specialized for binding to a particular molecule. For simplicity, not all of the known binding sites are shown. (C) The three-dimensional structure of the ninth and tenth type III fibronectin repeats, as determined by x-ray crystallography. Both the Arg-Gly-Asp (RGD) and the “synergy” sequences shown in red are important for binding to integrins on cell surfaces. (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.)

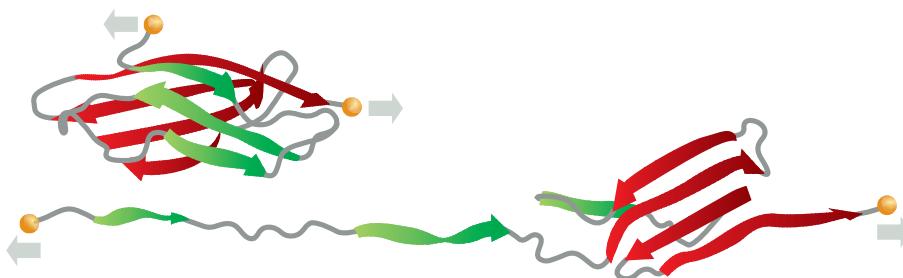


Figure 19–48 Tension-sensing by fibronectin. Some type III fibronectin repeats are thought to unfold when fibronectin is stretched. The unfolding exposes cryptic binding sites that interact with other fibronectin molecules resulting in the formation of fibronectin filaments like those shown in Figure 19–49. (From V. Vogel and M. Sheetz, *Nat. Rev. Mol. Cell Biol.* 7:265–275, 2006. With permission from Macmillan Publishers Ltd.)

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, which we describe in detail later. Each integrin 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.

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 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 cryptic binding sites in the fibronectin molecules (Figure 19–48). This allows them to bind directly to one another and to recruit additional fibronectin molecules to form a fibril (Figure 19–49). 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.

Many other extracellular matrix proteins contain multiple copies of the type III fibronectin repeat (see Figure 19–46), and it is possible that tension exerted on these proteins also uncovers cryptic binding sites and thereby influences their behavior.

The Basal Lamina Is a Specialized Form of Extracellular Matrix

Thus far in this section we have reviewed the general principles underlying the structure and function of the major classes of extracellular matrix components. We now describe how some of these components are assembled into a specialized type of extracellular matrix called the **basal lamina** (also known as the **basement membrane**). This exceedingly thin, tough, flexible sheet of matrix molecules is an essential underpinning of all epithelia. Although small 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, and it seems to have appeared very early in their evolution. The major molecular components of the basal lamina are among the most ancient extracellular matrix macromolecules.

Basal laminae are typically 40–120 nm thick. A sheet of basal lamina not only lies beneath epithelial cells but also surrounds individual muscle cells, fat cells,

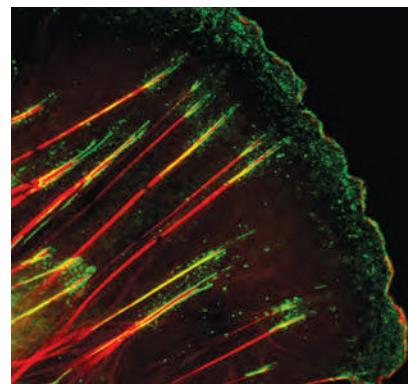
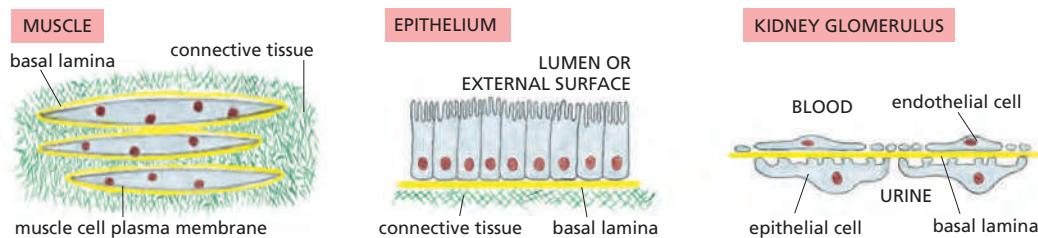


Figure 19–49 Organization of fibronectin into fibrils at the cell surface. This fluorescence micrograph shows the front end of a migrating mouse fibroblast. Extracellular fibronectin is stained green and intracellular actin filaments are stained red. The fibronectin is initially present as small dotlike aggregates near the leading edge of the cell. It accumulates at focal adhesions (sites of anchorage of actin filaments, discussed later) 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 (see Figure 19–55). 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.)



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 19–50). Basal laminae have more than simple structural and filtering roles, however. They are able to determine cell polarity; influence cell metabolism; organize the proteins 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 and Type IV Collagen Are Major Components 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 19–51). 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

Figure 19–50 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.

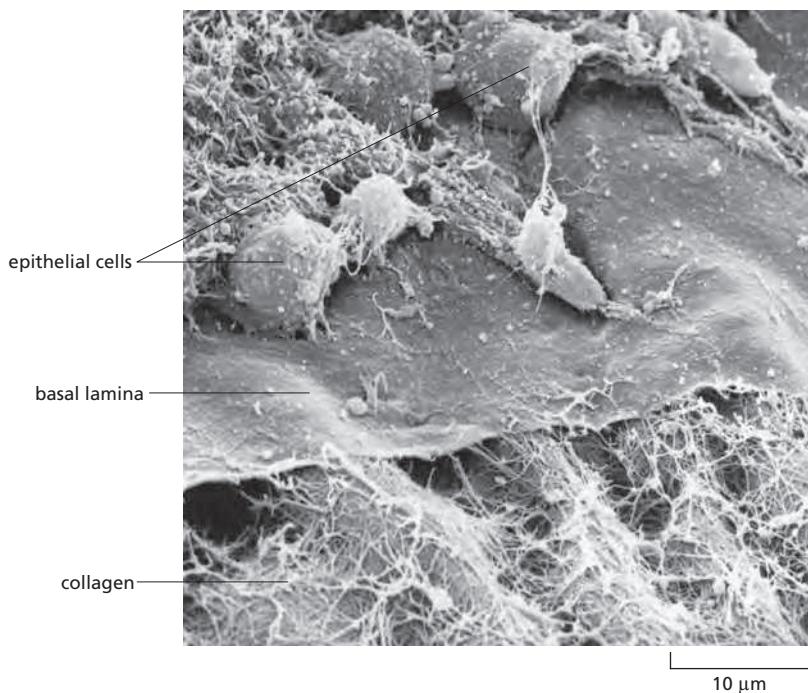
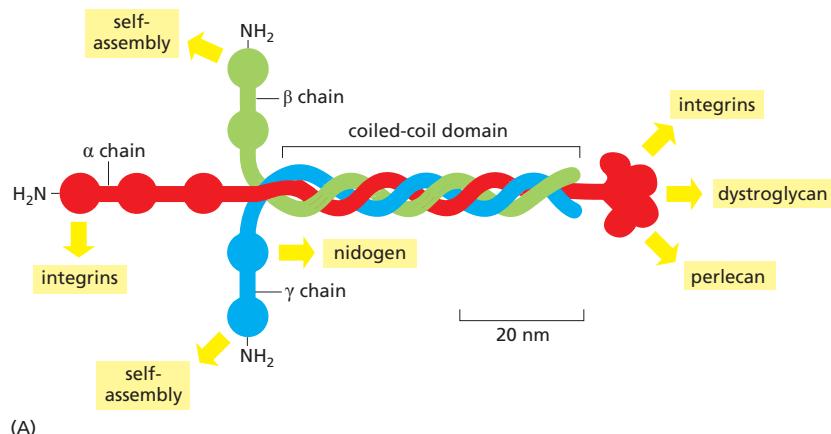


Figure 19–51 The basal lamina in the cornea of a chick embryo. In this scanning electron micrograph, some of the epithelial cells have been removed to expose the upper surface of the matlike basal lamina. A network of collagen fibrils in the underlying connective tissue interacts with the lower face of the lamina. (Courtesy of Robert Trelstad.)



typically contains the glycoproteins *laminin*, *type IV collagen*, and *nidogen*, along with the proteoglycan *perlecan*. Other common basal lamina components are fibronectin and *type XVIII collagen* (an atypical member of the collagen family, forming the core protein of a proteoglycan).

Laminin is the primary organizer of the sheet structure, and, early in development, basal laminae consist mainly of laminin molecules. Laminins comprise a large family of proteins, each composed of three 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 19–52). 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 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 $\gamma 1$ chain is, however, a component of most laminin heterotrimers; mice lacking it die during embryogenesis because they are unable to make basal laminae.

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, type IV collagen molecules consist of three separately synthesized long protein chains that twist together to form a ropelike superhelix; however, they differ from the fibrillar collagens in that the triple-stranded helical structure is interrupted in more than 20 regions, allowing multiple bends. Type IV collagen molecules interact via their terminal domains to assemble extracellularly into a flexible, feltlike network that gives the basal lamina tensile strength.

Laminin and type IV collagen interact with other basal lamina components, such as the glycoprotein *nidogen* and the proteoglycan *perlecan*, resulting in a highly cross-linked network of proteins and proteoglycans (Figure 19–53). 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 laminin. The cell-surface receptors are primarily members of the integrin family, but another important type of laminin receptor is *dystroglycan*, a proteoglycan with a core protein that spans the cell membrane, dangling its GAG 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 coordinates the assembly of the other basal lamina components.

Basal Laminae Have Diverse Functions

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 19–50). The proteoglycan in

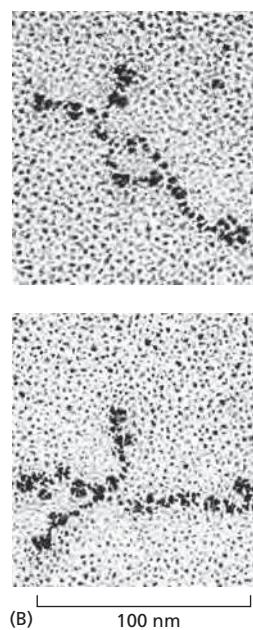
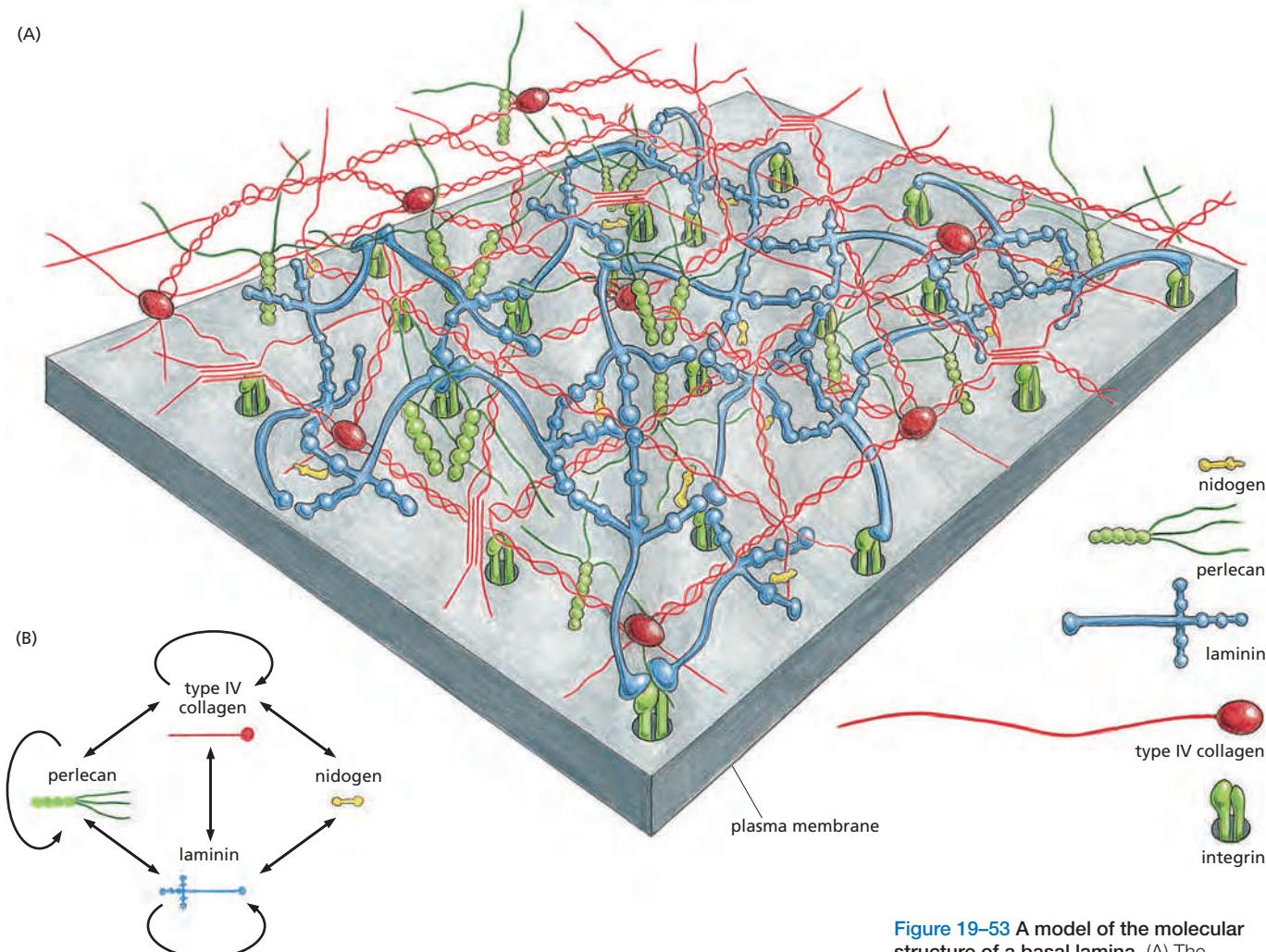


Figure 19–52 The structure of laminin. (A) The best-understood family member is laminin-111, shown here with some of its binding sites for other molecules (yellow boxes). Laminins are multidomain glycoproteins composed of three polypeptides (α , β , and γ) that are disulfide-bonded into an asymmetric crosslike structure. Each of the polypeptide chains is more than 1500 amino acids long. Five types of α chains, four types of β chains, and three types of γ chains are known, and various combinations of these subunits can assemble to form a large variety of different laminins, which are named according to numbers assigned to each of their three subunits: laminin-111, for example, contains $\alpha 1$, $\beta 1$, and $\gamma 1$ subunits. Each isoform tends to have a specific tissue distribution: laminin-332 is found in skin, laminin-211 in muscle, and laminin-411 in endothelial cells of blood vessels. Through their binding sites for other proteins, laminin molecules play a central part in organizing basal laminae 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.)

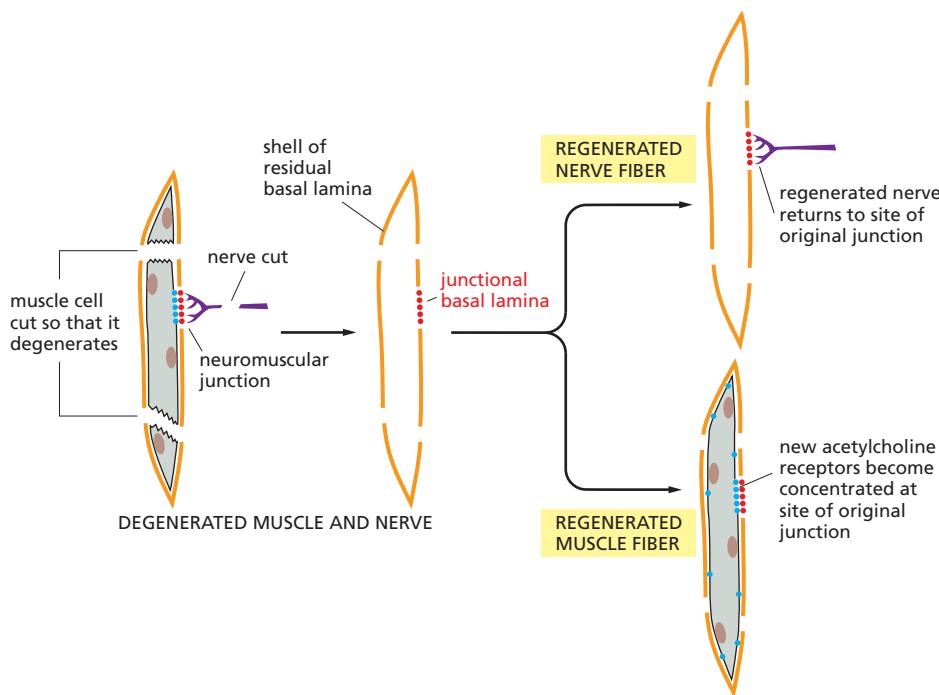


the basal lamina is important for this function: when its GAG chains are removed by specific enzymes, the filtering properties of the lamina are destroyed. Type IV collagen also has a role: in a human hereditary kidney disorder (*Alport syndrome*), mutations in a type IV collagen gene result in an irregularly 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 of 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,

Figure 19–53 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.)



with special isoforms of type IV collagen and laminin and a proteoglycan called *agrin*. After a nerve or muscle injury, the basal lamina at the synapse has a central role in reconstructing the synapse at the correct location (Figure 19–54). Defects in components of the basal lamina at the synapse are responsible for some forms of muscular dystrophy, in which muscles develop normally but then degenerate later in life.

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 remodeled so as to adapt to changes in the stresses on it.

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. Cells in connective tissues generally need to be able to stretch out in order to divide. If a cell lacks the enzymes needed to degrade the surrounding matrix, it 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 increase, and needed also when white blood cells migrate across the basal lamina of a blood vessel into tissues in response to infection or injury. Matrix degradation is important both for the spread of cancer cells through the body and for their ability to proliferate in the tissues that they invade (discussed in Chapter 20).

In general, matrix components are degraded by extracellular proteolytic enzymes (proteases) that act close to the cells that produce them. Many of these proteases belong to one of two general classes. The largest group, with about 50 members in vertebrates, is the **matrix metalloproteases**, which depend on bound Ca^{2+} or Zn^{2+} for activity. The second group is the **serine proteases**, which have a highly reactive serine in their active site. Together, metalloproteases and serine

Figure 19–54 Regeneration experiments demonstrating the special character of the junctional basal lamina at a neuromuscular junction. 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. When the nerve, but not the muscle, is allowed to regenerate (upper right), 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 right), the junctional basal lamina causes newly made acetylcholine receptors (blue) to accumulate at the original synaptic site. These experiments show that the junctional basal lamina controls the localization of synaptic components on both sides of the lamina. 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 acetylcholine 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 this molecule is likely to interact with specific ion channels on the presynaptic membrane of the neuron.

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 activities of the proteases that degrade the matrix must be tightly controlled, if the fabric of the body is not to collapse in a heap. Numerous mechanisms are therefore employed to ensure that matrix proteases are activated only at the correct time and place. Protease activity is generally confined to the cell surface by specific anchoring proteins, by membrane-associated activators, and by the production of specific protease inhibitors in regions where protease activity is not needed.

Matrix Proteoglycans and Glycoproteins Regulate the Activities of Secreted Proteins

The physical properties of extracellular matrix are important for its fundamental roles as a scaffold for tissue structure and as a substrate for cell anchorage and migration. The matrix also has an important impact on cell signaling. Cells communicate with each other by secreting signal molecules that diffuse through the extracellular fluid to influence other cells (discussed in Chapter 15). En route to their targets, the signal molecules encounter the tightly woven meshwork of the extracellular matrix, which contains a high density of negative charges and protein-interaction domains that can interact with the signal molecules, thereby altering their function in a variety of ways.

The highly charged heparan sulfate chains of proteoglycans, for example, interact with numerous secreted signal molecules, including *fibroblast growth factors* (*FGFs*) and *vascular endothelial growth factor* (*VEGF*), which (among other effects) stimulate a variety of cell types to proliferate. By providing a dense array of growth factor binding sites, proteoglycans are thought to generate large local reservoirs of these factors, limiting their diffusion and focusing their actions on nearby cells. Similarly, proteoglycans might help generate steep growth factor gradients in an embryo, which can be important in the patterning of tissues during development. FGF activity can also be enhanced by proteoglycans, which oligomerize the FGF molecules, enabling them to cross-link and activate their cell-surface receptors more effectively.

The importance of proteoglycans 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 function of several signal proteins during development is governed by interactions with the membrane-associated proteoglycans *Dally* and *Dally-like*. These members of the *glypican* family are thought to concentrate signal proteins in specific locations and act as co-receptors that collaborate with the conventional cell-surface receptor proteins; as a result, they promote signaling in the correct location and prevent it in the wrong locations. In the *Drosophila* ovary, for example, Dally is partly responsible for the restricted localization and function of a signaling protein called Dpp, which blocks differentiation of the germ-line stem cells: when the gene encoding Dally is mutated, Dpp activity is greatly reduced and oocyte development is abnormal.

Several matrix proteins also interact with signal proteins. The type IV collagen of basal laminae interacts with Dpp in *Drosophila*, for example. Fibronectin contains a type III fibronectin repeat that interacts with VEGF, and another domain that interacts with another growth factor called hepatocyte growth factor (HGF), thereby promoting the activities of these factors. As discussed earlier, many matrix glycoproteins contain extensive arrays of binding domains, and the arrangement of these domains is likely to influence the presentation of signal proteins to their target cells (see Figure 19–46).

Finally, many matrix glycoproteins contain domains that bind directly to specific cell-surface receptors, thereby generating signals that influence the behavior of the cells, as we describe in the next section.

Summary

Cells are embedded in an intricate extracellular matrix, which 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 network of glycosaminoglycan (GAG) chains. GAGs are negatively charged polysaccharide chains that (except for hyaluronan) are covalently linked to protein to form proteoglycan molecules. GAGs attract water and 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. The fibrillar collagens (types I, II, III, V, and XI) are ropelike, triple-stranded helical molecules that aggregate into long fibrils in the extracellular space, thereby providing tensile strength. They also form structures to which cells can be anchored, often via large multidomain glycoproteins, such as laminin and fibronectin, that bind to 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 basal lamina is a specialized form of extracellular matrix that underlies epithelial cells or is wrapped around certain other cell types, such as muscle cells. Basal laminae are organized on a framework of laminin molecules, which are linked together by their side-arms and bind to integrins and other receptors in the basal plasma membrane of overlying epithelial cells. Type IV collagen molecules, together with the protein nidogen and the large heparan sulfate proteoglycan perlecan, assemble into a sheetlike mesh that is an essential component of all mature basal laminae. Basal laminae provide mechanical support for epithelia; they form the interface and 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; and they guide cell migration during development and tissue regeneration.

CELL–MATRIX JUNCTIONS

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 proteins 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. They are no less important in connective-tissue cells, mediating the cells' 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 plasma 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 the matrix. 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 serve not only to transmit mechanical and molecular signals, but also to convert one type of signal into the other.

Integrins Are Transmembrane Heterodimers That Link the Extracellular Matrix to the Cytoskeleton

There are many varieties of integrins, 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 (Figure 19–55). The extracellular domains bind to specific amino acid sequence motifs in extracellular matrix proteins or, in some cases, in proteins on the surfaces of other cells. The best-understood binding site for integrins is the RGD sequence mentioned earlier (see Figure 19–47), which is found in fibronectin and other extracellular matrix proteins. Some integrins bind a Leu-Asp-Val (LDV) sequence in fibronectin and other proteins. Additional integrin-binding sequences, as yet poorly defined, exist in laminins and collagens.

Humans contain 24 types of integrins, formed from the products of 8 different β -chain genes and 18 different α -chain genes, dimerized in different combinations. Each integrin dimer 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 matrix 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 extracellular ligands.

The intracellular portion of an integrin dimer binds to a complex of several different proteins, which together form a linkage to the cytoskeleton. For all but one of the 24 varieties of human integrins, this intracellular linkage is to actin filaments. These linkages depend on proteins that assemble at the short cytoplasmic tails of the integrin subunits (see Figure 19–55). A large adaptor protein called *talin* is a component of the linkage in many cases, but numerous additional proteins are also involved. 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 establish strong attachments to the rigid surface of a culture dish, and the *myotendinous junctions* that attach muscle cells to their tendons.

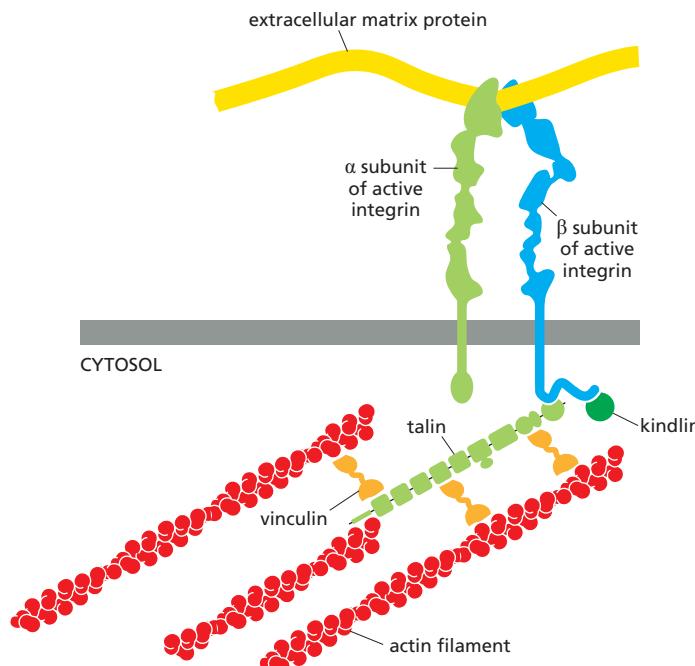
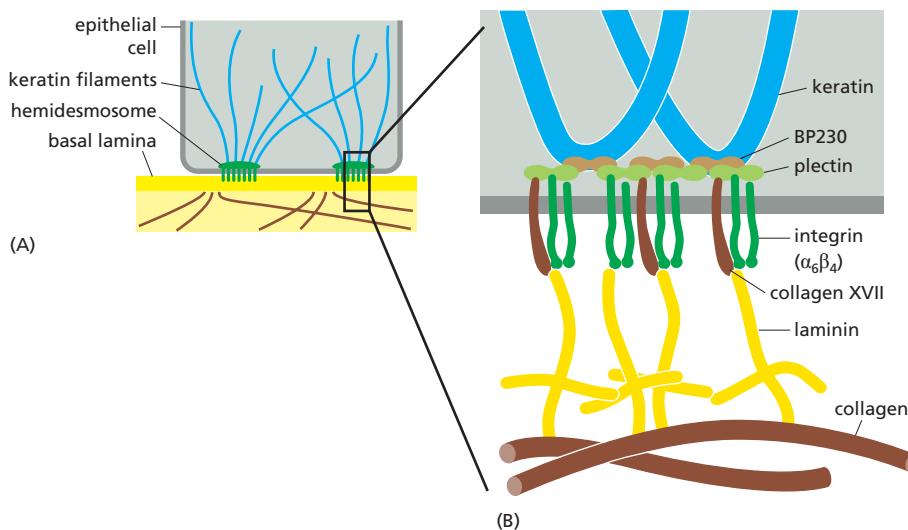


Figure 19–55 The subunit structure of an active integrin molecule, linking extracellular matrix to the actin cytoskeleton. The N-terminal heads of the integrin chains attach directly to an extracellular protein such as fibronectin; the C-terminal intracellular tail of the integrin β subunit binds to adaptor proteins that interact with filamentous actin. The best-understood adaptor is a giant protein called talin, which contains a string of multiple domains for binding actin and other proteins, such as vinculin, that help reinforce and regulate the linkage to actin filaments. One end of talin binds to a specific site on the integrin β subunit cytoplasmic tail; other regulatory proteins, such as kindlin, bind at another site on the tail.

**Figure 19-56 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 hemidesmosome. A specialized integrin ($\alpha_6\beta_4$ integrin) spans the membrane, attaching to keratin filaments intracellularly via adaptor proteins called plectin and BP230, and to laminin extracellularly. The adhesive complex also contains, in parallel with the integrin, an unusual collagen family member known as 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. One such disease, called *bullous pemphigoid*, is an autoimmune disease in which the immune system develops antibodies against collagen XVII or BP230.

In epithelia, the most prominent cell-matrix attachment sites are the hemidesmosomes, where a specific type of integrin anchors the cells to laminin in the basal lamina. Here, uniquely, the intracellular attachment is to keratin intermediate filaments, via the intracellular adaptor proteins plectin and BP230 (Figure 19-56).

Integrin Defects Are Responsible for Many Genetic Diseases

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 19-3 lists some varieties of integrins and the problems that result when individual integrin α or β chains are defective.

The β_1 subunit forms dimers with at least 12 distinct α subunits and is found on almost all vertebrate cells: $\alpha_5\beta_1$ is a fibronectin receptor and $\alpha_6\beta_1$ is a laminin

TABLE 19-3 Some Types of Integrins

Integrin	Ligand*	Distribution	Phenotype when α subunit is mutated	Phenotype when β subunit is mutated
$\alpha_5\beta_1$	Fibronectin	Ubiquitous	Death of embryo; defects in blood vessels, somites, neural crest	Early death of embryo (at implantation)
$\alpha_6\beta_1$	Laminin	Ubiquitous	Severe skin blistering; defects in other epithelia also	Early death of embryo (at implantation)
$\alpha_7\beta_1$	Laminin	Muscle	Muscular dystrophy; defective myotendinous junctions	Early death of embryo (at implantation)
$\alpha_L\beta_2$ (LFA1)	Ig superfamily counterreceptors (ICAM1)	White blood cells	Impaired recruitment of leucocytes	Leukocyte adhesion deficiency (LAD); impaired inflammatory responses; recurrent life-threatening infections
$\alpha_{IIb}\beta_3$	Fibrinogen	Platelets	Bleeding; no platelet aggregation (Glanzmann's disease)	Bleeding; no platelet aggregation (Glanzmann's disease); mild osteopetrosis
$\alpha_6\beta_4$	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.

receptor on many types of cells. Mutant mice that cannot make any β_1 integrins die early in embryonic development. Mice that are only unable to make the α_7 subunit (the partner for β_1 in muscle) survive but develop muscular dystrophy (as do mice that cannot make the laminin ligand for the $\alpha_7\beta_1$ integrin).

The β_2 subunit forms dimers with at least four types of α subunit and is expressed exclusively on the surface of white blood cells, where it has an essential role in enabling these cells to fight infection. The β_2 integrins mainly mediate cell-cell rather than cell-matrix interactions, binding to specific ligands on another cell, such as an endothelial cell. The ligands 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 ($\alpha_L\beta_2$, also known as LFA1) on white blood cells enables them to attach firmly to the Ig family protein ICAM1 on vascular endothelial cells at sites of infection (see Figure 19–28B). People with the genetic disease called *leukocyte adhesion deficiency* fail to synthesize functional β_2 subunits. As a consequence, their white blood cells lack the entire family of β_2 receptors, and they suffer from repeated bacterial infections.

The β_3 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 β_3 integrins, suffer from defective clotting and bleed excessively.

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. 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. Furthermore, if 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.

Structural studies, using a combination of electron microscopy and x-ray crystallography, suggest that integrins exist in multiple structural conformations that reflect different states of activity (Figure 19–57). In the inactive state, the external segments of the integrin dimer are folded together into a compact structure that cannot bind matrix proteins. In this state, the cytoplasmic tails of the dimer are

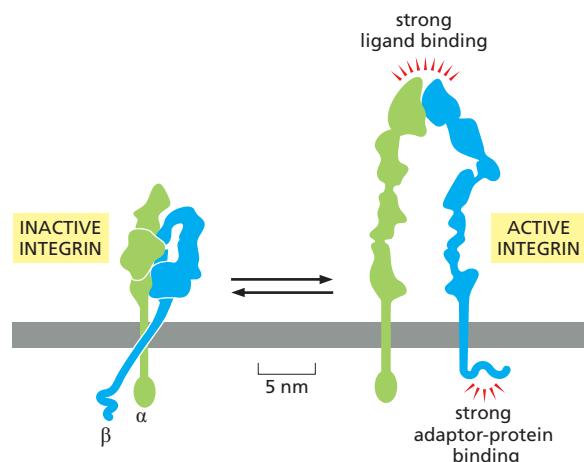


Figure 19–57 Integrins exist in two major activity states. Inactive (folded) and active (extended) structures of an integrin molecule, based on data from x-ray crystallography and other methods.

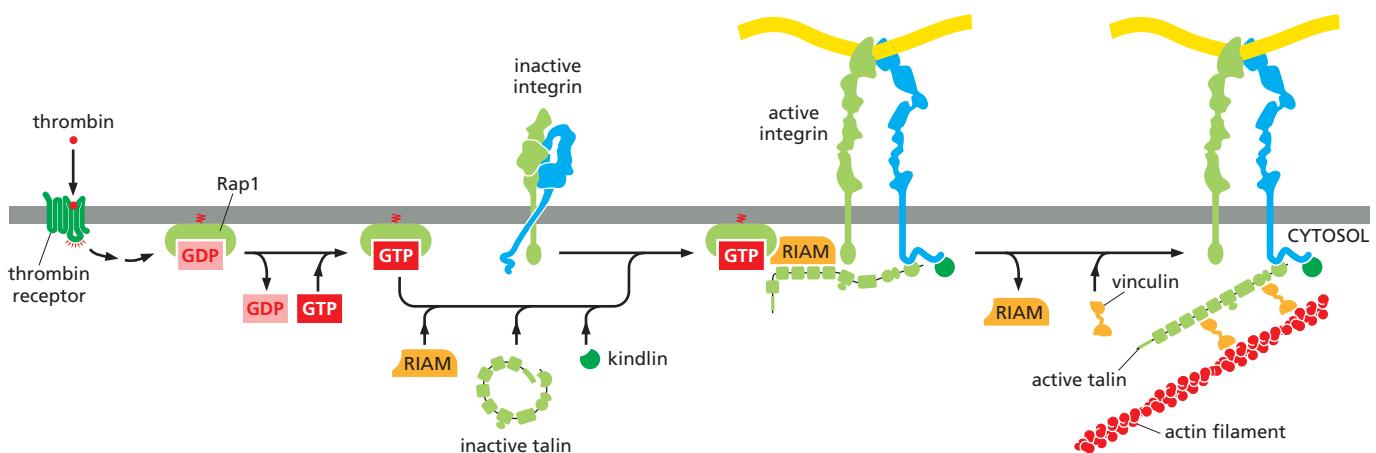


Figure 19–58 Activation of integrins by intracellular signaling. Signals received from outside the cell can act through various intracellular mechanisms to stimulate integrin activation. In platelets, as illustrated here, the extracellular signal protein thrombin activates a G-protein-coupled receptor on the cell surface, thereby initiating a signaling pathway that leads to activation of Rap1, a member of the monomeric GTPase family. Activated Rap1 interacts with the protein RIAM, which then recruits talin to the plasma membrane. Together with another protein called kindlin, talin interacts with the integrin β chain to trigger integrin activation. Talin then interacts with adaptor proteins such as vinculin, resulting in the formation of an actin linkage (see Figure 19–55).

Talin regulation depends in part on an interaction between its flexible C-terminal rod domain and the N-terminal head domain that contains the integrin-binding site. This interaction is thought to maintain talin in an inactive state when it is free in the cytoplasm. When talin is recruited by RIAM to the plasma membrane, the talin head domain interacts with a phosphoinositide called PI(4,5)P₂ (not shown here, but see Figure 15–28), resulting in dissociation of the rod domain. Talin unfolds to expose its binding sites for integrin and other proteins.

hooked together, preventing their interaction with cytoskeletal linker proteins.

In the active state, the two integrin subunits are unhooked at the membrane to expose the intracellular binding sites for cytoplasmic adaptor proteins, and the external domains unfold and extend, like a pair of legs, to expose a high-affinity matrix-binding site at the tips of the subunits. Thus, the switch from inactive to active states depends on a major conformational change that simultaneously exposes the external and internal ligand-binding sites at the ends of the integrin molecule. External matrix binding and internal cytoskeleton linkages are thereby coupled.

Switching between the inactive and active states is regulated by a variety of mechanisms that vary, depending on the needs of the cell. In some cases, activation occurs by an “outside-in” mechanism: the binding of an external matrix protein, such as the RGD sequence of fibronectin, can drive some integrins to switch from the low-affinity inactive state to the high-affinity active state. As a result, binding sites for talin and other cytoplasmic adaptor proteins are exposed on the tail of the β chain. The binding of these adaptor proteins then leads to attachment of actin filaments to the intracellular end of the integrin molecule (see Figure 19–55). In this way, when the integrin catches hold of its ligand outside the cell, the cell reacts by tying the integrin molecule to the cytoskeleton, so that force can be applied at the point of cell attachment.

The chain of cause and effect can also operate in reverse, from inside to outside. This “inside-out” integrin-activation process generally depends on intracellular regulatory signals that stimulate the ability of talin and other proteins to interact with the β chain of the integrin. 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 blocks the intracellular α - β linkage, allowing the two legs of the integrin molecule to spring apart.

The regulation of “inside-out” integrin activation is particularly well understood in platelets, where an extracellular signal protein called thrombin binds to a specific G-protein-coupled receptor (GPCR) on the cell surface and thereby activates an intracellular signaling pathway that leads to integrin activation (Figure 19–58). It is likely that similar signaling pathways govern integrin activation in numerous other cell types.

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 present at a 10–100-fold higher concentration on the cell surface. The Velcro principle, mentioned earlier in the context of cadherin adhesion (see Figure 19–6C), operates here too. Following their activation, integrins cluster together to create a dense plaque in which many integrin molecules are anchored to cytoskeletal filaments. The resulting protein structure can be remarkably large and complex, as seen in the focal adhesion made by a fibroblast on a fibronectin-coated surface culture dish.

The assembly of mature cell-matrix junctional complexes depends on the recruitment of dozens of different scaffolding and signaling proteins. Talin is a major component of many cell-matrix complexes, but numerous other proteins also make important contributions. These include the *integrin-linked kinase* (*ILK*) and its binding partners *pinch* and *parvin*, which together form a trimeric complex that serves as an organizing hub at many junctions. Cell-matrix junctions also employ several actin-binding proteins, such as *vinculin*, *zyxin*, *VASP*, and *α-actinin*, to promote the assembly and organization of actin filaments. Another critical component of many cell-matrix junctions is the *focal adhesion kinase* (*FAK*), which interacts with multiple components in the junction and serves an important function in signaling, as we describe next.

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.

Many cells will not grow or proliferate in culture 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 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. 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.

Our understanding of anchorage dependence has come mainly from studies of cells living on the surface of matrix-coated culture dishes. For connective-tissue cells that are normally surrounded by matrix on all sides, this is a far cry from the natural environment. Walking over a two-dimensional plain is very different from clambering through a three-dimensional 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 in cell behavior in 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 are crucial for cell proliferation and survival.

Integrins Recruit Intracellular Signaling Proteins at Sites of Cell-Matrix Adhesion

The mechanisms by which integrins signal into the cell interior are complex, involving several 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 15–49), for

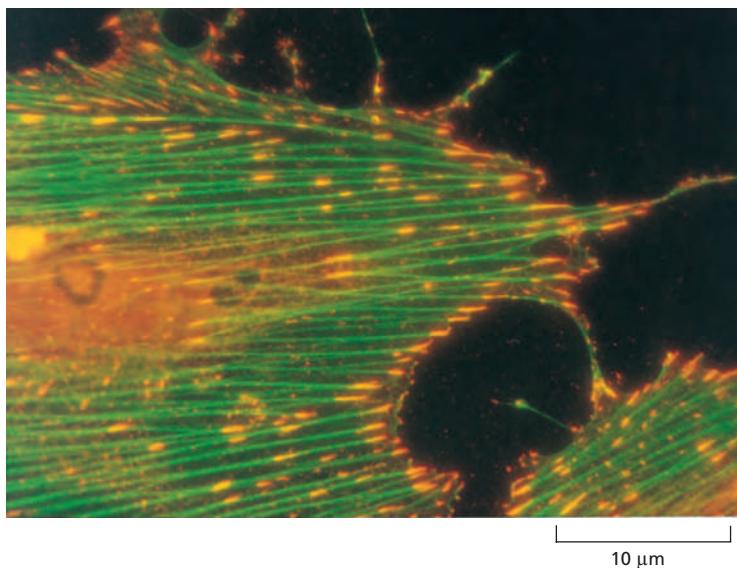


Figure 19–59 Tyrosine phosphorylation at focal adhesions. A fibroblast cultured on a fibronectin-coated substratum and stained with fluorescent antibodies: actin filaments are stained green and activated proteins that contain phosphotyrosine are red, giving orange where the two components overlap. The 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. (Courtesy of Keith Burridge.)

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 to promote cell survival (discussed in Chapters 15 and 18).

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 on plastic dishes, focal adhesions are often prominent sites of tyrosine phosphorylation (Figure 19–59), and FAK is one of the major tyrosine-phosphorylated proteins found at these sites. When integrins cluster at cell-matrix contacts, FAK is recruited to the integrin β subunit by intracellular adaptor proteins such as talin or *paxillin* (which binds to one type of integrin α subunit). The clustered FAK molecules 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 in much the same way as receptor tyrosine kinases generate signals (as discussed in Chapter 15).

Cell-Matrix Adhesions Respond to Mechanical Forces

Like the cell-cell junctions we described earlier, cell-matrix junctions can sense and respond to the mechanical forces that act on them. Most cell-matrix junctions, for example, are connected to a contractile actin network that tends to pull the junctions inward, away from the matrix. When cells are attached to a rigid matrix that strongly resists such pulling forces, the cell-matrix junction is able to sense the resulting high tension and trigger a response in which it recruits additional integrins and other proteins to increase the junction's ability to withstand that tension. Cell attachment to a relatively soft matrix generates less tension and therefore a less robust response. These mechanisms allow cells to sense and respond to differences in the rigidity of extracellular matrices in different tissues.

We saw earlier that mechanotransduction at cadherin-based cell-cell junctions likely depends on junctional proteins that change their structure when the junction is stretched by tension (see Figure 19–12). The same is true for cell-matrix junctions. The long C-terminal tail domain of talin, for example, includes a large number of binding sites for the actin-regulatory protein vinculin. Many of these sites are hidden inside folded protein domains but are exposed when those

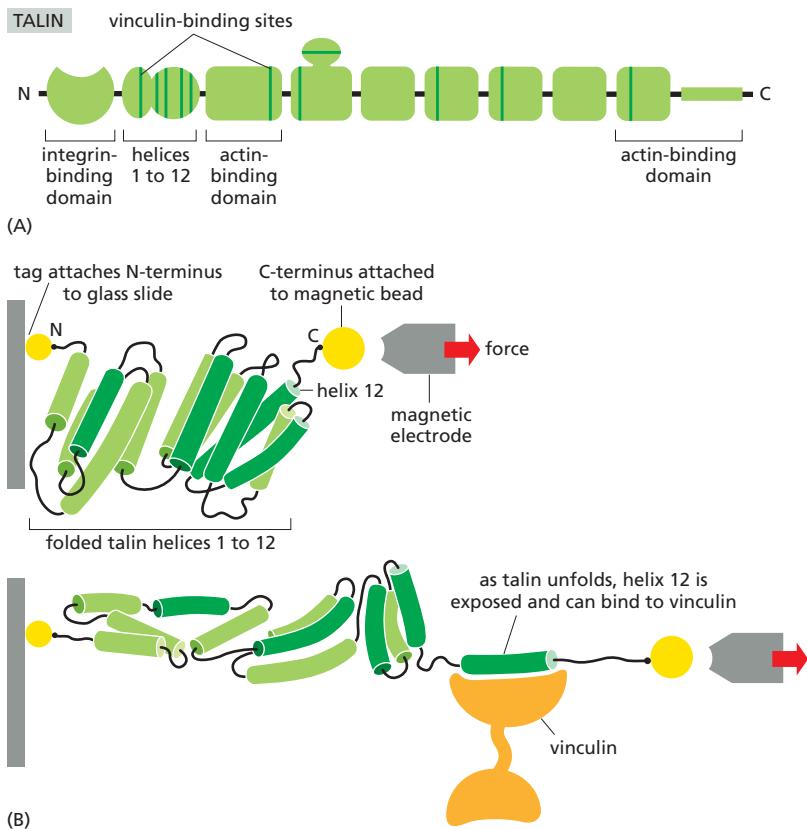


Figure 19–60 Talin is a tension sensor at cell–matrix junctions. Tension across cell–matrix junctions stimulates the local recruitment of vinculin and other actin-regulatory proteins, thereby strengthening the junction’s attachment to the cytoskeleton. The experiments presented here tested the hypothesis that tension is sensed by the talin adaptor protein that links integrins to actin filaments (see Figure 19–55). (A) The long, flexible, C-terminal region of talin is divided into a series of folded domains, some of which contain vinculin-binding sites (dark green lines) that are thought to be hidden and therefore inaccessible. One domain near the N-terminus, for example, comprises a folded bundle of 12 α helices containing five vinculin-binding sites. (B) This experiment tested the hypothesis that tension stretches the 12-helix domain, thereby exposing vinculin-binding sites. A fragment of talin containing this domain was attached to an apparatus in which the domain could be stretched, as shown here. The fragment was labeled at its N-terminus with a tag that sticks to the surface of a glass slide on a microscope stage. The C-terminal end of the fragment was bound to a tiny magnetic bead, so the talin fragment could be stretched using a small magnetic electrode. The solution around the protein contained fluorescently tagged vinculin proteins. After the talin protein was stretched, excess vinculin solution was washed away, and the microscope was used to determine if any fluorescent vinculin proteins were bound to the talin protein. In the absence of stretching (top), most talin molecules did not bind vinculin. When the protein was stretched (bottom), two or three vinculin molecules were bound (only one is shown here for clarity). (Adapted from A. del Rio et al., Science 323:638–641, 2009.)

domains are unfolded by stretching the protein (Figure 19–60). The N-terminal end of talin binds integrin and the C-terminal end binds actin (see Figure 19–55); thus, when actin filaments are pulled by myosin motors inside the cell, the resulting tension stretches the talin rod, thereby exposing vinculin-binding sites. The vinculin molecules then recruit and organize additional actin filaments. Tension thereby increases the strength of the junction.

Summary

Integrins are the principal cell-surface receptors used by animal cells to bind to the extracellular matrix: they function as transmembrane linkers between the extracellular matrix and the cytoskeleton. Most integrins connect to actin filaments, while those at hemidesmosomes bind to intermediate filaments. Integrin molecules are heterodimers, and the binding of extracellular matrix ligands or intracellular activator proteins such as talin results in a dramatic conformational switch from an inactive to an active state. 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. Complex assemblies of proteins become organized around the intracellular tails of activated 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. Integrin-based cell–matrix junctions are also capable of mechanotransduction: they can sense and respond to mechanical forces acting across the junction.

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

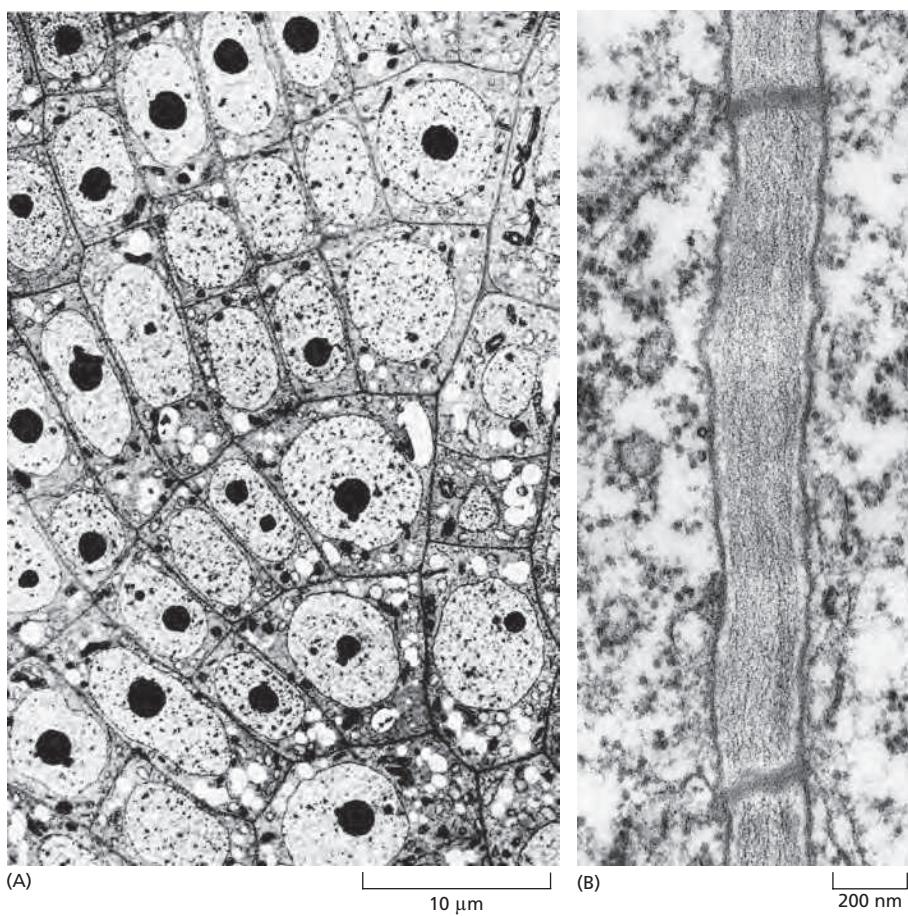


Figure 19–61 Plant cell walls. (A) Electron micrograph of the root tip of a rush, showing the organized pattern of cells that results from an ordered sequence of cell divisions in cells with relatively rigid cell walls. In this growing tissue, the cell walls are still relatively thin, appearing as fine 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 19–27). (A, courtesy of C. Busby and B. Gunning, *Eur. J. Cell Biol.* 21:214–223, 1980. With permission from Elsevier; B, courtesy of Jeremy Burgess.)

together to form the intact plant (Figure 19–61), 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*, 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 cell 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 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 CO₂, H₂O, 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. 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. 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 19–62). 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 that of steel. 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

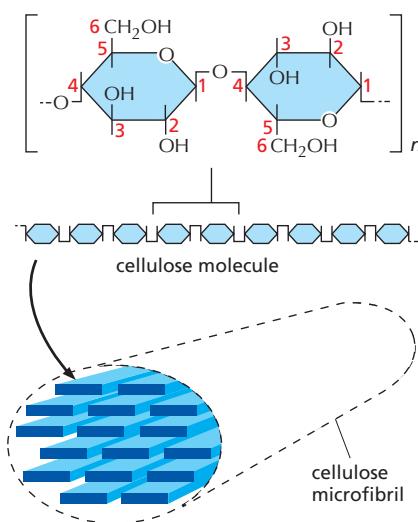


Figure 19–62 Cellulose. Cellulose molecules are long, unbranched chains of β 1,4-linked glucose units. Each glucose residue is inverted with respect to its neighbors, and the resulting disaccharide repeat occurs hundreds of times in a single cellulose molecule. About 16 individual cellulose molecules assemble to form a strong, hydrogen-bonded cellulose microfibril.

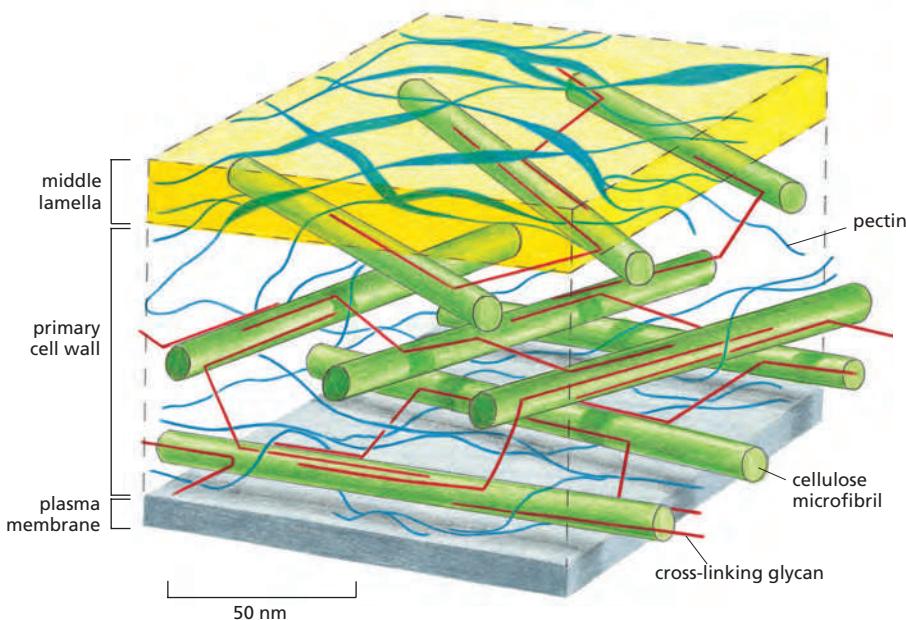


Figure 19–63 Scale model of a portion of a primary plant cell wall showing the two major polysaccharide networks. The orthogonally arranged layers of cellulose microfibrils (green) are tied into a network by the cross-linking glycans (red) that form hydrogen bonds with the microfibrils. This network is coextensive with a network of pectin 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.

glycan molecules, which 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 19–63).

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. 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 Figure 19–63). 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 19–33). 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 19–63); 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, contains 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.

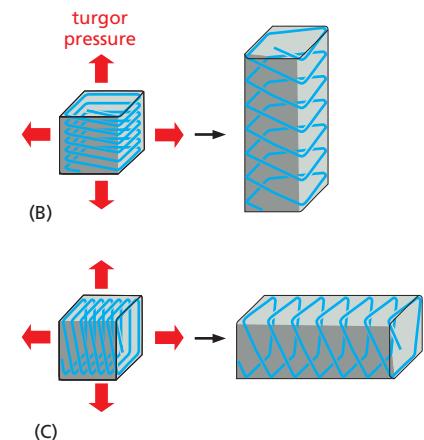
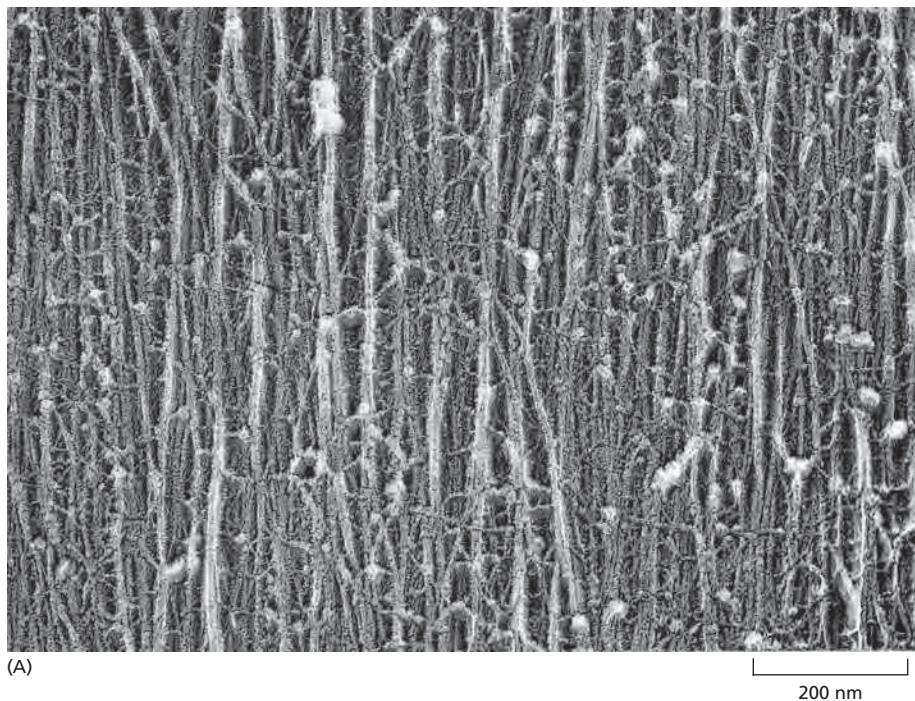


Figure 19–64 Cellulose microfibrils influence the direction of cell elongation.

(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 19–63). (B, C) The cells in (B) and (C) 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.)

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 19–64**).

Unlike most other matrix macromolecules, which are made in the endoplasmic reticulum and Golgi apparatus and are secreted, cellulose 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 sixfold symmetry (see Figure 19–65) and contains the protein products of three separate cellulose synthase (*CESA*) genes. Each *CESA* protein 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 19–63). 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 19–64B and C).

Microtubules Orient Cell Wall Deposition

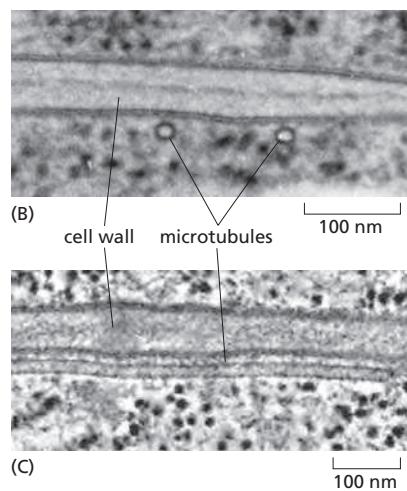
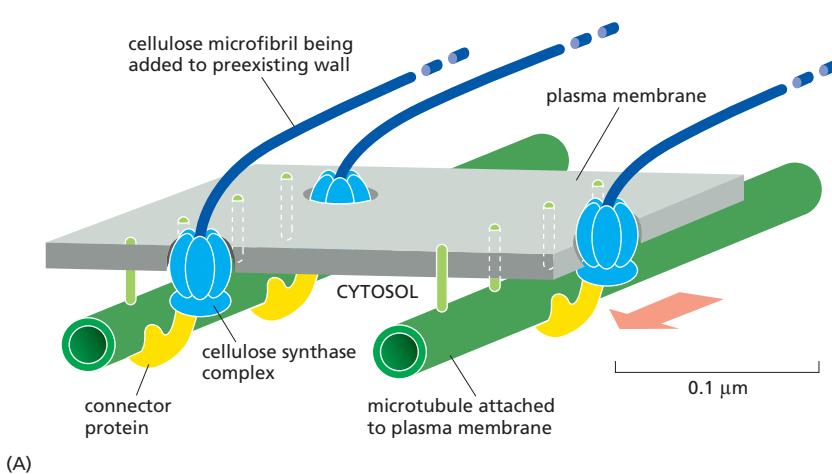
An important clue to the mechanism that dictates microfibril orientation came from observations of the microtubules in plant cells. These are frequently arranged in the cortical cytoplasm with the same orientation as the cellulose microfibrils that are currently being deposited in the cell wall in that region. These cortical microtubules form a *cortical array* close to the cytosolic face of the plasma membrane, held there by poorly characterized proteins. The 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 suggestion 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, the rosette moves 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 19–65). 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.

Figure 19–65 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, or rosettes, 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 microtubule in 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.)



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 plant growth regulators such as ethylene and auxins (discussed in Chapter 15).

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.

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 lifestyle. The wall is composed of a network of cellulose microfibrils and cross-linking glycans, embedded in a highly cross-linked matrix of pectin polysaccharides. In secondary cell walls, lignin may be deposited to make 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 plasma membrane domains to prevent the mixing of apical and basolateral membrane proteins.

19–3 The elasticity of elastin derives from its high content of α helices, which act as molecular springs.

19–4 Integrins can convert mechanical signals into intracellular molecular signals.

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

WHAT WE DON'T KNOW

- What are the regulatory mechanisms that control the rearrangement of cell–cell junctions in epithelia during early development? What roles do mechanical force and tension play in these rearrangements?
- How do extracellular matrix proteins and carbohydrates influence the localization and actions of extracellular signal molecules or their cell-surface receptors?
- How do intracellular adaptor proteins coordinate the activation of integrin proteins and their interactions with cytoskeletal components and their response to changes in mechanical force acting on cell–matrix junctions?
- Given that extracellular matrix molecules have the ability to present ordered arrays of signals to cells, might the exact spatial relationships between such signals carry a message beyond that of the individual signals themselves?

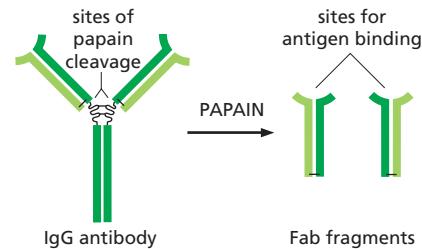


Figure Q19–1 Production of Fab fragments from IgG antibodies by digestion with papain (Problem 19–6).

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 the C-terminus of the toxin is bound to a claudin, the N-terminus can 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

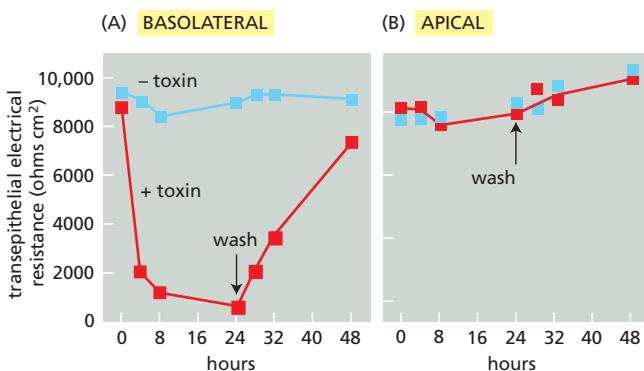


Figure Q19-2 Effects of *Clostridium* toxin on the barrier function of MDCK cells (Problem 19-7). (A) Addition of toxin from the basolateral side of the epithelial sheet. (B) Addition of toxin from the apical side of the epithelial sheet. For a given voltage, a higher resistance (ohms cm^2) gives less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

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 out the toxin (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).

- A.** How can it be that two tight-junction strands remain, even though all of the claudin-4 has disappeared?
B. 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 γ -1 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 γ -1, 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 γ -1 chain is thought to have no other function than to bind nidogen, and does not affect laminin structure or its ability to assemble into the 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?

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 γ -1	Nidogen binding-site deletion ($+/-$)	None
Laminin γ -1	Nidogen binding-site deletion ($-/-$)	Dead at birth

$+/-$ stands for heterozygous, $-/-$ stands for homozygous.

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.”

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 $\alpha_{IIb}\beta_3$ 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 $\alpha_{IIb}\beta_3$ integrin is held in its inactive state?

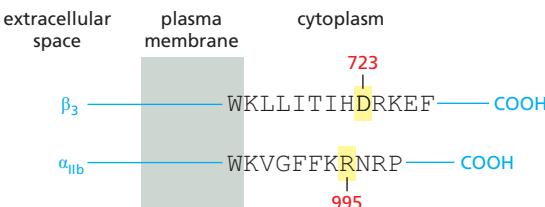


Figure Q19-3 Schematic representation of $\alpha_{IIb}\beta_3$ integrin (Problem 19-10). The D723 and R995 residues are indicated. (From P.E. Hughes et al., *J. Biol. Chem.* 271:6571–6574, 1996. With permission from American Society for Biochemistry and Molecular Biology.)

19-11 The glycosaminoglycan polysaccharide chains that are linked to specific core proteins to form the proteoglycan components of the extracellular space are highly negatively charged. How do you suppose these negatively charged polysaccharide chains help to establish a hydrated gel-like environment around the cell? How would the properties of these molecules differ if the polysaccharide chains were uncharged?

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 levels of D-aspartate that increase steadily with age?

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 cannot 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 A plant must be able to respond to changes in the water status of its surroundings. It does so by the flow of water molecules through water channels called aquaporins. The hydraulic conductivity of a single aquaporin is $4.4 \times 10^{-22} \text{ m}^3$ 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.
Hynes RO & Yamada KM (eds) (2011) Extracellular Matrix Biology (Cold Spring Harbor Perspectives in Biology). Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Cell–Cell Junctions

- Brasch J, Harrison OJ, Honig B & Shapiro L (2012) Thinking outside the cell: how cadherins drive adhesion. *Trends Cell Biol.* 22, 299–310.
Gomez GA, McLachlan RW & Yap AS (2011) Productive tension: force-sensing and homeostasis of cell-cell junctions. *Trends Cell Biol.* 21, 499–505.
Goodenough DA & Paul DL (2003) Beyond the gap: functions of unpaired connexon channels. *Nat. Rev. Mol. Cell Biol.* 4, 285–294.
Gumbiner BM (2005) Regulation of cadherin-mediated adhesion in morphogenesis. *Nat. Rev. Mol. Cell Biol.* 6, 622–634.
Harris TJ & Tepass U (2010) Adherens junctions: from molecules to morphogenesis. *Nat. Rev. Mol. Cell Biol.* 11, 502–514.
King N, Hittinger CT & Carroll SB (2003) Evolution of key cell signaling and adhesion protein families predates animal origins. *Science* 301, 361–363.
Leckband DE, le Duc Q, Wang N & de Rooij J (2011) Mechanotransduction at cadherin-mediated adhesions. *Curr. Opin. Cell Biol.* 23, 523–530.
Lecuit T, Lenne PF & Munro E (2011) Force generation, transmission, and integration during cell and tissue morphogenesis. *Annu. Rev. Cell Dev. Biol.* 27, 157–184.
Litjens SH, de Pereda JM & Sonnenberg A (2006) Current insights into the formation and breakdown of hemidesmosomes. *Trends Cell Biol.* 16, 376–383.
Maule AJ, Benitez-Alfonso Y & Faulkner C (2011) Plasmodesmata—membrane tunnels with attitude. *Curr. Opin. Plant Biol.* 14, 683–690.
McEver RP & Zhu C (2010) Rolling cell adhesion. *Annu. Rev. Cell Dev. Biol.* 26, 363–396.
Nakagawa S, Maeda S & Tsukihara T (2010) Structural and functional studies of gap junction channels. *Curr. Opin. Struct. Biol.* 20, 423–430.
Shin K, Fogg VC & Margolis B (2006) Tight junctions and cell polarity. *Annu. Rev. Cell Dev. Biol.* 22, 207–236.
Takeichi M (2007) The cadherin superfamily in neuronal connections and interactions. *Nat. Rev. Neurosci.* 8, 11–20.
Thomason HA, Scothern A, McHarg S & Garrod DR (2010) Desmosomes: adhesive strength and signalling in health and disease. *Biochem. J.* 429, 419–433.

The Extracellular Matrix of Animals

- 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 (2010) Transmembrane signaling proteoglycans. *Annu. Rev. Cell Dev. Biol.* 26, 89–114.
Domogatskaya A, Rodin S & Tryggvason K (2012) Functional diversity of laminins. *Annu. Rev. Cell Dev. Biol.* 28, 523–553.
Hynes RO (2009) The extracellular matrix: not just pretty fibrils. *Science* 326, 1216–1219.
Hynes RO & Naba A (2012) Overview of the matrisome—an inventory of extracellular matrix constituents and functions. *Cold Spring Harb. Perspect. Biol.* 4, a004903.
Kiely CM, Sherratt MJ & Shuttleworth CA (2002) Elastic fibres. *J. Cell Sci.* 115, 2817–2828.
Larsen M, Artym WV, Green JA & Yamada KM (2006) The matrix reorganized: extracellular matrix remodeling and integrin signaling. *Curr. Opin. Cell Biol.* 18, 463–471.
Lu P, Takai K, Weaver VM & Werb Z (2011) Extracellular matrix degradation and remodeling in development and disease. *Cold Spring Harb. Perspect. Biol.* 3, a005058.
Ricard-Blum S (2011) The collagen family. *Cold Spring Harb. Perspect. Biol.* 3, a004978.
Sasaki T, Fässler R & Hohenester E (2004) Laminin: the crux of basement membrane assembly. *J. Cell Biol.* 164, 959–963.
Toole BP (2001) Hyaluronan in morphogenesis. *Semin. Cell Dev. Biol.* 12, 79–87.
Yurchenco PD (2011) Basement membranes: cell scaffoldings and signaling platforms. *Cold Spring Harb. Perspect. Biol.* 3, a004911.

Cell–Matrix Junctions

- Calderwood DA, Campbell ID & Critchley DR (2013) Talins and kindlins: partners in integrin-mediated adhesion. *Nat. Rev. Mol. Cell Biol.* 14, 503–517.
Campbell ID & Humphries MJ (2011) Integrin structure, activation, and interactions. *Cold Spring Harb. Perspect. Biol.* 3, a004994.
Hoffman BD, Grashoff C & Schwartz MA (2011) Dynamic molecular processes mediate cellular mechanotransduction. *Nature* 475, 316–323.

- Hogg N, Patzak I & Willenbrock F (2011) The insider's guide to leukocyte integrin signalling and function. *Nat. Rev. Immunol.* 11, 416–426.
- Kanchanawong P, Shtengel G, Pasapera AM et al. (2010) Nanoscale architecture of integrin-based cell adhesions. *Nature* 468, 580–584.
- Luo BH & Springer TA (2006) Integrin structures and conformational signaling. *Curr. Opin. Cell Biol.* 18, 579–586.
- Moser M, Legate KR, Zent R & Fässler R (2009) The tail of integrins, talin, and kindlins. *Science* 324, 895–899.
- Ross TD, Coon BG, Yun S et al. (2013) Integrins in mechanotransduction. *Curr. Opin. Cell Biol.* 25, 613–618.
- Shattil SJ, Kim C & Ginsberg MH (2010) The final steps of integrin activation: the end game. *Nat. Rev. Mol. Cell Biol.* 11, 288–300.
- The Plant Cell Wall**
- Albersheim P, Darvill A, Roberts K et al. (2011) Plant Cell Walls: From Chemistry to Biology. New York: Garland Science.
- Braidwood L, Breuer C & Sugimoto K (2013) My body is a cage: mechanisms and modulation of plant cell growth. *New Phytol.* 210, 388–402.
- Keegstra K (2010) Plant cell walls. *Plant Physiol.* 154, 483–486.
- Li S, Lei L, Somerville C et al. (2011) Cellulose synthase interactive protein 1 (CSI1) links microtubules and cellulose synthase complexes. *Proc. Natl. Acad. Sci. USA* 109, 189–190.
- Lloyd C (2011) Dynamic microtubules and the texture of plant cell walls. *Int. Rev. Cell Mol. Biol.* 287, 287–329.
- McFarlane HE, Döring A & Perrson S (2014) The cell biology of cellulose synthesis. *Annu. Rev. Plant Biol.* 65, 69–94.
- Somerville C (2006) Cellulose synthesis in higher plants. *Annu. Rev. Cell Dev. Biol.* 22, 53–78.
- Szymanski DB & Cosgrove DJ (2009) Dynamic Coordination of cytoskeletal and cell wall systems during cell wall biogenesis. *Curr. Biol.* 19, R800–R811.
- Wightman R & Turner SR (2008) The roles of the cytoskeleton during cellulose deposition at the secondary cell wall. *Plant J.* 54, 794–805.
- Wolf S, Hématy K & Höfte H (2012) Growth control and cell wall signaling in plants. *Annu. Rev. Plant Biol.* 63, 381–407.

Cancer

CHAPTER 20

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. These transgressions help to reveal what the normal rules are and how they are enforced. As a result, cancer research helps to illuminate the fundamentals of cell biology—especially cell signaling (Chapter 15), the cell cycle and cell growth (Chapter 17), programmed cell death (apoptosis, Chapter 18), and the control of tissue architecture (Chapters 19 and 22). Of course, with a deeper understanding of these normal processes, we also gain a deeper understanding of the disease and better tools to treat it.

In this chapter, we first consider what cancer is and describe the natural history of the disease from a cellular standpoint. We then discuss the molecular changes that make a cell cancerous. And we end the chapter by considering how our enhanced understanding of the molecular basis of cancer is leading to improved methods for its prevention and treatment.

CANCER AS A MICROEVOLUTIONARY PROCESS

The body of an animal operates as a society or ecosystem, whose individual members are cells that reproduce by cell division and organize themselves into collaborative assemblies called *tissues*. This ecosystem 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 the support of the germ cells, which alone have a chance of continued survival (discussed in Chapter 21). There is no 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 cells 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 slightly more vigorously and survive more readily than its neighbors and in this way 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

IN THIS CHAPTER

CANCER AS A MICROEVOLUTIONARY PROCESS

CANCER-CRITICAL GENES: HOW THEY ARE FOUND AND WHAT THEY DO

CANCER PREVENTION AND TREATMENT: PRESENT AND FUTURE

prospering at the expense of its neighbors. In the end—as the clone grows, evolves, and spreads—it can destroy the entire cellular society ([Movie 20.1](#)).

In this section, we discuss the development of cancer as a microevolutionary process that takes place within the course of a human life-span in a subpopulation of cells in the body. But the process depends on the same principles of mutation and natural selection that have driven the evolution of living organisms on Earth for billions of years.

Cancer Cells Bypass Normal Proliferation Controls 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 have not yet become invasive, however, the tumor is said to be **benign**. For most types of such neoplasms, removing or destroying the mass locally usually achieves a complete cure. A tumor is considered a true cancer if it is **malignant**; that is, when 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](#)). In general, the more widely a cancer spreads, the harder it becomes to eradicate. It is generally metastases that kill the cancer patient.

Cancers are traditionally 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. They account for about 80% of cases, perhaps because most of the cell proliferation in adults occurs in epithelia. In addition, epithelial tissues are the most likely to be exposed to the various forms of physical and chemical damage that favor the development of cancer. **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 rates. Each broad category has many subdivisions according to the specific cell type, the 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 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. Basal-cell carcinomas are readily cured by surgery or local irradiation, whereas malignant melanomas, once they have metastasized widely, are usually fatal.

Later, we shall see that there is also a different way to classify cancers, one that cuts across the traditional classification by site of origin: we can classify them in terms of the mutations that make the tumor cells cancerous. The final section of the chapter will show how this information can be crucial to the design and choice of treatments.

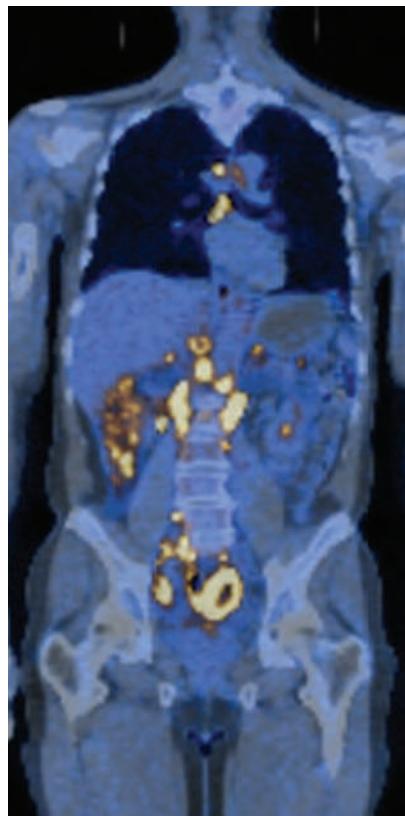


Figure 20–1 Metastasis. Malignant tumors typically give rise to metastases, making the cancer hard to eradicate. Shown in this fusion image is a whole-body scan of a patient with metastatic non-Hodgkin's lymphoma (NHL). The background image of the body's tissues was obtained by CT (computed x-ray tomography) scanning. Overlaid on this image, a PET (positron emission tomography) scan reveals the tumor tissue (yellow), detected by its unusually high uptake of radioactively labeled fluorodeoxyglucose (FDG). High FDG uptake occurs in cells with unusually active glucose uptake and metabolism, which is a characteristic of cancer cells (see [Figure 20–12](#)). The yellow spots in the abdominal region reveal multiple metastases. (Courtesy of S. Gambhir.)

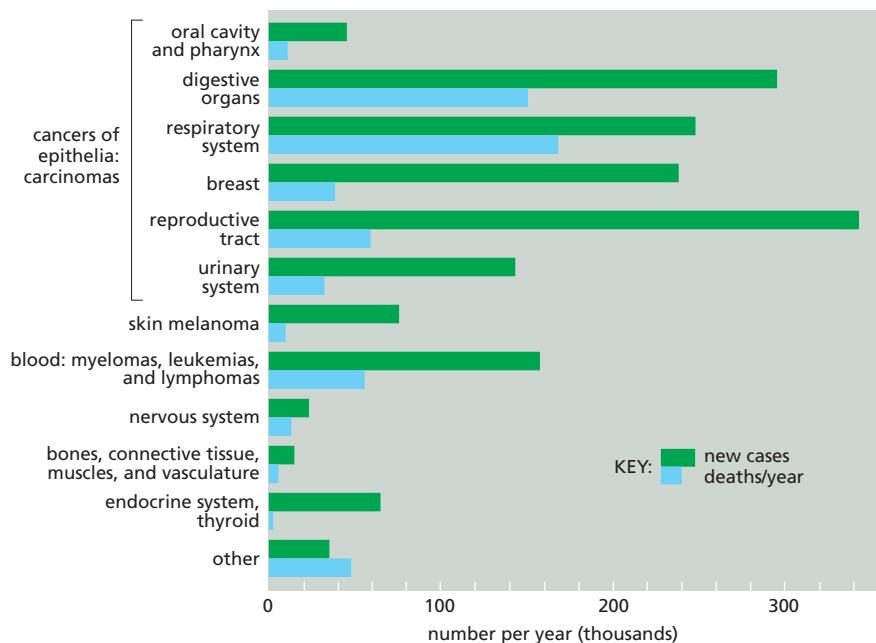


Figure 20–2 Cancer incidence and mortality in the United States. The total number of new cases diagnosed in 2012 in the United States was 1,665,540, and total cancer deaths were 585,720. Note that deaths reflect cases diagnosed at many different times and that somewhat less than half of the people who 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.

The data for the United Kingdom are similar. However, incidences are different in some other parts of the world, reflecting widespread exposures to different infectious agents and environmental toxins. (Data from American Cancer Society, Cancer Facts and Figures, 2014.)

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 a billion cancer cells or more (Figure 20–4). Tumors will usually also contain a variety of other cell types; for example, fibroblasts will be present in the supporting connective tissue associated with a carcinoma, in addition to inflammatory and vascular endothelial cells. How can we be sure that the cancer cells are the clonal descendants of a single abnormal cell?

One way of proving clonal origin is through molecular analysis of the chromosomes in tumor cells. In almost all patients with *chronic myelogenous leukemia* (*CML*), for example, we can distinguish the leukemic white blood cells from the patient's 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

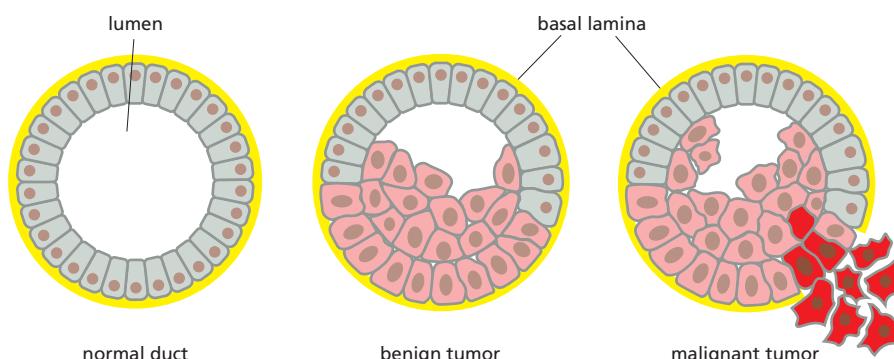


Figure 20–3 Benign versus malignant tumors. A benign glandular tumor (pink cells; an adenoma) remains inside the basal lamina (yellow) that marks the boundary of the normal structure (a duct, in this example). In contrast, a malignant glandular tumor (red cells; an adenocarcinoma) can develop from a benign tumor cell, and it destroys the integrity of the tissue, as shown. There are many different forms that such tumors may take.

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. However, particularly virulent tumors may grow much more rapidly.

how this particular 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.

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. Thus, the development of a clone of cancer cells depends on genetic changes. The tumor cells contain **somatic mutations**: they 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. (The mutations are called *somatic* because they occur in the soma, or body cells, not in the germ line). Cancers are also driven by *epigenetic changes*—persistent, heritable changes in gene expression that result from modifications of chromatin structure without alteration of the cell's DNA sequence. But somatic mutations that alter DNA sequence appear to be a fundamental and universal feature, and cancer is in this sense a genetic disease.

Factors that cause genetic changes tend to provoke the development of cancer. Thus, **carcinogenesis** (the generation of cancer) can be linked to **mutagenesis** (the production of a change in the DNA sequence). This correlation is particularly clear for two classes of external 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 ultraviolet (UV) light (which causes specific DNA base alterations).

As would be expected, people who have inherited a genetic defect in one of several DNA repair mechanisms, causing their cells to accumulate mutations at an elevated rate, run a heightened risk of cancer. Those 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.

A Single Mutation Is Not Enough to Change a Normal Cell into a Cancer Cell

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 lifespan, 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 pp. 237–238). 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 mutation in a single gene were enough to convert a typical healthy cell into a cancer cell, 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 and epigenetic accidents occur in the lineage that emanates from a single cell. One such indication comes from epidemiological studies of the incidence of cancer as a function of age (Figure 20–6). If a

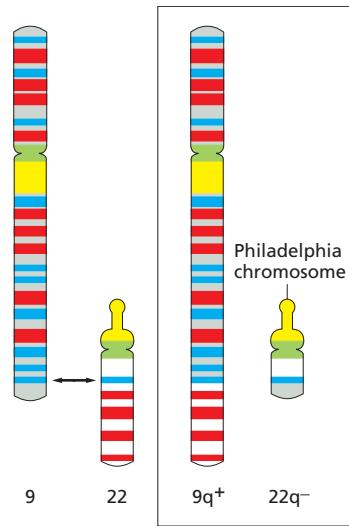
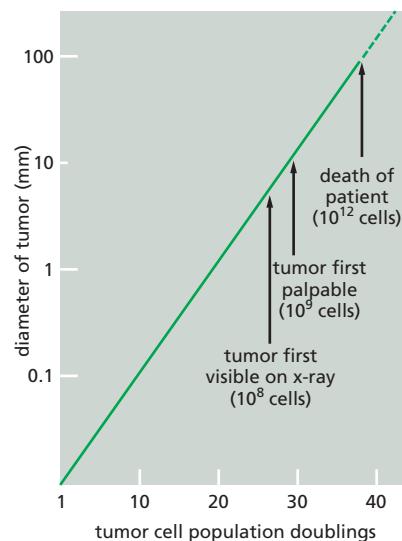
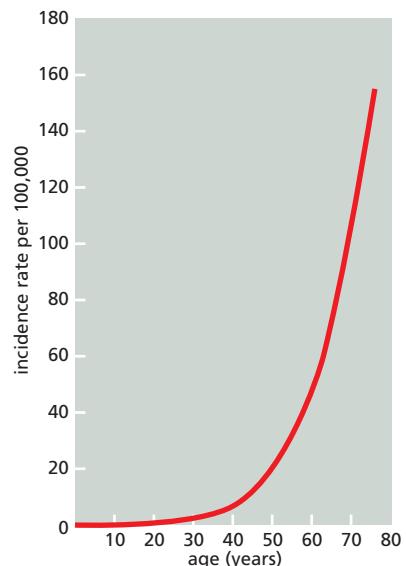


Figure 20–5 The translocation between chromosomes 9 and 22 responsible for chronic myelogenous leukemia. The normal structures of chromosomes 9 and 22 are shown at the left. When a translocation occurs between them at the indicated site, the result is the abnormal pair at the right. The smaller of the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

Figure 20–6 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, 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 of this cancer would be the same at all ages. Analyses of this type suggest that the development of a solid tumor instead requires five to eight independent accidents ("hits") that occur randomly over time. This calculation assumes that the mutation rate remains constant as a cancer evolves, where in fact it often increases (see p. 1097). (Data from C. Muir et al., *Cancer Incidence in Five Continents*, Vol. V. Lyon: International Agency for Research on Cancer, 1987.)

single mutation were responsible for cancer, 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, random accumulation of a set of mutations in a single lineage of cells.

As discussed later, these indirect arguments have now been confirmed by systematically sequencing the genomes of the tumor cells from individual cancer patients and cataloging the mutations that they contain.



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 decades of heavy smoking (Figure 20–7). 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. During this long incubation period, the prospective cancer cells undergo a succession of changes, and the same presumably applies 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 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 again 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

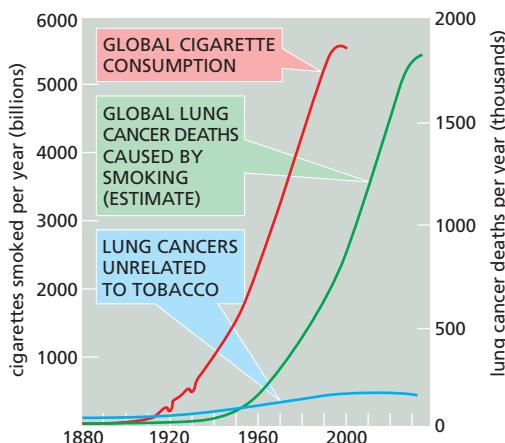


Figure 20–7 Smoking and the onset of lung cancer. A major increase in cigarette smoking (red line) has caused a dramatic rise in lung cancer deaths (green line), with a lag time of about 35 years. Because global cigarette smoking peaked in 1990, global lung cancer deaths are expected to decline after a similar lag. (Data from R.N. Proctor, *Nat. Rev. Cancer* 1:82–86, 2001).

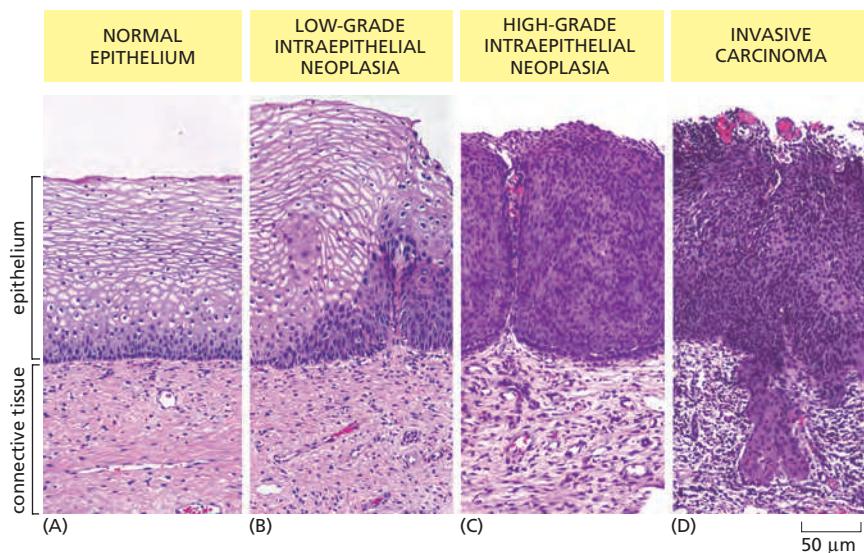


Figure 20–8 Stages of progression in the development of cancer of the epithelium of the uterine cervix. Pathologists use standardized terminology to classify the types of disorders they see, so as to guide the choice of treatment. (A) In a stratified squamous epithelium, dividing cells are confined to the basal layer. (B) In this low-grade intraepithelial neoplasia (right half of image), 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.)

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 that cells from the initial mutant clone have undergone further mutations that make them proliferate even more vigorously, so that they come to outnumber both the normal blood cells and their ancestors with the primary chromosomal translocation.

Carcinomas and other solid tumors evolve in a similar way (Figure 20–8). 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 and, as we shall see later, to relate them to specific genetic changes.

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

From all the evidence, therefore, 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. Correspondingly, 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 may be why the majority of us will die of causes other than cancer.

At each stage of progression, some individual 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 larger the number of tumor cells, the higher the chance that at least one of them will undergo a change that favors it over its neighbors. Thus, as the tumor grows, progression accelerates. The offspring of the best-adapted cells continue to divide, eventually producing the dominant clones in the developing lesion (Figure 20–9).

Just as in the evolution of plants and animals, a kind of speciation often occurs: the original cancer cell lineage can diversify to give many genetically different vigorous subclones of cells. These may coexist in the same mass of tumor tissue; or they may migrate and colonize separate environments suited to their individual quirks, where they settle, thrive, and progress as independently evolving metastases. As new mutations arise within each tumor mass, different subclones may gain an advantage and come to predominate, only to be overtaken by others or outgrown by their own sub-subclones. The increasing genetic diversity as a cancer progresses is one of the chief factors that make cures difficult.

Figure 20–9 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 that is 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. Not shown here is the fact that, over time, a variety of competing subclones will often arise in a tumor. As we will discuss later, this heterogeneity complicates cancer therapies (see Figure 20–30).

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**. The extent of this instability and its molecular origins differ from cancer to cancer and from patient to patient, as we shall discuss in a later section. The basic phenomenon was evident even before modern molecular analyses. For example, the cells of many cancers show grossly abnormal sets of chromosomes, with duplications, deletions, and translocations that are visible at mitosis (Figure 20–10). When the cells are maintained in culture, these patterns of chromosomal disruption can often be seen to evolve rapidly and in a seemingly haphazard way. And for many years, pathologists have used an abnormal appearance of the cell nucleus to identify and classify cancer cells in tumor biopsies; in particular, cancer cells can contain an unusually large amount of heterochromatin—a condensed form of interphase chromatin that silences genes (see pp. 194–195). This suggested that epigenetic changes of chromatin structure can also contribute to the cancer cell phenotype, as recently confirmed by molecular analysis.

The genetic instability observed in cancer cells can arise from defects in the ability to repair DNA damage or to correct replication errors of various kinds. These alterations lead to changes in DNA sequence and produce rearrangements such as DNA translocations and duplications. Also common are defects in chromosome segregation during mitosis, which provide another possible source of chromosome instability and changes in karyotype.

From an evolutionary perspective, none of this should be a surprise: anything that increases the probability of random changes in gene function heritable from one cell generation to the next—and that is not too deleterious—is likely to speed the evolution of a clone of cells toward malignancy, thereby causing this property to be selected for during tumor progression.

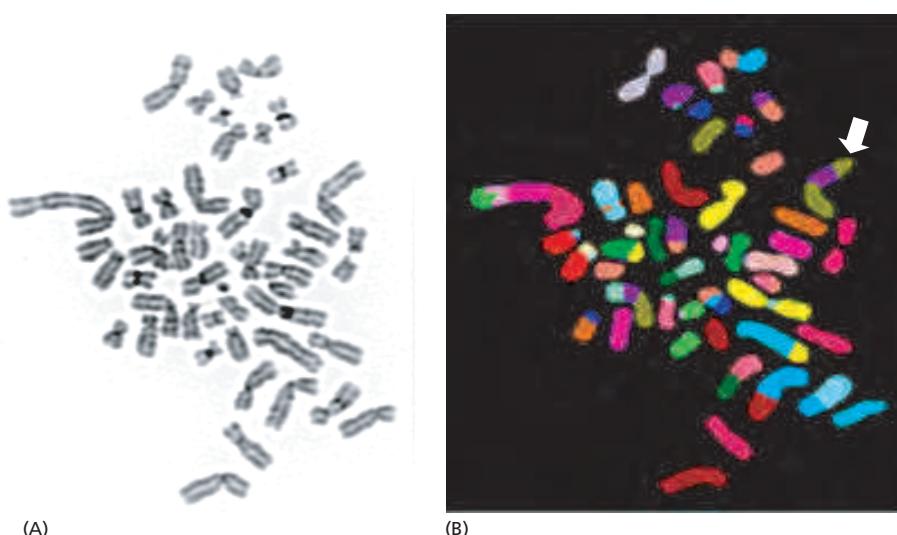
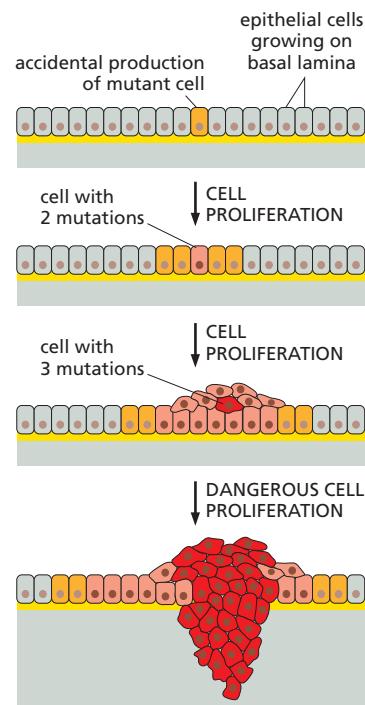


Figure 20–10 Chromosomes from a breast tumor displaying abnormalities in structure and number. Chromosomes were prepared from a breast tumor cell in metaphase, spread on a glass slide, and stained with (A) a general DNA stain or (B) a combination of fluorescently labeled DNA molecules that color each normal human chromosome differently (see Figure 4–10). The staining (displayed in false color) shows multiple translocations, including a doubly translocated chromosome (white arrow) that is made up of two pieces of chromosome 8 (green-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.)

Cancer Cells Display an Altered Control of Growth

Mutability and large cell population numbers create the opportunities for mutations to occur, but the driving force for development of a cancer has to come from some sort of selective advantage possessed by the mutant cells. Most obviously, a mutation or epigenetic change can confer such an advantage by increasing the rate at which a clone of cells proliferates or by enabling it to continue proliferating when normal cells would stop. Cancer cells that can be grown in culture, or cultured cells artificially engineered to contain the types of mutations encountered in cancers, typically show a **transformed** phenotype. They are abnormal in their shape, their motility, their responses to growth factors in the culture medium, and, most characteristically, in the way they react to contact with the substratum and with one another. Normal cells will not divide unless they are attached to the substratum; transformed cells will often divide even if held in suspension. Normal cells become inhibited from moving and dividing when the culture reaches confluence (where the cells are touching one another); transformed cells continue moving and dividing even after confluence, and so pile up in layer upon layer in the culture dish (**Figure 20–11**). In addition, transformed cells no longer require all of the positive signals from their surroundings that normal cells require.

Their behavior in culture gives a hint of the ways in which cancer cells may misbehave in their natural environment, embedded in a tissue. But cancer cells in the body show other peculiarities that mark them out from normal cells, beyond those just described.

Cancer Cells Have an Altered Sugar Metabolism

Given sufficient oxygen, normal adult tissue cells will generally fully oxidize almost all the carbon in the glucose they take up to CO_2 , which is lost from the body as a waste product. A growing tumor needs nutrients in abundance to provide the building blocks to make new macromolecules. Correspondingly, most tumors have a metabolism more similar to that of a growing embryo than to that of normal adult tissue. Tumor cells consume glucose avidly, importing it from the blood at a rate that can be as much as 100 times higher than neighboring normal cells. Moreover, only a small fraction of this imported glucose is used for production of ATP by oxidative phosphorylation. Instead, a great deal of lactate is produced, and many of the remaining carbon atoms derived from glucose are diverted for use as raw materials for synthesis of the proteins, nucleic acids, and lipids required for tumor growth (**Figure 20–12**).

This tendency of tumor cells to de-emphasize oxidative phosphorylation even when oxygen is plentiful, while at the same time taking up large quantities of glucose, can be shown to promote cancer cell growth and is called the *Warburg*

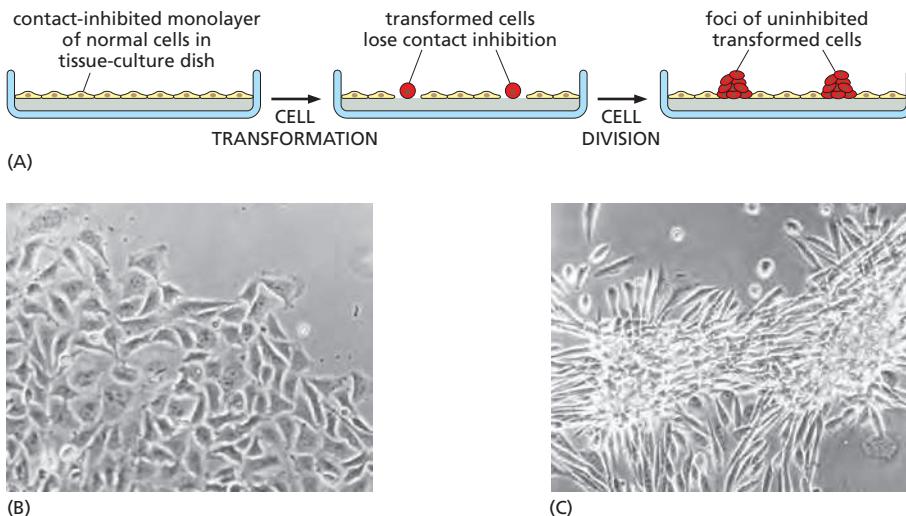


Figure 20–11 Loss of contact inhibition by cancer cells in cell culture. Most normal cells stop proliferating once they have carpeted the dish with a single layer of cells: proliferation seems to depend on contact with the dish, and to be inhibited by contacts with other cells—a phenomenon known as “contact inhibition.” Cancer cells, in contrast, usually disregard these restraints and continue to grow, so that they pile up on top of one another, as shown (**Movie 20.2**). (A) Schematic drawing. (B and C) Light micrographs of normal (B) and transformed (C) fibroblasts. (B and C, courtesy of Lan Bo Chen.)

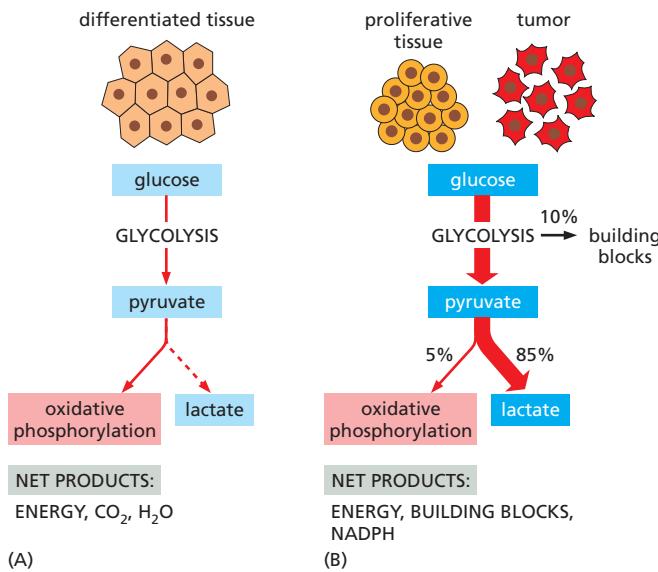


Figure 20–12 The Warburg effect in tumor cells reflects a dramatic change in glucose uptake and sugar metabolism.

(A) Cells that are not proliferating will normally oxidize nearly all of the glucose that they import from the blood to produce ATP through the oxidative phosphorylation that takes place in their mitochondria. Only when deprived of oxygen will these cells generate most of their ATP from glycolysis, converting the pyruvate produced to lactate in order to regenerate the NAD^+ that they need to keep glycolysis going (see Figure 2–47). (B) Tumor cells, by contrast, will generally produce abundant lactate even in the presence of oxygen. This results from a greatly increased rate of glycolysis that is fed by a very large increase in the rate of glucose import. In this way, tumor cells resemble the rapidly proliferating cells in embryos (and during tissue repair), which likewise require for biosynthesis a large supply of the small-molecule building blocks that can be produced from imported glucose (see also Figure 20–26).

effect—so named because Otto Warburg first noticed the phenomenon in the early twentieth century. It is this abnormally high glucose uptake that allows tumors to be selectively imaged in whole-body scans (see Figure 20–1), thereby providing a way to monitor cancer progression and responses to treatment.

Cancer Cells Have an Abnormal Ability to Survive Stress and DNA Damage

In a large multicellular organism, there are powerful safety mechanisms that guard against the trouble that can be caused by damaged and deranged cells. For example, internal disorder gives rise to danger signals in the faulty cell, activating protective devices that can eventually lead to apoptosis (see Chapter 18). To survive, cancer cells require additional mutations to elude or break through these defenses against cellular misbehavior.

Cancer cells are found to contain mutations that drive the cell into an abnormal state, where metabolic processes may be unbalanced and essential cell components may be produced in ill-matched proportions. States of this type, where the cell's homeostatic mechanisms are inadequate to cope with an imposed disturbance, are loosely referred to as states of *cell stress*. As one example, chromosome breakage and other forms of DNA damage are commonly observed during the development of cancer, reflecting the genetic instability that cancer cells display. Thus, to survive and divide without limit, a prospective cancer cell must accumulate mutations that disable the normal safety mechanisms that would otherwise induce a cell that is stressed, in this or in other ways, to commit suicide. In fact, 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 (Figure 20–13).

While cancer cells tend to avoid apoptosis, this does not mean that they rarely die. On the contrary, in the interior of a large solid tumor, cell death often occurs on a massive scale: living conditions are difficult, with severe competition among the cancer cells for oxygen and nutrients. Many die, but typically much more by necrosis than by apoptosis (Figure 20–14). The tumor grows because the cell birth rate outpaces the cell death rate, but often by only a small margin. For this reason, the time that a tumor takes to double in size can be far longer than the cell-cycle time of the tumor cells.

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

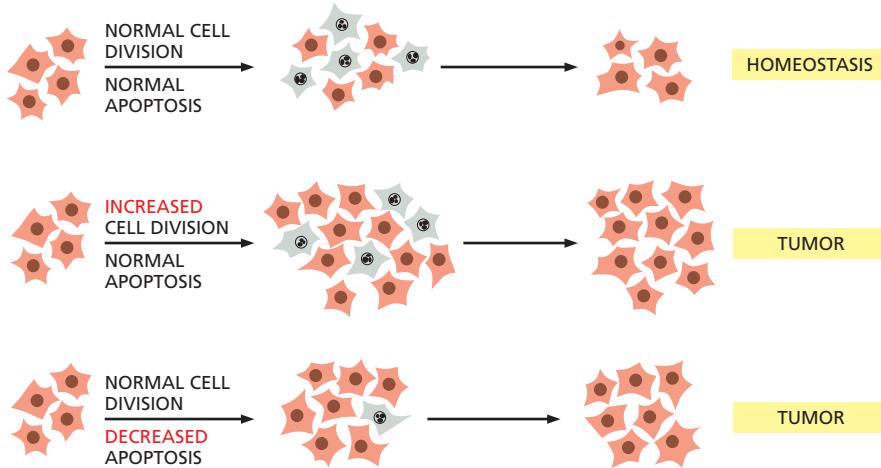


Figure 20–13 Both increased cell division and decreased apoptosis can contribute to tumorigenesis. In normal tissues, apoptosis balances cell division to maintain homeostasis (see Movie 18.1). During the development of cancer, either an increase in cell division or an inhibition of apoptosis can lead to the increased cell numbers important for tumorigenesis. The cells fated to undergo apoptosis are gray in this diagram. Both an increase in cell division and a decrease in apoptosis normally contribute to tumor growth.

after a certain number of population doublings (25–50 for human fibroblasts, for example). This cell-division-counting mechanism is termed **replicative cell senescence**, and it generally depends on the progressive shortening of the telomeres at the ends of chromosomes, a process that eventually changes their structure (discussed in Chapter 17). As discussed in Chapter 5, the replication of telomere DNA during S phase depends on the enzyme *telomerase*, which maintains a special telomeric DNA sequence that promotes the formation of protein cap structures to protect chromosome ends. Because many proliferating human cells (stem cells being an exception) are deficient in telomerase, their telomeres shorten with every division, and their protective caps deteriorate, creating a DNA damage signal. Eventually, the altered chromosome ends can trigger a permanent cell-cycle arrest, causing a normal cell to die.

Human cancer cells avoid replicative cell senescence in one of two ways. They can maintain the activity of telomerase as they proliferate, so that their telomeres do not shorten or become uncapped, or they can evolve an alternate mechanism based on homologous recombination (called ALT) for elongating their chromosome ends. Regardless of the strategy used, the result is that the cancer cells continue to proliferate under conditions when normal cells would stop.

The Tumor Microenvironment Influences Cancer Development

While the cancer cells in a tumor are the bearers of dangerous mutations and are often grossly abnormal, the other cells in the tumor—especially those of the supporting connective tissue, or **stroma**—are far from passive bystanders. The

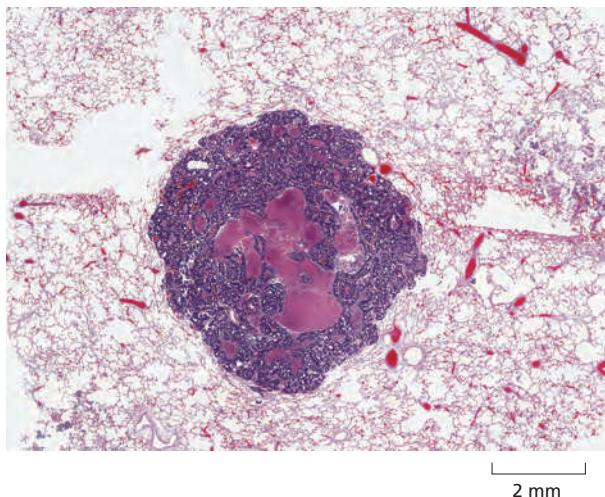


Figure 20–14 Cross-section of a colon adenocarcinoma that has metastasized to the lung. This tissue slice shows well-differentiated colorectal cancer cells forming cohesive glands in the lung. The metastasis has central pink areas of necrosis where dying cancer cells have outgrown their blood supply. Such anoxic regions are common in the interior of large tumors. (Courtesy of Andrew J. Connolly.)

development of a tumor relies on a two-way communication between the tumor cells and the tumor stroma, just as the normal development of epithelial organs relies on communication between epithelial cells and mesenchymal cells (discussed in Chapter 22).

The stroma provides a framework for the tumor. It is composed of normal connective tissue containing fibroblasts and inflammatory white blood cells, as well as the endothelial cells that form blood and lymphatic vessels with their attendant pericytes and smooth muscle cells (Figure 20–15). As a carcinoma progresses, the cancer cells induce changes in the stroma by secreting signal proteins that alter the behavior of the stromal cells, as well as proteolytic enzymes that modify the extracellular matrix. The stromal cells in turn act back on the tumor cells, secreting signal proteins that stimulate cancer cell growth and division as well as proteases that further remodel the extracellular matrix. In these ways, the tumor and its stroma evolve together, like weeds and the ecosystem that they invade, and the tumor becomes dependent on its particular stromal cells. Experiments using mice indicate that the growth of some transplanted carcinomas depends on the tumor-associated fibroblasts and normal fibroblasts will not do. Such environmental requirements help to protect us from cancer, as we discuss next in considering the critical phenomenon called metastasis.

Cancer Cells Must Survive and Proliferate in a Foreign Environment

Cancer cells generally need to spread and multiply at new sites in the body in order to kill us, through a process called metastasis. This is the most deadly—and least understood—aspect of cancer, being responsible for 90% of cancer-associated deaths. By spreading through the body, a cancer becomes almost impossible to eradicate by either surgery or local irradiation. **Metastasis** is itself a multi-step process: the cancer cells first have to invade local tissues and vessels, move through the circulation, leave the vessels, and then establish new cellular colonies at distant sites (Figure 20–16). Each of these events is complex, and most of the molecular mechanisms involved are not yet clear.

For a cancer cell to become dangerous, 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 20–8). 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. For carcinomas, this change resembles the *epithelial-mesenchymal transition (EMT)* that occurs in some epithelial tissues during normal development (see p. 1042).

The next step in metastasis—the establishment of colonies in distant organs—begins with entry into the circulation: the invasive cancer cells must penetrate the

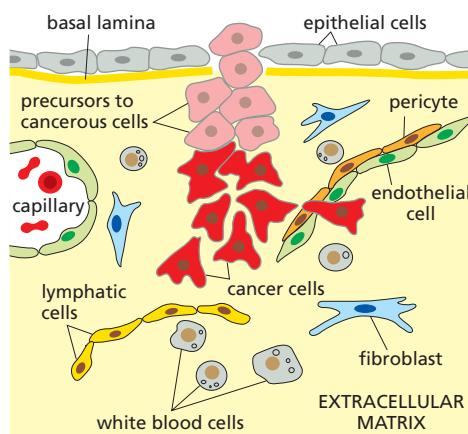


Figure 20–15 The tumor microenvironment plays a role in tumorigenesis. Tumors consist of many cell types, including cancer cells, endothelial cells, pericytes (vascular smooth muscle cells), fibroblasts, and inflammatory white blood cells. Communication among these and other cell types plays an important part in tumor development. Note, however, that only the cancer cells are thought to be genetically abnormal in a tumor.

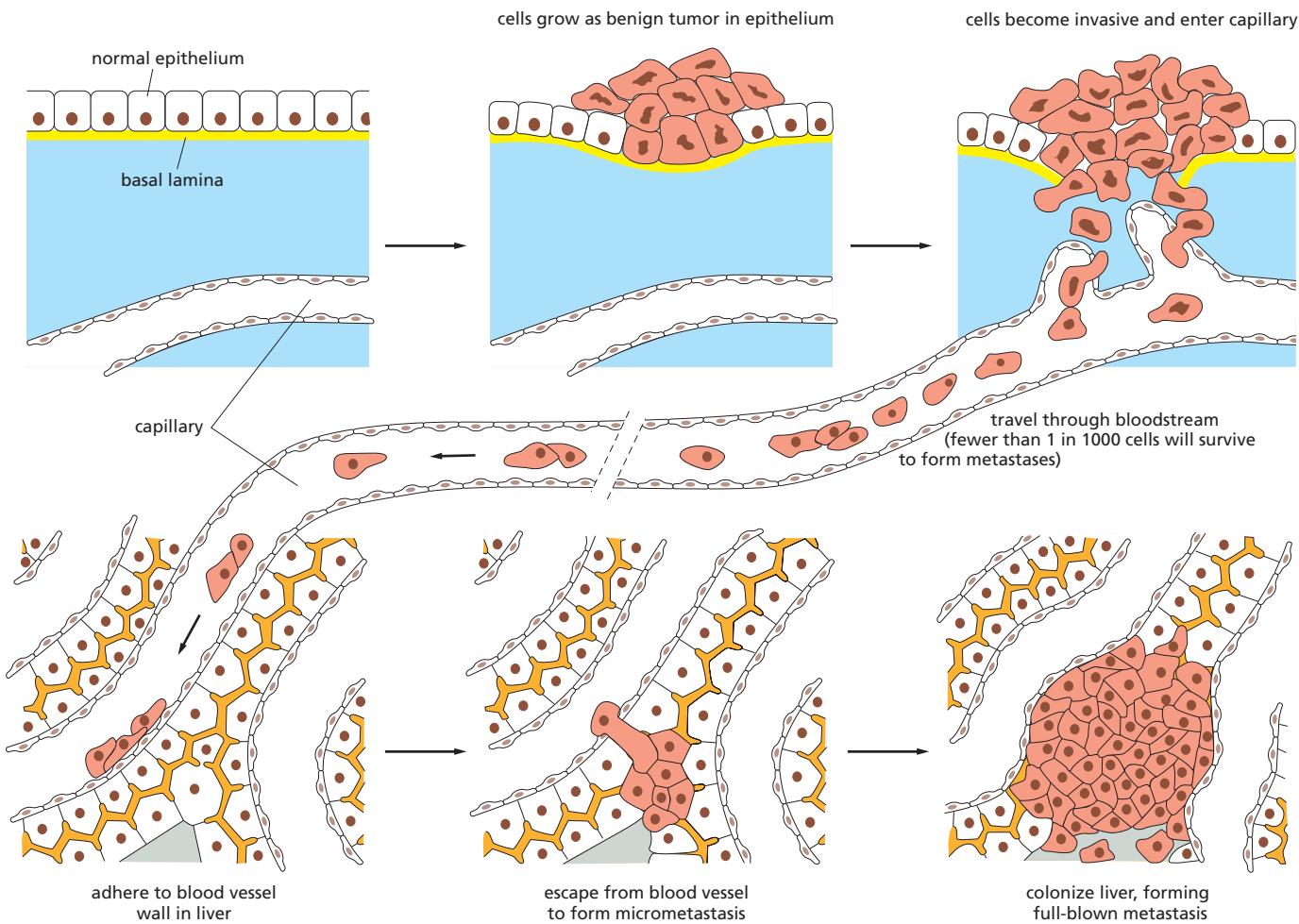


Figure 20–16 Steps in the process of metastasis. This example illustrates the spread of a tumor from an organ such as the bladder to the liver. Tumor cells may enter the bloodstream directly by crossing the wall of a blood vessel, as diagrammed here, or, more commonly perhaps, by crossing the wall of 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.

wall of a blood or lymphatic vessel. Lymphatic vessels, being larger and having more flimsy walls than blood vessels, allow cancer cells to enter in small clumps; such clumps may then become trapped in lymph nodes, giving rise to lymph-node metastases. The cancer cells that enter blood vessels, in contrast, seem to do so singly. With modern techniques for sorting cells according to their surface properties, it has become possible in some cases to detect these *circulating tumor cells* (CTCs) in samples of blood from cancer patients, even though they are only a minute fraction of the total blood-cell population. These cells, in principle at least, provide a useful sample of the tumor-cell population for genetic analysis.

Of the cancer cells that enter the lymphatics or bloodstream, only a tiny proportion succeed in making their exit, settling in new sites, and surviving and proliferating there as founders of metastases. Experiments show that fewer than one in thousands, perhaps one in millions, manage this feat. The final step of colonization seems to be the most difficult: like the Vikings who landed on the inhospitable shores of Greenland, the migrant cells may fail to survive in the alien environment; or they may only thrive there for a short while to found a little colony—a *micrometastasis*—that then dies out ([Movie 20.3](#)).

Many cancers are discovered before they have managed to found metastatic colonies and can be cured by destruction of the primary tumor. But on occasion,

an undetected micrometastasis will remain dormant for many years, only to reveal its presence by erupting into growth to form a large secondary tumor long after the primary tumor has been removed.

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, cancers all share some common features. By definition, they all ignore or misinterpret normal social controls so as to proliferate and spread where normal cells would not. These defining properties are commonly combined with other features that help the miscreants to arise and thrive. A list of the key attributes of cancer cells in general would include the following, all of which we have just discussed:

1. They grow (biosynthesize) when they should not, aided by a metabolism shifted from oxidative phosphorylation toward aerobic glycolysis.
2. They go through the cell-division cycle when they should not.
3. They escape from their home tissues (that is, they are invasive) and survive and proliferate in foreign sites (that is, they metastasize).
4. They have abnormal stress responses, enabling them to survive and continue dividing in conditions of stress that would arrest or kill normal cells, and they are less prone than normal cells to commit suicide by apoptosis.
5. They are genetically and epigenetically unstable.
6. They escape replicative cell senescence, either by producing telomerase or by acquiring another way of stabilizing their telomeres.

In the next section of the chapter, we examine the mutations and molecular mechanisms that underlie these and other properties of cancer cells.

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 additional mutations and epigenetic events, to become cancerous. 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. Their mutant genomes enable them to grow and divide in defiance of the signals that normally keep cell proliferation under tight control. As part of the evolutionary process of tumor progression, cancer cells acquire a collection of additional abnormalities, including defects in the controls that permanently stop cell division or induce apoptosis in response to cell stress or DNA damage, and in the mechanisms that normally keep cells from straying from their proper place. All of these changes increase the ability of cancer cells to survive, grow, and divide in their original tissue and then to metastasize, founding new colonies in foreign environments. The evolution of a tumor also depends on other cells present in the tumor microenvironment, collectively called stromal cells, that the cancer attracts and manipulates.

Since many changes are needed to confer this collection of asocial behaviors, it is not surprising that most cancer cells are genetically and/or epigenetically unstable. This instability is thought to be 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.

CANCER-CRITICAL GENES: HOW THEY ARE FOUND AND WHAT THEY DO

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 those genes with the cancer-promoting changes among all the other genes in the cancerous cells? A typical cancer depends on a whole set of mutations and epigenetic changes—usually a somewhat different set in each individual patient. In addition, a given cancer cell will also contain a large number of somatic mutations that are accidental by-products—so-called *passengers* rather than *drivers*—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 of the genes that are repeatedly altered in human cancers have been identified over the past 40 years. We will call such genes, for want of a better term, **cancer-critical genes**, meaning all genes whose alteration contributes to the causation or evolution of cancer by driving tumorigenesis.

In this section, we shall first discuss how cancer-critical genes are identified. We shall then examine their functions and the parts they play in conferring on cancer cells the properties outlined in the first part of the chapter. We shall 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.

The Identification of Gain-of-Function and Loss-of-Function Cancer Mutations Has Traditionally Required 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**. In either case, the mutation may lead toward cancer directly (by causing cells to proliferate when they should not) or indirectly—for example, by causing genetic or epigenetic instability and so hastening the occurrence of other inherited changes that directly stimulate tumor growth. Those genes whose alteration results in genomic instability represent a subclass of cancer-critical genes that are sometimes called *genome maintenance genes*.

As we shall see, mutations in oncogenes and tumor suppressor genes can have similar effects in promoting the development of cancer; overproduction of a signal for cell proliferation, for example, can result from either kind of mutation. 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 that led to the discovery of these two categories of genes, however, are quite different.

The 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 20–17A**). 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 or experimental animal. 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 20–17B**). This calls for a different experimental approach, one focusing on discovering what is *missing* in the cancer cell.

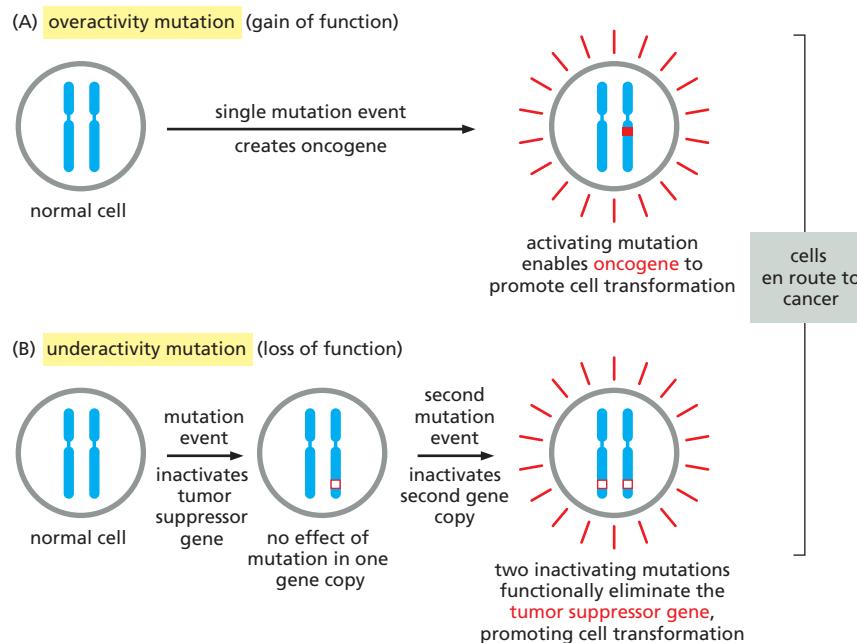


Figure 20–17 Cancer-critical mutations fall into two readily distinguishable categories, dominant and recessive.

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.

We begin by discussing some examples of each class of cancer-critical genes to illustrate basic principles. These examples are chosen also for their historical importance: the experiments that led to their discovery—at different times and by different methods—marked turning points in the understanding of cancer.

Retroviruses Can Act as Vectors for Oncogenes That Alter Cell Behavior

The search for the genetic causes of human cancer took a devious route, beginning with clues that came from the study of **tumor viruses**. Although viruses are involved only in a minority of human cancers, a set of viruses that infect animals provided critical early tools for studying cancer.

One of the first animal viruses to be implicated in cancer was discovered over 100 years ago in chickens, when an infectious agent that causes connective-tissue tumors, or sarcomas, was characterized as a virus—the *Rous sarcoma virus*. Like all the other *RNA tumor viruses* discovered since, it is a **retrovirus**. When it infects a cell, its RNA genome 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. Something in the DNA inserted by the Rous sarcoma virus made the host cells cancerous, but what was it? The answer was a surprise. It turned out to be a piece of DNA that was unnecessary for the virus's own survival or reproduction; instead, it was a passenger, a gene called *v-Src*, that the virus had picked up on its travels. *v-Src* was unmistakably similar, but not identical, to a gene—*c-Src*—that was discovered in the normal vertebrate genome. *c-Src* had evidently been caught up accidentally by the retrovirus from the genome of a previously infected host cell, and it had undergone mutation in the process to become an oncogene (*v-Src*).

This Nobel Prize-winning finding was followed by a flood of discoveries of other viral oncogenes carried by retroviruses that cause cancer in nonhuman animals. Each such oncogene turned out to have a counterpart proto-oncogene in the normal vertebrate genome. As was the case for *Src*, these other oncogenes generally differed from their normal counterparts, either in structure or in level of expression. But how did this relate to typical human cancers, most of which are not infectious and in which retroviruses play no part?

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

In an attempt to answer the above question, other researchers searched directly for oncogenes in the genomes of human cancer cells. They did this by searching for DNA fragments from cancer cells that could provoke uncontrolled proliferation when introduced into noncancerous cell lines. As tester cells for the assay, cell lines derived from mouse fibroblasts were used. These cells had been previously selected for their ability to proliferate indefinitely in culture, and they are thought to already contain alterations that take them part of the way toward malignancy. For this reason, the addition of a single oncogene can sometimes be enough to produce a dramatic effect.

When DNA was extracted from the human tumor cells, broken into fragments, and introduced into the cultured cells, occasional colonies of abnormally proliferating cells began to appear in the culture dish. These cells showed a transformed phenotype, outgrowing the untransformed cells in the culture and piling up in layer upon layer (see Figure 20–11). Each colony was a clone originating from a single cell that had incorporated a DNA fragment that drove cancerous behavior. This fragment, which carried markers of its human origin, could be isolated from the transformed cultured mouse cells. And once isolated and sequenced, it could be recognized: it contained a human version of a gene already known from study of a retrovirus that caused tumors in rats—an oncogene called *v-Ras*.

The newly discovered oncogene was clearly derived by mutation from a normal human gene, one of a small family of proto-oncogenes called *Ras*. 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 transformed 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 (see Movie 15.7). 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. One or another of the three human *Ras* family members is mutated in perhaps 30% of all human cancers. *Ras* genes are thus among the most important of all cancer-critical genes.

Genes Mutated in Cancer Can Be Made Overactive in Many Ways

Figure 20–18 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

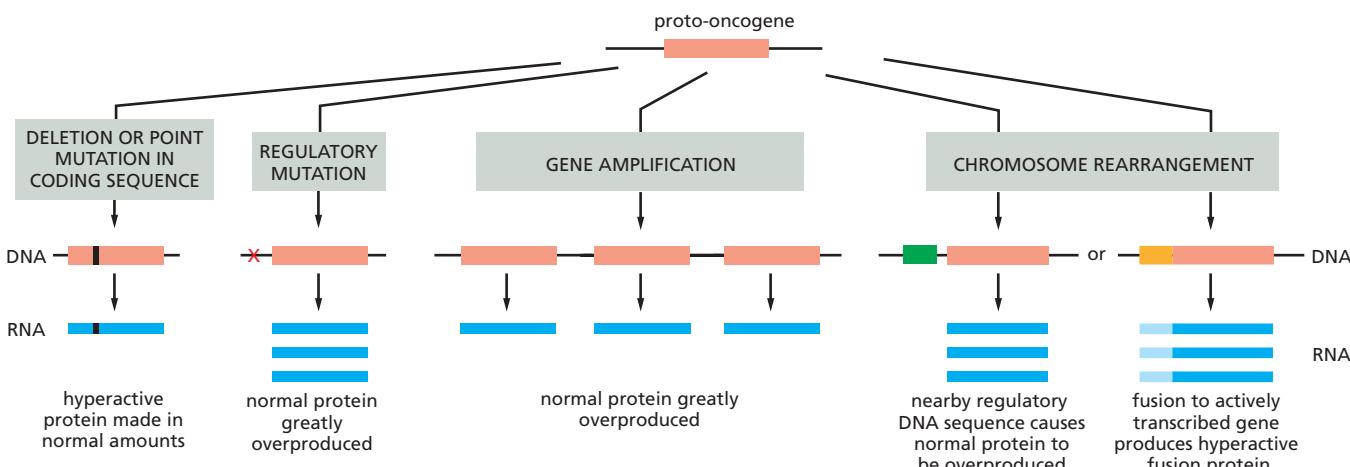


Figure 20–18 The types of accidents that can convert a proto-oncogene into an oncogene.

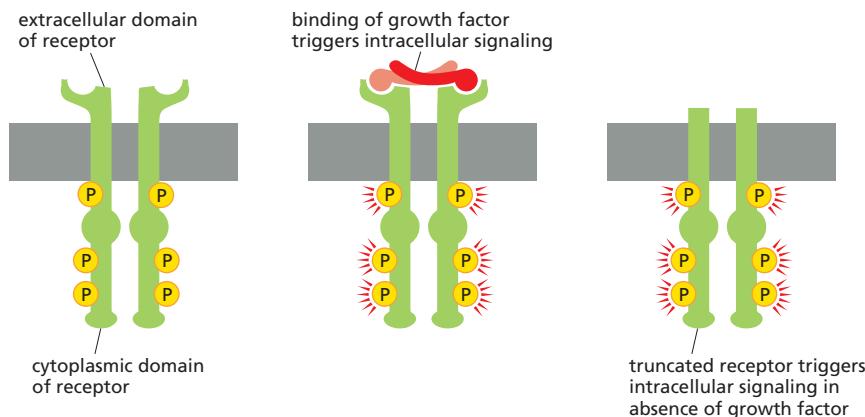


Figure 20-19 Mutation of the epidermal growth factor (EGF) receptor can make it active even in the absence of EGF, and consequently oncogenic. Only one of the possible types of activating mutations is illustrated here.

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. (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.

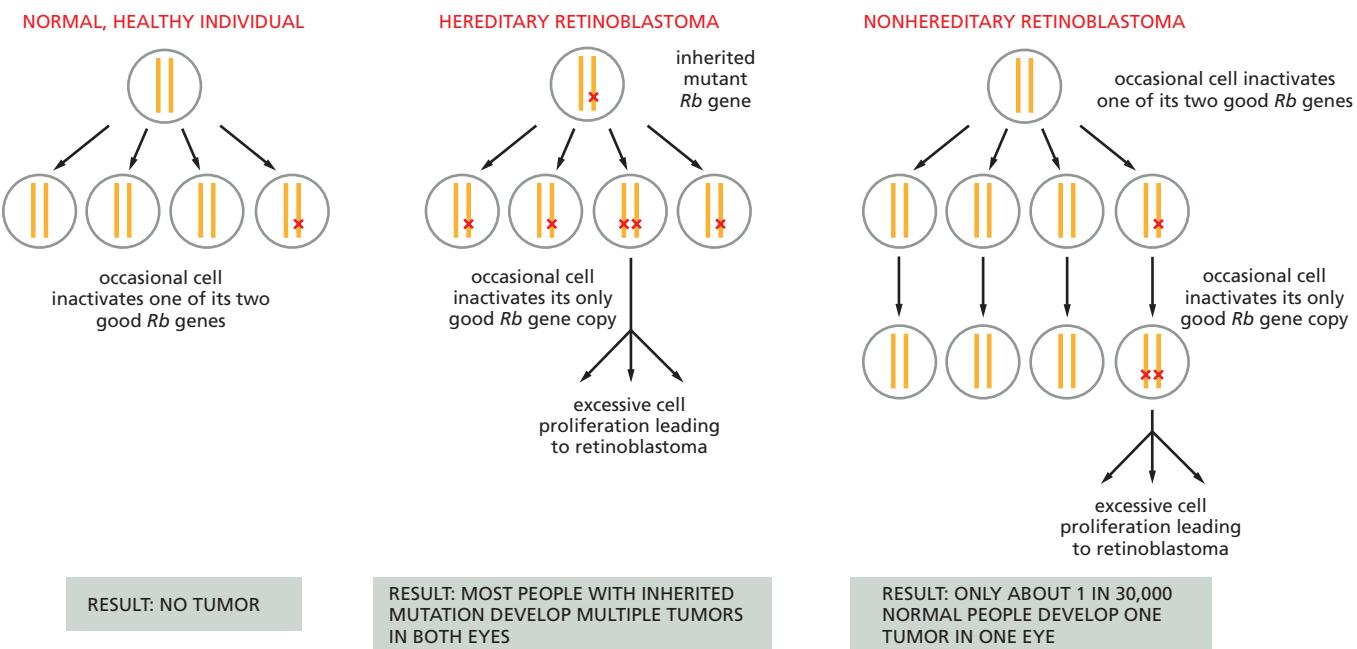
As one example, the receptor for the extracellular signal protein *epidermal growth factor (EGF)* can be activated by a deletion that removes part of its extracellular domain, causing it to be active even in the absence of EGF (Figure 20-19). It thus produces an inappropriate stimulatory signal, like a faulty doorbell that rings even when nobody is pressing the button. Mutations of this type are frequently found in the most common type of human brain tumor, called glioblastoma.

As another example, the *Myc protein*, which acts in the nucleus to stimulate cell growth and division (see Chapter 17), generally contributes to cancer by being overproduced in its normal form. 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. Or a point mutation can stabilize the protein, which normally turns over very rapidly. 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. Different specific chromosome translocations are common in other cancers.

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 importance.

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 20–20).

The *Rb* gene is also missing in several common types of sporadic cancer, including carcinomas of 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 17–61). 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.

Both Genetic and Epigenetic Mechanisms Can Inactivate Tumor Suppressor Genes

For tumor suppressor genes, it is their inactivation that is dangerous. This inactivation can occur in many ways, with different combinations of mishaps serving 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 mechanism:

Figure 20–20 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 a single line of cells.

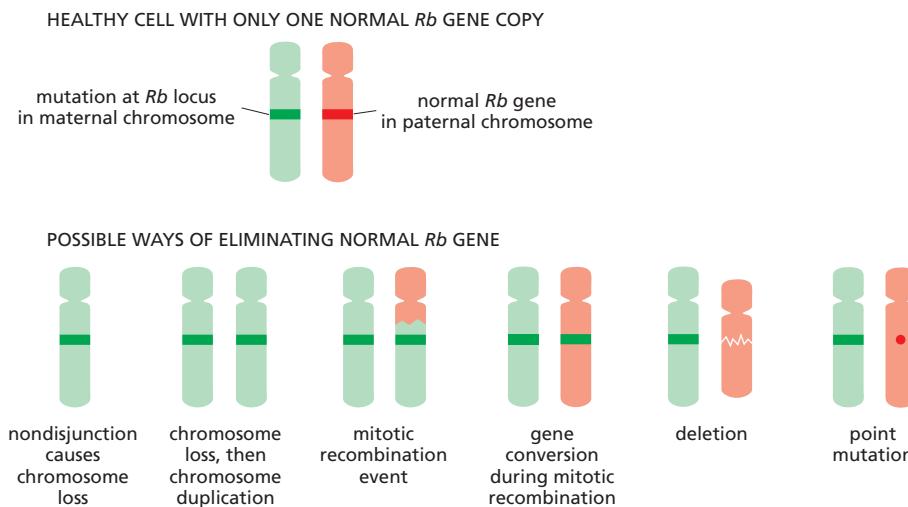


Figure 20-21 Six ways of losing the remaining good copy of a tumor suppressor gene through a change in DNA sequences. A cell that is defective in only one of its two copies of a tumor suppressor gene—for example, the *Rb* gene—usually behaves as a normal, healthy cell; the diagrams below show how this 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 20-22. (After W.K. Cavenee et al., *Nature* 305:779–784, 1983. With permission from Macmillan Publishers Ltd.)

the chromosome carrying the remaining normal copy may be lost from the cell through errors in chromosome segregation; or the normal gene, along with neighboring genetic material, may be replaced by a mutant version through either a *mitotic recombination* event or a *gene conversion* that accompanies it (see p. 286).

Figure 20-21 summarizes the range of ways in which the remaining good copy of a tumor suppressor gene can be lost through a DNA sequence change, 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.

Epigenetic changes provide another important way to permanently inactivate a tumor suppressor gene. Most commonly, the gene may become packaged into heterochromatin and/or the C nucleotides in CG sequences in its promoter may become methylated in a heritable manner (see pp. 404–405). These mechanisms can irreversibly silence the gene in a cell and in all of its progeny. Analysis of methylation patterns in cancer genomes shows 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 20-22**).

Systematic Sequencing of Cancer Cell Genomes Has Transformed Our Understanding of the Disease

Methods such as those we have described above shone a spotlight on a set of cancer-critical genes that were identified in a piecemeal fashion. Meanwhile, the rest of the cancer cell genome remained in darkness: it was a mystery how many other mutations might lurk there, of what types, in which varieties of cancer, at what frequencies, with what variations from patient to patient, and with what consequences. With the sequencing of the human genome and the dramatic advances in DNA sequencing technology (see Panel 8-1, pp. 478–481), it has become possible to see the whole picture—to view cancer cell genomes in their entirety. This transforms our understanding of the disease.

Cancer cell genomes can be scanned systematically in several different ways. At one extreme—the most costly, but no longer prohibitively so—one can determine a tumor's complete genome sequence. More cheaply, one can focus just on the 21,000 or so genes in the human genome that code for protein (the so-called *exome*), looking for mutations in the cancer cell DNA that alter the amino acid sequence of the product or prevent its synthesis (**Figure 20-23**). There are also efficient techniques to survey the genome for regions that have undergone

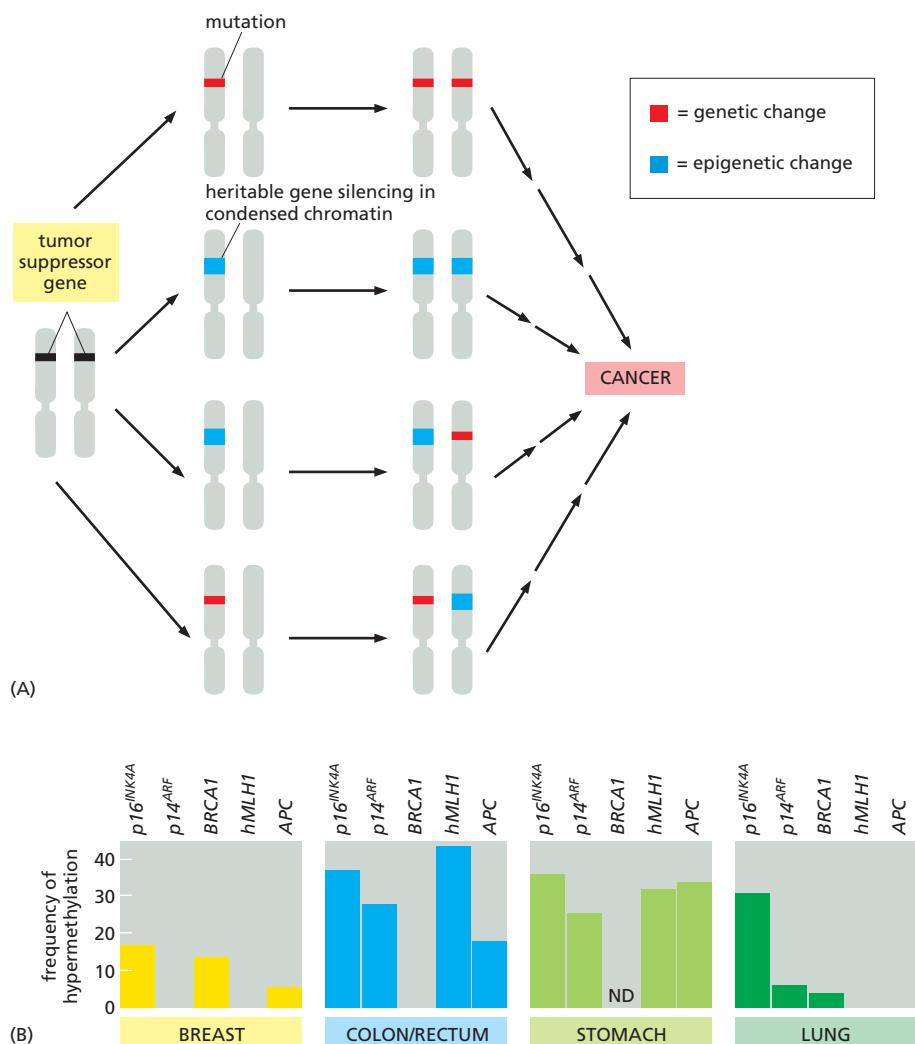


Figure 20-22 The pathways leading to loss of tumor suppressor gene function in cancer involve both genetic and epigenetic changes. (A) As indicated, the changes that silence tumor suppressor genes can occur in any order. Both DNA methylation and the packaging of a gene into condensed chromatin can prevent its expression in a way that is inherited when a cell divides (see Figure 4-44). (B) The frequency of gene silencing by hypermethylation observed in four different types of cancer. The five genes listed at the top can all function as tumor suppressor genes; *BRCA1* and *hMLH1* affect genome stability and are in the subclass known as genome maintenance genes. ND, no data. (Adapted from M. Esteller et al., *Cancer Res.* 61:3225–3229, 2001.)

deletion or duplication, without the need for complete sequence information. The genome can be scanned for epigenetic changes. And finally, alterations in levels of gene expression can be systematically determined by analysis of mRNAs (see Figure 7-3). These approaches generally involve comparing cancer cells with normal controls—ideally, noncancerous cells originating in the same tissue and from the same patient.

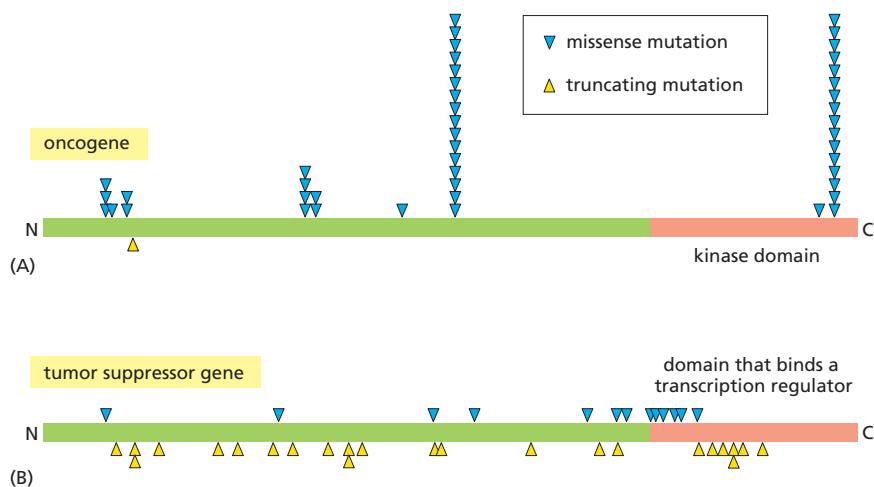


Figure 20-23 The distinct types of DNA sequence changes found in oncogenes compared to tumor suppressor genes. In this diagram, mutations that change an amino acid are denoted by blue arrowheads, whereas mutations that truncate the polypeptide chain are marked by yellow arrowheads. (A) As in this example, oncogene mutations can be detected by the fact that the same nucleotide change is repeatedly found among the missense mutations in a gene. (B) For tumor suppressor genes, by contrast, missense mutations that abort protein synthesis by creating stop codons predominate. (Adapted from B. Vogelstein et al., *Science* 339:1546–1558, 2013.)

Many Cancers Have an Extraordinarily Disrupted Genome

Cancer genome analysis reveals, first of all, the scale of gross genetic disruption in cancer cells. This varies greatly from one type of cancer and one cancer patient to another, both in severity and in character. In some cases, the karyotype—the set of chromosomes as they appear at mitosis—is normal or nearly so, but many point mutations are detected in individual genes, suggesting a failure of the repair mechanisms that normally correct local errors in the replication or maintenance of DNA sequences. Often, however, the karyotype is severely disordered, with many chromosome breaks and rearrangements. In some breast cancers, for example, genome sequencing reveals an astonishing scene of genetic chaos (**Figure 20–24**), with hundreds of chromosome breaks and translocations, resulting in many deletions, duplications, and amplifications of parts of the genome. In such cells, the normal machinery for avoidance or repair of DNA double-strand breaks is evidently somehow defective, destabilizing the genome by giving rise to broken chromosomes whose fragments then rejoin in random combinations. From the pattern of changes, one can infer that this disruptive process has occurred repeatedly during the evolution of the tumor, with a progressive increase of genetic disorder. Breast cancers showing the most extreme chromosome disorder are usually hard to treat and have a gloomy prognosis.

One survey of more than 3000 individual cancer specimens showed that on average 24 separate blocks of genetic material were duplicated in each tumor, amounting to 17% of the normal genome, and 18 blocks were deleted, amounting to 16% of the normal genome. Many of these changes were found repeatedly, suggesting that they contain cancer-critical genes whose loss (tumor suppressor genes) or gain (oncogenes) confers a selective advantage.

Whole-genome analysis also helps to explain some cancers that seem, at first sight, to be exceptions to the general rules. An example is retinoblastoma, with its early onset during childhood. If cancers in general require an accumulation of many genetic changes and are thus diseases of old age, what makes retinoblastoma different? Whole-genome sequencing confirms that in retinoblastoma, the tumor cells contain loss-of-function mutations in the *Rb* gene; but, amazingly, they contain practically no mutations or genome rearrangements that affect any other oncogene or tumor suppressor gene. Instead, they contain many epigenetic modifications, which alter the level of expression of many known cancer-critical genes—as many as 15 in one well-analyzed case.

Many Mutations in Tumor Cells are Merely Passengers

Cancer cells generally contain many mutations in addition to gross chromosome abnormalities: point mutations can be scattered over the genome as a whole at a rate of about one per million nucleotide pairs, in addition to the abnormalities

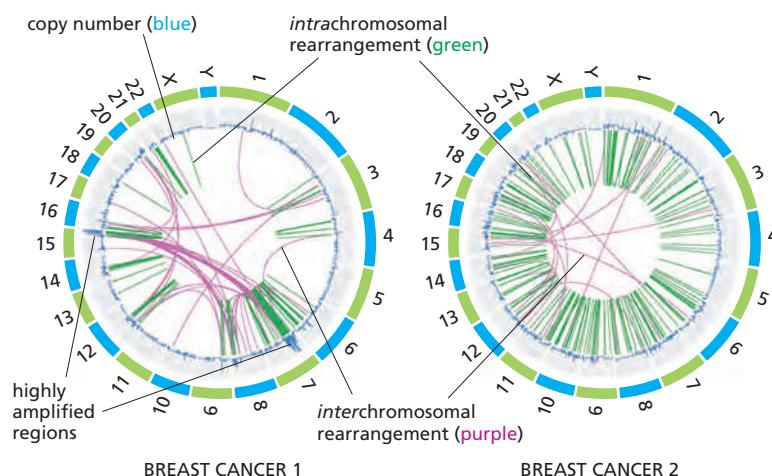


Figure 20–24 The chromosomal rearrangements in breast cancer cells. The results of an extensive DNA sequencing analysis performed on two different primary tumors are displayed as “Circos plots.” In each plot, the reference DNA sequences of the 22 autosomes and single sex chromosome (X) of a normal human female (3.2 billion nucleotide pairs) are aligned end-to-end to form a circle. Colored lines within the circle are then used to indicate the chromosome alterations found in the particular primary tumor. As indicated, purple lines connect sites at which two different chromosomes have become joined to create an interchromosomal rearrangement, while green lines connect the sites of rearrangements found within a single chromosome. The intrachromosomal rearrangements can be seen to predominate, and most join neighboring sections of DNA that were originally located within 2 million nucleotide pairs of each other. The increases in copy number, shown in blue, reveal the amplified DNA sequences (see the highly amplified regions indicated). (Adapted from P.J. Stephens et al., *Nature* 462:1005–1010, 2009.)

attributed to chromosome breakage and rejoining. Systematic surveys of the protein-coding genes in common solid tumors—such as those of the breast, colon, brain, or pancreas—have revealed that an average of 33 to 66 genes have undergone somatic mutation affecting the sequence of their protein product. Mutations in noncoding regions of the genome are much more numerous, as one would expect from the much larger fraction of the genome that noncoding DNA represents. But they are considerably more difficult to interpret.

The high frequency of mutations testifies to the genetic instability of many cancer cells, but it leaves us with a difficult problem. How can we discover which of the mutations are **drivers** of cancer—that is, causal factors in the development of the disease—and which are merely **passengers**—mutations that happen to have occurred in the same cell as the driver mutations, thanks to genetic instability, but are irrelevant to the development of the disease? A simple criterion is based on frequency of occurrence. Driver mutations affecting a gene that plays a part in the disease will be seen repeatedly, in many different patients. In contrast, passenger mutations, occurring at more-or-less random locations in the genome and conferring no selective advantage on the cancer cell, are unlikely to be found in the same genes in different patients.

Figure 20–25 shows the results of an analysis of this sort for a large sample of colorectal cancers. The different sites in the genome are laid out on a two-dimensional array, with chromosome serial number along one axis and position within each chromosome along the other. The frequency with which mutations are encountered is shown by height above this plane, creating a mutation “landscape” with mountains (sites where mutations are found in a large proportion of the tumors in the sample), hills (where mutations are found less frequently but still more often than would be expected for a random scattering over the genome), and hillocks (sites of occasional mutations, occurring at a frequency no higher than would be expected for mutations scattered at random in each individual tumor). The mountains and the hills are strong candidates to be the sites of driver mutations—in other words, sites of cancer-critical genes; the hillocks are likely to correspond to passengers. Indeed, many of the mountains and hills turn out to be sites of known oncogenes or tumor suppressor genes, whereas the hillocks mostly correspond to genes that have no known or probable role in causation of cancer. Of course, some hillocks may correspond to genes that are mutated in only a few rare patients but are nevertheless cancer-critical for them.

About One Percent of the Genes in the Human Genome Are Cancer-Critical

From studies such as the one just described, it is estimated that the number of driver mutations for an individual case of cancer (the sum of meaningful epigenetic and genetic changes in both coding sequences and regulatory regions) is typically on the order of 10, explaining why cancer progression generally involves an increase in genetic and/or epigenetic instability that enhances the rate of such changes.

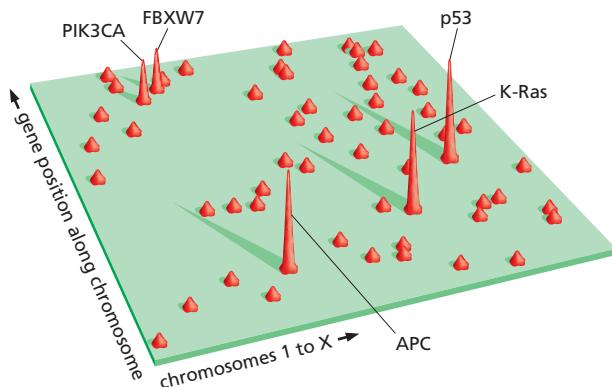


Figure 20–25 The mutation landscape in colorectal cancer. In this two-dimensional representation of the human genome, the green surface depicts the 22 human autosomes plus the X sex chromosome as being laid out side-by-side in numerical order from left to right, with the DNA sequence of each chromosome running from back to front. The mountains represent the locations of genes mutated with high frequency in different, independent tumors. As indicated, these are suspected driver mutations in the adenomatous polyposis coli (APC), K-Ras, p53, phosphoinositide 3-kinase (PIK3CA), and ubiquitin ligase (FBXW7) proteins. (Adapted from L.D. Wood et al., *Science* 318:1108–1113, 2007.)

By compiling the data for different types of cancer, each with its own range of identified driver mutations, we can develop a comprehensive catalog of genes that are strongly suspected to be cancer-critical. Current estimates put the total number of such genes at about 300, about 1% of the genes in the human genome. These cancer-critical genes are amazingly diverse. Their products include secreted signal proteins, transmembrane receptors, GTP-binding proteins, protein kinases, transcription regulators, chromatin modifiers, DNA repair enzymes, cell-cell adhesion molecules, cell-cycle controllers, apoptosis regulators, scaffold proteins, metabolic enzymes, components of the RNA splicing machinery, and more besides. All these are susceptible to mutations that can contribute, in one way or another, in one tissue or another, to the evolution of cells with the cancerous properties that we listed earlier on page 1103.

Clearly, the molecular changes that cause cancer are complex. As we now explain, however, the complexity is not quite as daunting as it may initially seem.

Disruptions in a Handful of Key Pathways Are Common to Many Cancers

Some genes, like *Rb* and *Ras*, are mutated in many cases of cancer and in cancers of many different types. The involvement of genes such as *Rb* and *Ras* in cancer is no surprise, now that we understand their normal functions: they control fundamental processes of cell division and growth. But even these common culprits feature in considerably less than half of individual cases. What is happening to the control of these processes in the many cases of cancer where, for example, *Rb* is intact or *Ras* is not mutated? What part do mutations in the hundreds of other cancer-critical genes play in the development of the disease? With our increasing knowledge of the normal functions of the genes in the human genome, it is becoming easier to see patterns in the catalogued driver mutations and to give some simplifying answers to these questions.

Glioblastoma—the commonest type of human brain tumor—provides a good example. Analysis of the genomes of tumor cells from 91 patients identified a total of at least 79 genes that were mutated in more than one individual. The normal functions of most of these genes were known or could be guessed, allowing them to be assigned to specific biochemical or regulatory pathways. Three functional groupings stood out, accounting for a total of 21 of the recurrently mutated genes. One of these groupings consisted of genes in the *Rb pathway* (that is, *Rb* itself, along with genes that directly regulate *Rb*); this pathway governs initiation of the cell-division cycle. Another consisted of genes in the same regulatory subnetwork as *Ras*—a more loosely defined system of genes referred to as the *RTK/Ras/PI3K pathway*, after three of its core components; this pathway serves to transmit signals for cell growth and cell division from the cell exterior into the heart of the cell. The third grouping consisted of genes in a pathway regulating responses to stress and DNA damage—the *p53 pathway*. We shall have more to say about each of these pathways below.

Out of all tumors, 74% had identifiable mutations in all three pathways. If one were to trace these three pathways further upstream and include all the components, known and unknown, on which they depend, this percentage would almost certainly be even higher. In other words, in almost every case of glioblastoma, there are mutations that disrupt each of three fundamental controls: the control of cell growth, the control of cell division, and the control of responses to stress and DNA damage.

Strikingly, in any given tumor-cell clone, there is a strong tendency for no more than one gene to be mutated in each pathway. Evidently, what matters for tumor evolution is the disruption of the control mechanism, and not the genetic means by which that is achieved. Thus, for example, in a patient whose tumor cells have no mutation in *Rb* itself, there is generally a mutation in some other component of the *Rb* pathway, producing a similar biological effect.

Similar patterns are seen in other types of cancers. A survey of many specimens of the major variety of ovarian cancer, for example, identified 67% of patients as

having mutations in the Rb pathway, 45% in the Ras/PI3K pathway (defined more narrowly than in the glioblastoma study), and more than 96% in the p53 pathway. Allowing for additional pathway components not included in the analysis, it seems that most cases of this type of cancer, too, have mutations disrupting the same three controls, leading to misregulated cell growth, misregulated cell proliferation, and abnormal disregard of stress and DNA damage. It seems that these three fundamental controls are subverted in one way or another in virtually every type of cancer.

We have devoted an entire chapter to the cell cycle and growth controls (Chapter 17). Some important details of the other two control pathways are reviewed next.

Mutations in the PI3K/Akt/mTOR Pathway Drive Cancer Cells to Grow

Cell proliferation is not simply a matter of 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 die or become too small to divide. Cells appear to require two separate signals to grow and divide (Figure 20–26). Cancer depends, therefore, not only on a loss of restraints on cell-cycle progression, but also on disrupted control of cell growth.

The phosphoinositide 3-kinase (PI 3-kinase)/Akt/mTOR 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,

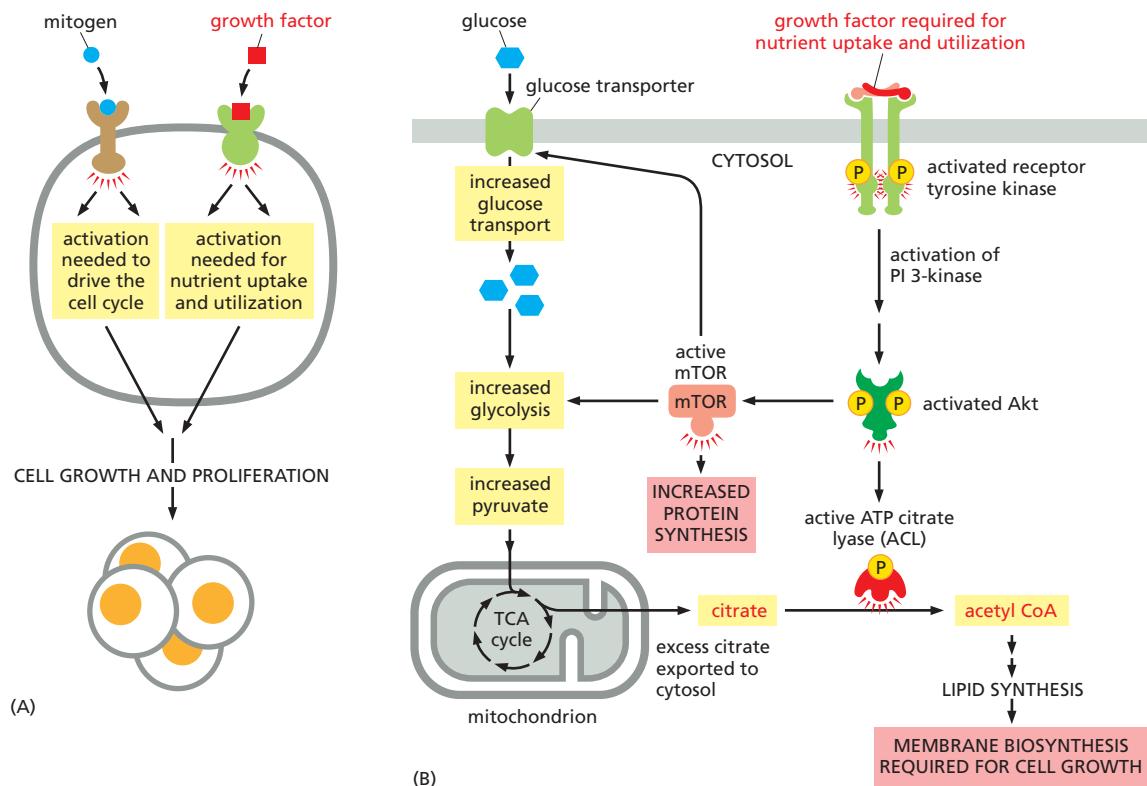


Figure 20–26 Cells seem to require two types of signals to proliferate. (A) In order to multiply successfully, most normal cells are suspected to require both extracellular signals that drive cell-cycle progression (shown here as blue mitogen) and extracellular signals that drive cell growth (shown here as red growth factor). How mitogens activate the Rb pathway to drive entry into the cell cycle is described in Figure 17–61. (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. TCA cycle indicates the tricarboxylic acid cycle (citric acid cycle).

normally activate 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 kinases Akt and mTOR not only stimulates protein synthesis (see Figure 17–64), but also greatly increases both glucose uptake and the production of the acetyl CoA in the cytosol required for cell lipid synthesis, as outlined in Figure 20–26B.

The abnormal activation of the PI 3-kinase/Akt/mTOR pathway, which normally occurs early in the process of tumor progression, helps to explain the excessive rate of glycolysis that is observed in tumor cells, known as the Warburg effect, as discussed earlier (see Figure 20–12). As expected from our previous discussion, cancers can activate this pathway in many different ways. Thus, for example, a growth factor receptor can become abnormally activated, as in Figure 20–19. Also very common in cancers is the loss of the PTEN phosphatase, an enzyme that normally suppresses the PI 3-kinase/Akt/mTOR pathway by dephosphorylating the PI (3,4,5) P₃ molecules that the PI 3-kinase forms (see pp. 859–861). *PTEN* is thus a common tumor-suppressor gene.

Of course, mutation is not the only way to overactivate the pathway: high levels of insulin in the circulation can have a similar effect. This may explain why the risk of cancer is significantly increased, by a factor of two or more, in people who are obese or have type 2 diabetes. Their insulin levels are abnormally high, driving cancer cell growth without need of mutation in the PI 3-kinase/Akt/mTOR pathway.

Mutations in the p53 Pathway Enable Cancer Cells to Survive and Proliferate Despite Stress and DNA Damage

That cancer cells must break the normal rules governing cell growth and cell division is obvious: that is part of the definition of cancer. It is not so obvious why cancer cells should also be abnormal in their response to stress and DNA damage, and yet this too is an almost universal feature. The gene that lies at the center of this response, the *p53* gene, is mutated in about 50% of all cases of cancer—a higher proportion than for any other known cancer-critical gene. When we include with *p53* the other genes that are closely involved in its function, we find that most cases of cancer harbor mutations in the *p53* pathway. Why should this be? To answer, we must first consider the normal function of this pathway.

In contrast to Rb, most cells in the body 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 has a function that is required only in special circumstances. In fact, cells raise their concentration of p53 protein in response to a whole range of conditions that have only one obvious thing in common: they are, from the cell's point of view, pathological, putting the cell in danger of death or serious injury. These conditions include DNA damage, putting the cell at risk from a faulty genome; telomere loss or shortening (see p. 1016), also dangerous to the integrity of the genome; hypoxia, depriving the cell of the oxygen it needs to keep its metabolism going; osmotic stress, causing the cell to swell or shrivel; and oxidative stress, generating dangerous levels of highly reactive free radicals.

Yet another form of stress that can activate the p53 pathway arises, it seems, when regulatory signals are so intense or uncoordinated as to drive the cell beyond its normal limits and into a danger zone where its mechanisms of control and coordination break down, as in an engine driven badly or too fast. The p53 concentration rises, for example, when *Myc* is overexpressed to oncogenic levels.

All these circumstances call for desperate action, which may take either of two forms: the cell can block any further progress through the division cycle in order to take time out to repair or recover from the pathological condition; or it can accept that it must die, and do so in a way that minimizes damage to the organism. A good death, from this point of view, is a death by apoptosis. In apoptosis,

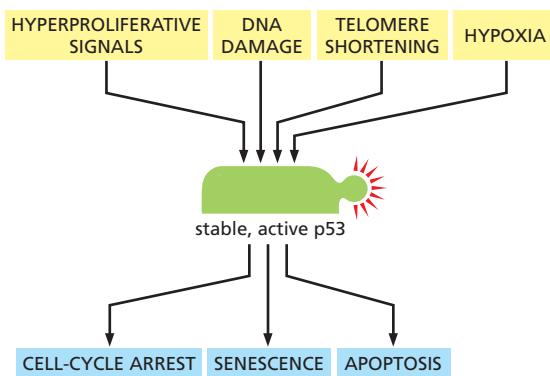


Figure 20–27 Modes of action of the p53 tumor suppressor. The p53 protein is a cellular stress sensor. In response to hyperproliferative signals, DNA damage, hypoxia, telomere shortening, and various other stresses, the p53 levels in the cell rise. As indicated, this may either arrest cell cycling in a way that allows the cell to adjust and survive, trigger cell suicide by apoptosis, or cause cell “senescence”—an irreversible cell-cycle arrest that stops damaged cells from dividing.

the cell is phagocytosed by its neighbors and its contents are efficiently recycled. A bad death is a death by necrosis. In necrosis, the cell bursts or disintegrates and its contents are spilled into the extracellular space, inducing inflammation.

The p53 pathway, therefore, behaves as a sort of antenna, sensing the presence of a wide range of dangerous conditions, and when any are detected, triggering appropriate action—either a temporary or permanent arrest of cell cycling (senescence), or suicide by apoptosis (Figure 20–27). These responses serve to prevent deranged cells from proliferating. Cancer cells are indeed generally deranged, and their survival and proliferation thus depend on inactivation of the p53 pathway. If the p53 pathway were active in them, they would be halted in their tracks or die (Movie 20.4).

The p53 protein performs its job mainly by acting as a transcription regulator (see Movie 17.8). 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. Because p53 binds to DNA as a tetramer, a single mutant subunit within a tetrameric complex can be enough to block its function. Thus, mutations in *p53* can have a dominant negative effect, causing loss of p53 function even when the cell also contains a wild-type version of the gene. For this reason, in contrast with other tumor suppressor genes such as *Rb*, the development of cancer does not always require that both copies of *p53* be knocked out.

As discussed in Chapter 17, the p53 protein exerts its inhibitory effects on the cell cycle, in part at least, by inducing the transcription of *p21*, which encodes a protein that binds to and inhibits the cyclin-dependent kinase (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 progressing through S phase and replicating its DNA.

The mechanism by which p53 induces apoptosis includes stimulation of the expression of many pro-apoptotic genes, and it will be described in Chapter 18.

Genome Instability Takes Different Forms in Different Cancers

If the p53 pathway is functional, a cell with unrepaired DNA damage will stop dividing or die; it cannot proliferate. Mutations in the p53 pathway are, therefore, generally present in cancer cells showing genome instability—which is to say, the majority. But how does this genome instability originate? Here too, cancer genome studies are illuminating.

In ovarian cancers, for example, chromosome breaks, translocations, and deletions are very common, and these aberrations correlate with a high frequency of mutations and epigenetic silencing in the genes needed for repair of DNA double-strand breaks by homologous recombination, especially *Brca1* and *Brca2* (see pp. 281–282). In a subset of colorectal cancers with DNA mismatch repair defects, on the other hand, one instead finds many point mutations scattered throughout the genome (see pp. 250–251). In both kinds of cancer, the genome is commonly destabilized, but different types of mutations can bring this about.

Cancers of Specialized Tissues Use Many Different Routes to Target the Common Core Pathways of Cancer

Mutations in core components of the machinery that regulates cell growth, division, and survival, such as Rb, Ras, PTEN, or p53, are not the only way to pervert the control of these processes. Specialized tissues depend on a variety of pathways, as discussed in Chapter 15, to relay environmental signals to the core control machinery, and each pathway lays the cells open to subversion in a different set of ways. Thus, in different cancers, we can find examples of driver mutations in practically all the major signaling pathways through which cells communicate during development and tissue maintenance (discussed in Chapters 21 and 22).

In glioblastoma, for example, most patients have mutations in one or other of a set of cell-surface receptor tyrosine kinases, especially the EGF receptor mentioned earlier (linking into the Ras/PI3K pathway), suggesting that the cells from which the cancer originates are normally controlled by this route. The cells of the prostate gland, on the other hand, respond to the androgen hormone testosterone, and in prostate cancer, components of the androgen receptor signaling pathway (a variety of nuclear hormone receptor signaling; see Chapter 15) are often mutated. In the normal gut lining, Wnt signaling is critical, and Wnt pathway mutations are present in most colorectal cancers. Pancreatic cancers generally have mutations in the transforming growth factor- β (TGF β) signaling pathway. Activating mutations in the Notch pathway are present in more than 50% of T cell acute lymphocytic leukemias, and so on.

Cells are generally regulated by several different types of external signals that must act in combination, representing a “fail-safe” control mechanism that protects the organism as a whole from cancer. These signals are different in different tissues. As expected, therefore, the corresponding cancers often have mutations in several signaling pathways concurrently. This is true of the examples we have just listed, which commonly have mutations in other signaling pathways in addition to the ones that we have singled out.

Studies Using Mice Help to Define the Functions of Cancer-Critical Genes

The ultimate test of a gene’s role in cancer has to come from investigations in the intact, mature organism. The most favored organism for such studies, apart from humans themselves, is the mouse. To explore the function of a candidate oncogene or tumor suppressor gene, one can make a transgenic mouse that overexpresses it or a knockout mouse that lacks it. Using the techniques described in Chapter 8, one can engineer mice in which the misexpression or deletion of the gene is restricted to a specific set of cells, or in which expression of the gene can be switched on at will at a chosen point in time, or both, to see whether and how tumors develop. Moreover, to follow the growth of tumors from day to day in the living organism, the cells of interest can be genetically marked and made visible by expression of a fluorescent or luminescent reporter (Figure 20–28). In these ways, one can begin to clarify the part that each cancer-critical gene plays in cancer initiation or progression.

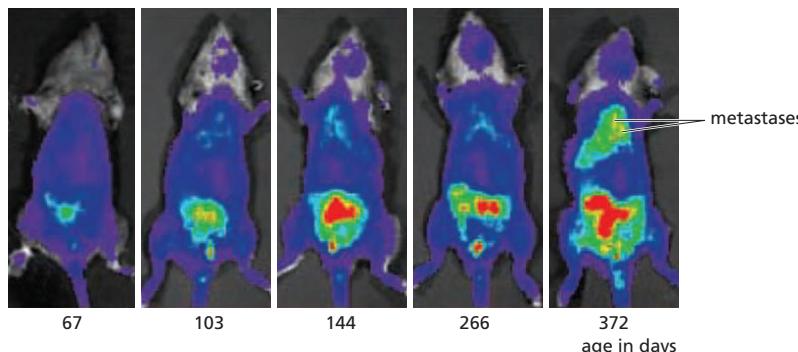
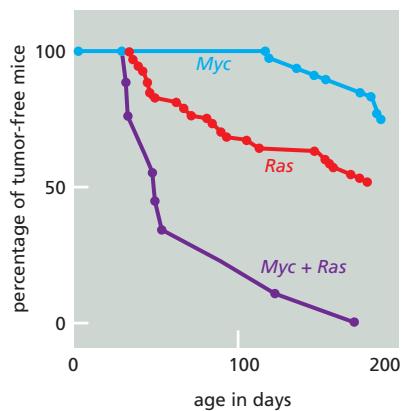


Figure 20–28 Monitoring tumor growth and metastasis in a mouse with a luminescent reporter. A mouse was genetically engineered in a way that allows both copies of its *PTEN* tumor suppressor gene to be inactivated in the prostate gland, simultaneously with the prostate-specific activation of a gene engineered to produce the enzyme luciferase (derived from fireflies). After an injection of luciferin (the substrate molecule for luciferase) into the mouse’s bloodstream, the cells in the prostate emit light and can be detected by their bioluminescence in a live mouse, as seen in the 67-day-old animal at the left. Cells lacking the PTEN phosphatase enzyme contain elevated amounts of the Akt activator, PI(3,4,5)P₃, and this causes the prostate cells to proliferate abnormally, progressing over time to form a cancer. In this way, the process of metastasis could be followed in the same animal over the course of a year. The light intensity in these experiments is proportional to the number of prostate-cell descendants, increasing from light blue to green, to yellow, to red in this representation. (Adapted from C.-P. Liao et al., *Cancer Res.* 67:7525–7533, 2007.)

Figure 20–29 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.)



Transgenic mouse studies confirm, for example, that a single oncogene is generally not enough to turn a normal cell into a cancer cell. Thus, in mice engineered to express a *Myc* or *Ras* oncogenic transgene, 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 oncogene, 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 20–29); 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. This strongly suggests that multiple mutations are required for tumorigenesis, as supported by a great deal of other evidence discussed earlier. Experiments using mice with deletions of tumor suppressor genes lead to similar conclusions.

Cancers Become More and More Heterogeneous as They Progress

From simple histology, looking at stained tissue sections, it is clear that some tumors contain distinct sectors, all clearly cancerous, but differing in appearance because they differ genetically: the cancer cell population is heterogeneous. Evidently, within the initial clone of cancerous cells, additional mutations have arisen and thrived, creating diverse subclones. Today, the ability to analyze cancer genomes lets us look much deeper into the process.

One approach involves taking samples from different regions of a primary tumor and from the metastases that it has spawned. With modern methods, it is even possible to take representative single cells and analyze their genomes. Such studies reveal a classic picture of Darwinian evolution, occurring on a time scale of months or years rather than millions of years, but governed by the same rules of natural selection (Figure 20–30).

One such investigation compared the genomes of 100 individual cells from different regions of a primary tumor of the breast. A large fraction—just over half—of the chosen cells was genetically normal or nearly so: these were connective-tissue cells and other cell types, such as those of the immune system, that were mixed up with the cancer cells. The cancer cells themselves were distinguished by their severely disrupted genomes. The detailed pattern of gene deletions and amplifications in each such cell revealed how closely it was related to the others, and from this data one could draw up a family tree (Figure 20–30B). In this case, three main branches of the tree were seen; that is, the cancer consisted of three major

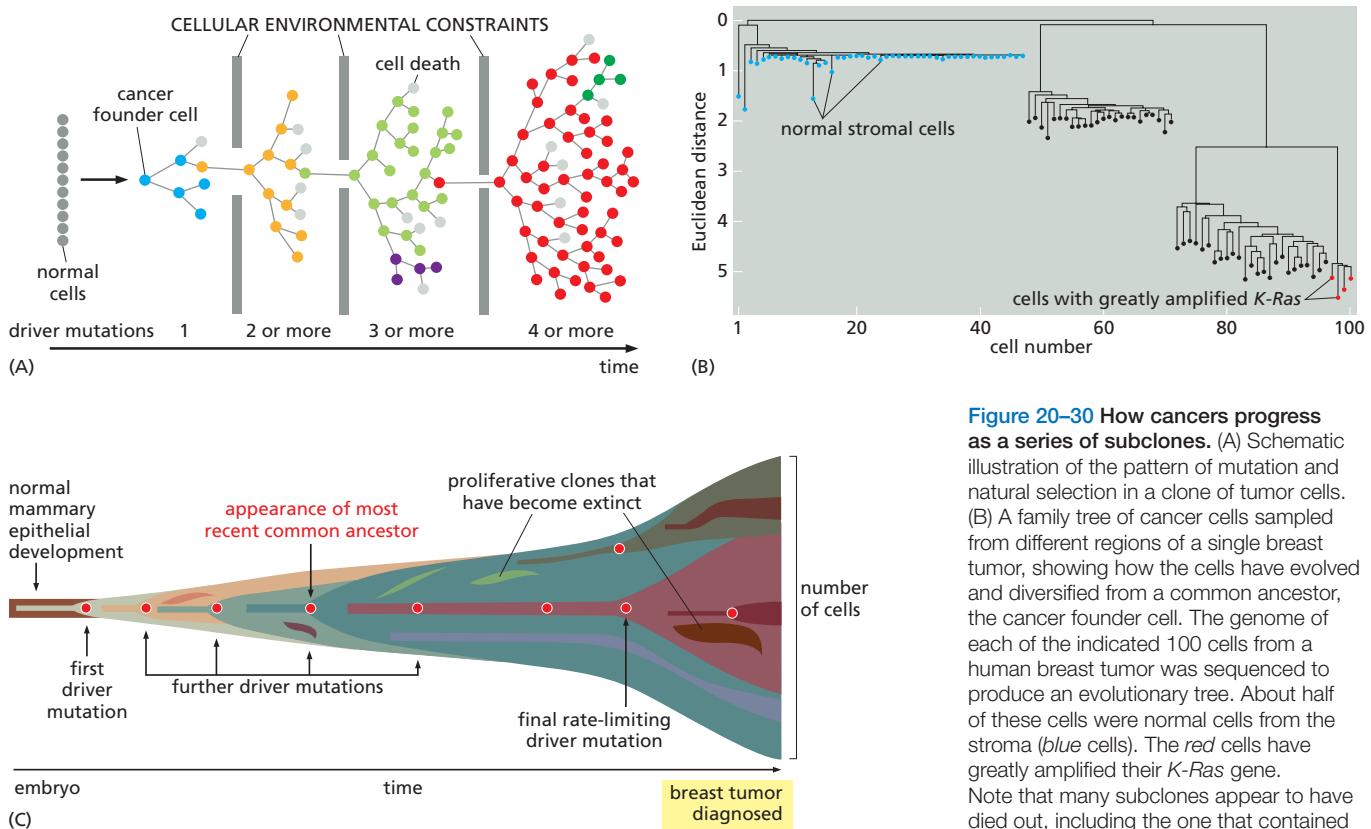


Figure 20–30 How cancers progress as a series of subclones. (A) Schematic illustration of the pattern of mutation and natural selection in a clone of tumor cells. (B) A family tree of cancer cells sampled from different regions of a single breast tumor, showing how the cells have evolved and diversified from a common ancestor, the cancer founder cell. The genome of each of the indicated 100 cells from a human breast tumor was sequenced to produce an evolutionary tree. About half of these cells were normal cells from the stroma (blue cells). The red cells have greatly amplified their *K-Ras* gene. Note that many subclones appear to have died out, including the one that contained the founder cells for the three subclones that survive.

(C) A depiction of how driver mutations are thought to cause cancer progression over long periods of time, before producing a large enough clone of proliferating cells to be detected as a tumor. The data indicate that driver mutations occur only rarely in a background of long-lived subclones of cells that continually accumulate passenger mutations without gaining a growth advantage. (A, adapted from M. Greaves, *Semin. Cancer Biol.* 20:65–70, 2010; B, adapted from N. Navin et al., *Nature* 472:90–94, 2011; C, adapted from S. Nik-Zainal et al., *Cell* 149:994–1007, 2012.)

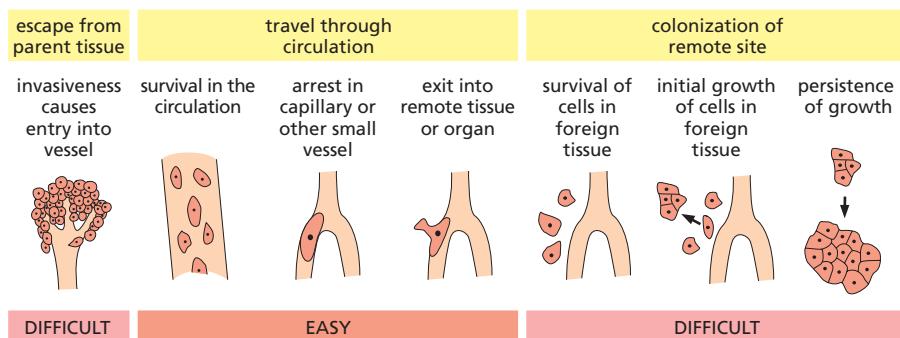
subclones. From the shared abnormalities, one could deduce that their last common ancestor—the presumed founder of the cancer—was already very different from a normal cell, but that the first split between branches occurred early, when the tumor was small. This was followed by a large amount of additional change within each branch. A hint of the future could be seen in the smallest of the three major subclones: its cells were distinguished by a massive amplification of a *Ras* oncogene. Given more time, perhaps they would have out-competed the other cancer cells and taken over the whole tumor.

Similar results have been obtained with other cancers. Clearly, cancer cells are constantly mutating, multiplying, competing, evolving, and diversifying as they exploit new ecological niches and react to the treatments that are used against them (Figure 20–30C). Diversification accelerates as they metastasize and colonize new territories, where they encounter new selection pressures. The longer the evolutionary process continues, the harder it becomes to catch them all in the same net and kill them.

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. For a start, it is not clear exactly what new properties a cancer cell must acquire to become metastatic. In some cases, it is possible that invasion and metastasis require no further genetic changes beyond those needed to violate the normal controls on cell growth, cell division, and cell death. On the other hand, it may be that, for some cancers, metastasis requires a large number of additional mutations and epigenetic changes. Clues are coming from comparisons of the genomes of cells of primary tumors with the cells of metastases that they have spawned. The results appear complex and variable from one cancer to another. Nevertheless, some general principles have emerged.

As we discussed earlier, it is helpful to distinguish three phases of tumor progression required for a carcinoma to metastasize (see Figure 20–16). First, the cells



must escape the normal confines of their parent epithelium and begin to invade the tissue immediately beneath. Second, they must travel via the blood or lymph to lodge in distant sites. Third, they must survive there and multiply. It is the first and last steps in this sequence that are the most difficult to accomplish for most cancers (**Figure 20-31**).

The first step, local invasiveness, requires a relaxation of the mechanisms that normally hold epithelial cells together. As mentioned earlier, this step resembles the normal developmental process known as the *epithelial-mesenchymal transition (EMT)*, in which epithelial cells undergo a shift in character, becoming less adhesive and more migratory (discussed in Chapter 19). A key part of the EMT process involves switching off expression of the *E-cadherin* gene. The primary function of the transmembrane *E-cadherin* protein is in cell-cell adhesion, binding epithelial cells together through adherens junctions (see Figure 19-13). In some carcinomas of the stomach and of the breast, *E-cadherin* has been identified as a tumor suppressor gene, and a loss of *E-cadherin* 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 as they grow larger and become hypoxic in their interior. This process, called *angiogenesis*, is caused by the secretion of angiogenic factors that promote the growth of blood vessels, such as vascular endothelial growth factor (VEGF; see Figure 22-26). An abnormal fragility and leakiness of the new vessels that form 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 (**Movie 20.5**). 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. Very few establish full-blown metastases.

What, if anything, distinguishes the survivors from the failures? A clue may come from the fact that in many types of tumors, the cancer cells show a kind of heterogeneity that resembles the heterogeneity seen among the cells of those normal tissues that renew themselves continually by a stem-cell strategy, as we discuss next.

A Small Population of Cancer Stem Cells May Maintain Many Tumors

Self-renewing tissues, where cell division continues throughout life, are the breeding ground for the great majority of human cancers. They include the epidermis

Figure 20-31 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 20-16, 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 cells from highly metastatic tumors are observed to have much greater success than cells from a nonmetastatic source. 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.)

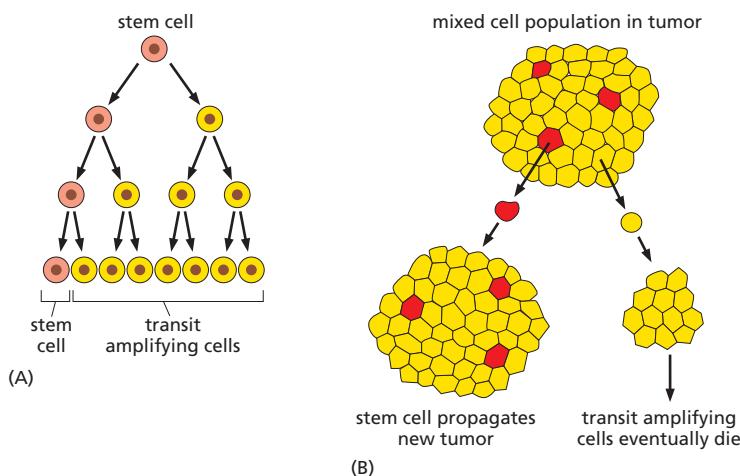


Figure 20–32 Cancer stem cells can be responsible for tumor growth and yet remain only a small part of the tumor-cell population. (A) How stem cells produce transit amplifying cells. (B) How a small proportion of cancer stem cells can maintain a tumor. Suppose, for example, that each daughter of a cancer stem cell has a probability slightly greater than 50% of retaining stem-cell potential and a probability slightly less than 50% of becoming a transit amplifying cell that is committed to a program of cell divisions that stops after 10 division cycles. While the number of cancer stem cells will increase slowly but steadily to give a growing tumor, the non-stem cells that they give rise to will always outnumber the stem cells by a large factor—in this example, by a factor of about 1000. (If the cell-division-cycle and survival times for the two classes of cells are equal.)

(the outer epithelial layer of the skin), the lining of the digestive and reproductive tracts, and the bone marrow, where blood cells are generated (see Chapter 22). In almost all these tissues, renewal depends on the presence of stem cells, which divide to give rise to terminally differentiated cells, which do not divide. This creates a mixture of cells that are genetically identical and closely related by lineage, but are in different states of differentiation. Many tumors seem likewise to consist of cells in varied states of differentiation, with different capacities for cell division and self-renewal.

To see the implications, it is helpful to consider how normal stem-cell systems operate. When a normal stem cell divides, each daughter cell has a choice—it can remain a stem cell, or it can commit to a pathway leading to differentiation. A stem-cell daughter remains in place to generate more cells in the future. A committed daughter typically undergoes some rounds of cell proliferation (as a so-called *transit amplifying cell*) but then stops dividing, terminally differentiates, and eventually is discarded and replaced (it may die by apoptosis, with recycling of its materials, or be shed from the body). On average, the two fates—stem cell or differentiating cell—normally occur with equal probability, so that half the daughters of stem-cell divisions take the one path and half take the other. In a healthy body, feedback controls regulate the process, adjusting this balance of cell-fate choices to correct for any departure from the proper cell population numbers. Thus, the number of stem cells remains approximately constant, and the terminally differentiated cells are continually replaced at a steady rate. Because of the divisions undergone by the transit amplifying cells, the stem cells may be vastly outnumbered by the cells that are committed to terminal differentiation and have lost the capacity for self-renewal. But the stem cells, though few and far between and often relatively slowly dividing, carry the whole responsibility for maintenance of the tissue in the long term.

Some cancers seem to be organized in a similar way: they consist of rare **cancer stem cells** capable of dividing indefinitely, together with much larger numbers of dividing transit amplifying cells that are derived from the cancer stem cells but have a limited capacity for self-renewal (Figure 20–32). These non-stem cells appear to constitute the great majority of the cell population in some tumors.

The Cancer Stem-Cell Phenomenon Adds to the Difficulty of Curing Cancer

Evidence for the cancer stem-cell phenomenon comes chiefly from experiments in which individual cells from a cancer are tested for their ability to give rise to fresh tumors: a standard assay is to implant the cells into an immunodeficient mouse (Figure 20–33). It has been known for half a century that there is usually only a small chance—typically much less than 1%—that a tumor cell chosen at random and tested in this way will generate a new tumor. This by itself does not prove that

the tumor cells are heterogeneous: like seeds scattered on difficult ground, each of them may have only a small chance of finding a spot where it can survive and grow. Modern technologies for sorting cells have shown, however, that in some cancers at least, the rate of success in founding new tumors is even lower than it would otherwise be because the cancer cells are heterogeneous in their state of differentiation, and only a small subset of them—the cancer stem cells—have the special properties needed for tumor propagation. For example, in several types of cancer, including breast cancers and leukemias, one can fractionate the tumor cells using monoclonal antibodies that recognize a particular cell-surface marker that is present on the normal stem cells in the tissue of origin of the cancer. The purified cancer cells expressing this marker are found to have a greatly enhanced ability to found new tumors. And the new tumors consist of mixtures of cells that express the marker and cells that do not, all generated from the same founder cell that expressed the marker.

Experiments with breast cancer cells have revealed that, instead of following a rigid program from stem cell to transit amplifying cell to terminally differentiated cell, these cancer cells can randomly switch to and fro—with a certain low transition probability—between different states of differentiation that express different molecular markers. In one state, they behave like stem cells, dividing slowly but capable of founding new tumors; in other states, they behave like transit amplifying cells, dividing rapidly but unable to found new tumors in a standard transplant assay. But a single cell in any of these states—given time in culture, or a congenial environment in the body—will give rise to a mixed population that includes all the other states as well.

The cancer stem-cell phenomenon, whatever its basis, implies that even when the tumor cells are genetically similar, they are phenotypically diverse. A treatment that wipes out those in one state is likely to allow survival of others that remain a danger. Radiotherapy or a cytotoxic drug, for example, may selectively kill off the rapidly dividing cells, reducing the tumor volume to almost nothing, and yet spare a few slowly dividing cells that go on to resurrect the disease. This greatly adds to the difficulty of cancer therapy, and it is part of the reason why treatments that seem at first to succeed often end in relapse and disappointment.

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 and epigenetic changes. We have discussed what some of these 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. We take **colorectal cancer** as our example.

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 22 (pp. 1217–1221). For both the small and large intestine, the epithelium is 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 22. Mutations that disrupt these signals begin the process of tumor progression for most colorectal cancers ([Movie 20.6](#)).

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 (a fiber



Figure 20–33 An immunodeficient mouse, as used in transplantation assays to test human cancer cells for their ability to found new tumors. This nude mouse has a mutation that blocks development of the thymus and, as a side effect, robs it of hair. Because it has practically no T cells, it tolerates grafts of cells even from other species. (Courtesy of Harlan Sprague Dawley.)

optic device for viewing the interior of the colon and rectum) often reveals a small benign tumor, or adenoma, of the gut epithelium in the form of a protruding mass of tissue called a *polyp* (see Figure 22–4). 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 much lower: according to some studies, less than a quarter of what it would be otherwise.

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, some tumor cells become invasive in a small fraction of the polyps, first breaking through the epithelial basal lamina, then spreading through the layer of muscle that surrounds the gut, 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), in about 40% of cases; *p53*, in about 60% of cases; and the tumor suppressor gene *Apc* (discussed below), in more than 80% of cases. Others are involved in smaller numbers of colon cancers, and some of these are listed in **Table 20–1**.

The role of *Apc* first came to light through study of certain families showing a rare type of hereditary predisposition to colorectal cancer, called *familial*

TABLE 20–1 Some Genetic Abnormalities Detected in Colorectal Cancer Cells

Gene	Class	Pathway affected	Human colon cancers (%)
<i>K-Ras</i>	Oncogene	Receptor tyrosine kinase signaling	40
<i>β-Catenin</i> ¹	Oncogene	Wnt signaling	5–10
<i>Apc</i> ¹	Tumor suppressor	Wnt signaling	>80
<i>p53</i>	Tumor suppressor	Response to stress and DNA damage	60
<i>TGFβ receptor II</i> ²	Tumor suppressor	<i>TGFβ</i> signaling	10
<i>Smad4</i> ²	Tumor suppressor	<i>TGFβ</i> signaling	30
<i>MLH1</i> and other DNA mismatch repair genes (often silenced by DNA methylation)	Tumor suppressor (genetic stability)	DNA mismatch repair	15

^{1,2}The genes with the same superscript numeral act in the same pathway, and therefore only one of the components is mutated in an individual cancer.

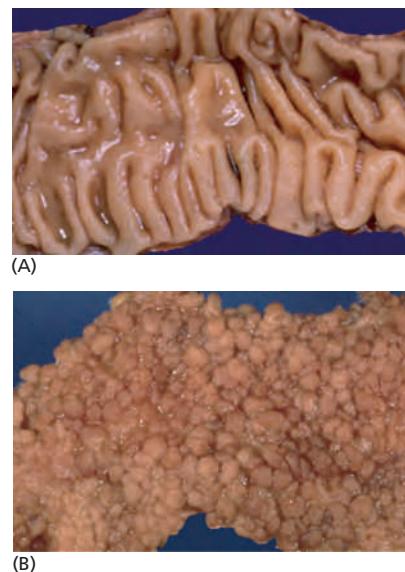
Figure 20–34 Colon of familial adenomatous polyposis coli patient compared with normal colon. (A) The normal colon wall is a gently undulating but smooth surface. (B) The polyposis colon is completely covered by hundreds of projecting polyps, each resembling a tiny cauliflower when viewed with the naked eye. (Courtesy of Andrew Wyllie and Mark Arends.)

adenomatous polyposis coli (FAP). In this syndrome, hundreds or thousands of polyps develop along the length of the colon (Figure 20–34). 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. Nevertheless, in more than 80% of the cases, their cancer cells (but not their normal cells) have inactivated both copies of the *Apc* gene through mutations acquired during the patient's lifetime. Thus, by a route similar to that which we discussed for retinoblastoma, mutation of the *Apc* gene was identified as one of the central ingredients of colorectal cancer.

The *Apc* protein, as we now know, is 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 the protein's degradation. By inhibiting β -catenin in this way, *Apc* prevents the β -catenin from migrating to the nucleus, where it would act as a transcriptional regulator to drive cell proliferation and maintain the stem-cell state (see Figure 15–60). Loss of *Apc* results in an excess of free β -catenin and thus leads to an uncontrolled expansion of the stem-cell population. This causes massive increase in the number and size of the intestinal crypts (see Figure 22–4).

When the β -catenin gene was sequenced in a collection of colorectal tumors, it was discovered that, many of the tumors that did not have *Apc* mutations 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 being so, why is the *Apc* gene in particular so often the most common culprit in colorectal cancer? The *Apc* protein is large and it interacts not only with 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.



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 more common 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 20–35).

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. These genes are homologous in structure and function to the *MutL* and *MutS* genes in bacteria and yeast (see Figure 5–19). Only one of the two copies of the involved gene is defective, so the repair system is still able to remove

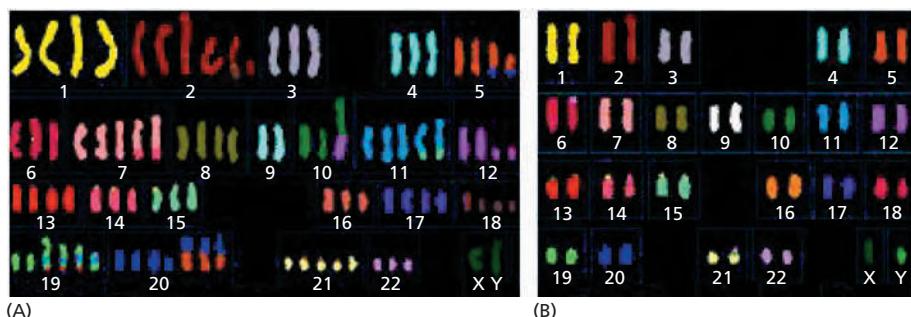


Figure 20-35 Chromosome complements (karyotypes) of colon cancers showing different kinds of genetic instability. (A) The karyotype of a typical cancer shows many gross abnormalities in chromosome number and structure. Considerable variation can also exist from cell to cell (not shown). (B) The karyotype of a tumor that has a stable chromosome complement with few chromosomal anomalies; the genetic abnormalities are mostly invisible, having been created by defects in DNA mismatch repair. All of the chromosomes in this figure were stained as in Figure 4-10, the DNA of each human chromosome being marked with a different combination of fluorescent dyes. (Courtesy Wael Abdel-Rahman and Paul Edwards.)

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 then 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, the epigenetic silencing or mutation of mismatch repair genes was 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 by-product of malignant behavior but a contributory cause—and that cancer cells can acquire this instability in multiple ways.

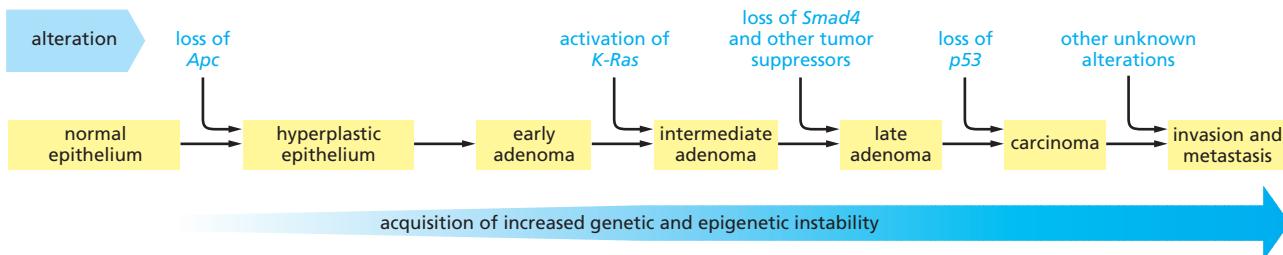
The Steps of Tumor Progression Can Often Be Correlated with Specific Mutations

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. 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.

Inactivating mutations in *p53* are thought to come later still, as they are rare in polyps but common in carcinomas (Figure 20-36). We have seen that loss of *p53* function allows cancer cells to endure stress and to avoid apoptosis and cell-cycle arrest. Additionally, loss of *p53* is related to the heightened activation



of oncogenes such as *Ras*. Experiments in mice show that an initial low level of oncogene activation can give rise to a slowly growing tumor even while *p53* is functional: genes such as *Ras* are, after all, part of the normal machinery of growth control, and moderate activation is not stressful for a cell and does not call the *p53* protein into play. Progression of a tumor from slow to rapid, malignant growth, however, involves activation of oncogenes beyond normal physiological limits to a higher, stressful level. If the *p53* protein is present and functional, this should lead to cell-cycle arrest or death. Only by losing *p53* function can the cancer cells with hyperactive oncogenes survive and progress.

The steps we have just described are only part of the picture. It is important to emphasize that each case of colorectal cancer is different, with its own detailed combination of mutations, and that even for the mutations that are commonly shared, the sequence of occurrence may vary. The same is true for cancers in general.

Advances in molecular biology have recently provided 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 in the next section, such information promises to become as important for the diagnosis and treatment of cancer as was the breakthrough of being able to identify microorganisms for the treatment of infectious diseases.

Summary

The molecular analysis of cancer cells reveals two classes of cancer-critical genes: oncogenes and tumor suppressor genes. A set of these genes becomes altered by a combination of genetic and epigenetic accidents to drive tumor progression. Many cancer-critical genes code for components of the social control pathways that regulate when cells grow, divide, differentiate, or die. In addition, a subclass of tumor suppressors can be categorized as “genome maintenance genes,” because their normal role is to help maintain genome integrity.

*The inactivation of the *p53* pathway, which occurs in nearly all human cancers, 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 for protecting us against cancer.*

The sequencing of cancer cell genomes reveals that—except for the cancers of childhood—many cancers acquire 10 or so driver mutations over the long course of tumor progression, along with a considerably larger number of passenger mutations of no consequence. The same methods reveal how subclones of cells arise and die out as a tumor ages. Tumors thus contain a heterogeneous mixture of cells, some—the so-called cancer stem cells—being much more dangerous than others.

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 a good 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 arise. Nevertheless, many of the same changes are encountered repeatedly, suggesting that there are a limited number of ways to breach our defenses against cancer.

Figure 20–36 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 are related. But many other mutations are generally involved, and different colon cancers can progress through different sequences of mutations (and/or epigenetic changes).

CANCER PREVENTION AND TREATMENT: PRESENT AND FUTURE

We can apply the growing understanding of the molecular biology of cancer to sharpen our attack on the disease at three levels: prevention, diagnosis, and treatment. Prevention is always better than cure, and indeed many cancers can be prevented, especially by avoiding smoking. Highly sensitive molecular assays promise new opportunities for earlier and more precise diagnosis, with the aim of detecting primary tumors while they are still small and have not yet metastasized. Cancers caught at these early stages can often be nipped in the bud by surgery or radiotherapy, as we saw for colorectal polyps. Nevertheless, full-blown malignant disease will continue to be common for many years to come, and cancer treatments will continue to be needed.

In this section, we first examine the preventable causes of cancer and then consider how advances in our understanding at a molecular level are beginning to transform the treatment of the disease.

Epidemiology Reveals That Many Cases of Cancer Are Preventable

A certain irreducible background incidence of cancer is to be expected regardless of circumstances. As discussed in Chapter 5, mutations can never be absolutely avoided because they are an inescapable consequence of fundamental limitations on the accuracy of DNA replication and repair. If a person 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.

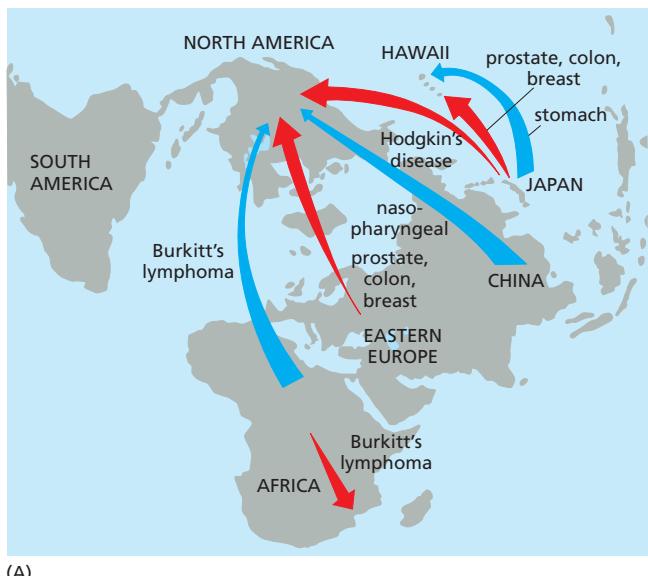
Nevertheless, environmental factors seem to play a large part in determining the risk for cancer. This is demonstrated most clearly by a comparison of cancer incidence in different countries: for almost every cancer that is common in one country, there is another country where the incidence is much lower. Because migrant populations tend to adopt the pattern of cancer incidence typical of their new host country, the differences are thought to be due mostly to environmental, not genetic, factors. From such findings, it has been suggested that 80–90% of cancers should be avoidable, or at least postponable (**Figure 20–37**).

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, approximately one in five people will die of cancer. But the incidence of cancer among strict Mormons in Utah—who avoid alcohol, coffee, cigarettes, drugs, and casual sex—is only about half the incidence for non-practicing members of the same family or for 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 tobacco as a striking exception—to pinpoint the specific environmental factors responsible for these large population differences or to establish how they act. Nevertheless, several important classes of environmental cancer risk factors have been identified (Figure 20–37B). One thinks first of mutagens. But there are also many other influences—including the amount of food we eat, the hormones that circulate in our bodies, and the irritations, infections, and damage to which we expose our tissues—that are no less important and favor development of the disease in other ways.

Sensitive Assays Can Detect Those Cancer-Causing Agents that Damage DNA

Many quite disparate chemicals are carcinogenic when they are fed to experimental animals or painted repeatedly on their skin. Examples include a range



cause	cancers caused (percent of total)	number of deaths in US (annual)	magnitude of reduction possible (percent)
smoking	33	189,000	75
diet, overweight, and obesity	25	143,000	50
lack of exercise	5	28,600	85
viruses	5	28,600	100
alcohol	3	17,200	50
UV and ionizing radiation	2	11,400	50
occupational carcinogens	5	28,600	50

(B)

Figure 20–37 Cancer incidence is related to environmental influences.

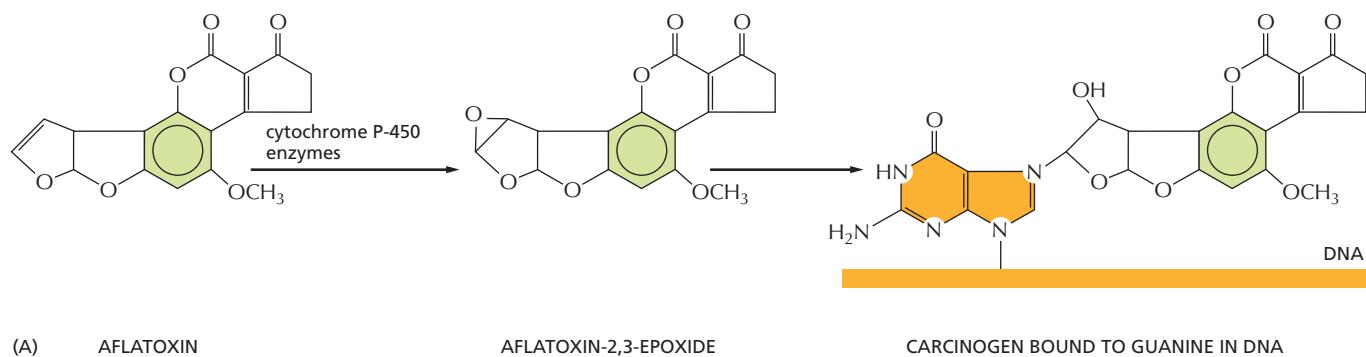
(A) This map of the world shows the rates of cancer increasing (red arrows) or decreasing (blue arrows) when specific populations move from one location to another. Such observations suggest the importance of environmental factors, including diet, in dictating cancer risk. (B) Some estimated effects of environment and lifestyle on cancer in the United States (US). The table shows both the yearly deaths in the US attributable to each cancer and the estimated percentage of that cancer that could be eliminated through prevention. (B, data from G.A. Colditz, K.Y. Wolin and S. Gehlert, *Sci. Transl. Med.* 4:127rv4, 2012.)

of aromatic hydrocarbons and derivatives of them such as aromatic amines, nitrosamines, and alkylating agents such as mustard gas. Although these **chemical carcinogens** are diverse in structure, a large proportion of them have at least one shared property—they cause mutations. In one common test for mutagenicity (the *Ames test*), the carcinogen is mixed with an activating extract prepared from rat liver cells (to mimic the biochemical processing that occurs in an intact animal). The mixture is then added to a culture of specially designed test bacteria and the bacterial mutation rate measured. 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.

A few of these carcinogens act directly on DNA. But generally the more potent ones are relatively inert chemically; these chemicals become damaging only after they have been converted to a more reactive molecule 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. Unhappily, their activity on certain chemicals generates products that are highly mutagenic. Examples of carcinogens activated in this way include *benzo[a]pyrene*, a cancer-causing chemical present in coal tar and tobacco smoke and the fungal toxin *aflatoxin B1* (Figure 20–38).

Fifty Percent of Cancers Could Be Prevented by Changes in Lifestyle

Tobacco smoke is the most important carcinogen in the world today. Even though many other chemical carcinogens have been identified, none of these appear to be responsible for anything like the same numbers of human cancer deaths. 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, age-adjusted death rates for most common human cancers have stayed much the same over the past half-century, or, in some cases, have declined significantly (Figure 20–39). Survival rates, moreover, have improved. Thirty years ago, less than 50% of patients lived more than five years from the time of diagnosis; now, more than two-thirds do so.

**Figure 20–38** 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. *Aflatoxin* is an important cause of liver cancer in the tropics. (B, data from Cancer and the Environment: Gene Environment Interactions, National Academies Press, 2002.)

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

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 20–38). 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.

Except for tobacco, chemical toxins and mutagens are of lesser importance as contributory causes of cancer than other factors that are more a matter of personal choice. One important factor is the quantity of food we eat: as mentioned earlier, the risk of cancer is greatly increased in people who are obese. In fact, it is estimated that as many as 50% of all cancers could be avoided by simple, identifiable changes in lifestyle (see Figure 20–37B).

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

Cancer in humans is not an infectious disease, and most human cancers do not have any infectious cause. However, 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. Evidence for their involvement comes partly from the detection of viruses in cancer patients and partly from epidemiology. Thus, cancer of the uterine cervix is associated with infection with a papillomavirus, while liver cancer is very common in parts of the world (Africa and Southeast Asia) where hepatitis-B viral infections are common. Chronic infection

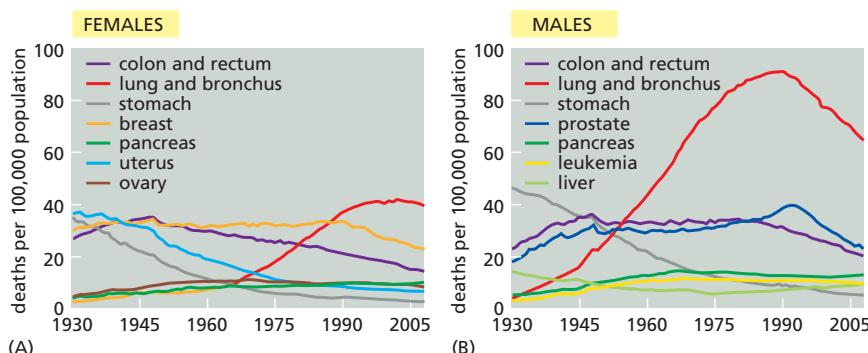


Figure 20–39 Age-adjusted cancer death rates, United States, 1930–2008. Selected death rates, adjusted to the age distribution of the US population, are plotted for (A) females and (B) males. Note the dramatic rise in lung cancer for both sexes, following the pattern of tobacco smoking, and the fall in deaths from stomach cancer, thought to be related to a fall in rates of infection with *Helicobacter pylori*. 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 as people live longer, on average. (Adapted from Cancer Facts and Figures, 2012. Data from U.S. Mortality Volumes 1930 to 1959, U.S. Mortality Data 1960 to 2008, National Center for Health Statistics, Centers for Disease Control and Prevention. © 2012, American Cancer Society, Inc., Surveillance Research.)

TABLE 20-2 Viruses Associated with Human Cancers

Virus	Associated cancer	Areas of high incidence
DNA viruses		
<i>Papovavirus family</i>		
Papillomavirus (many distinct strains)	Warts (benign)	Worldwide
	Carcinoma of the uterine cervix	Worldwide
<i>Hepadnavirus family</i>		
Hepatitis-B virus	Liver cancer (hepatocellular carcinoma)	Southeast Asia, tropical Africa
<i>Herpesvirus family</i>		
Epstein–Barr virus	Burkitt's lymphoma (cancer of B lymphocytes)	West Africa, Papua New Guinea
	Nasopharyngeal carcinoma	Southern China, Greenland
Human herpesvirus 8	Kaposi's sarcoma	Central and Southern Africa
RNA viruses		
<i>Retrovirus family</i>		
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 (via human herpesvirus 8)	Central and Southern Africa
<i>Flavivirus family</i>		
Hepatitis-C virus	Liver cancer (hepatocellular carcinoma)	Worldwide
For all these viruses, the number of people infected is much larger than the number who develop cancer: the viruses must act in conjunction with other factors. As described in the text, different viruses contribute to cancer in different ways.		

with hepatitis-C virus, which has infected 170 million people worldwide, is also clearly associated with the development of liver cancer.

The main culprits, as shown in **Table 20-2**, are the DNA viruses. The **DNA tumor viruses** cause cancer by the most direct route—by interfering with controls of the cell cycle and apoptosis. To understand this type of viral carcinogenesis, it is important to review the life history of viruses. Many DNA viruses use the host cell's DNA replication machinery to replicate their own genomes. However, to produce a large number of 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 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 (either integrated into the host genome, or as an extrachromosomal plasmid) in the course of ordinary cell-division cycles. These viruses will switch between two modes of existence according to circumstances, 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.

Neither of these conditions converts the host cell to a cancerous character, nor is it in the interest of the virus to do so. 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. As described in the example below, this type of accident can switch on the persistent proliferation of the host cell itself, leading to cancer.

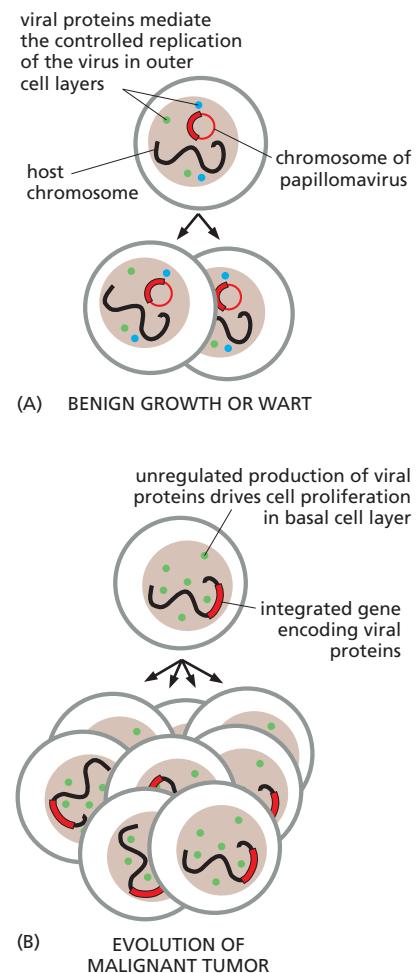
Cancers of the Uterine Cervix Can Be Prevented by Vaccination Against Human Papillomavirus

The **papillomaviruses** are a prime example of DNA tumor viruses. They are responsible for human warts and are especially important as a cause of carcinoma of the uterine cervix: this is the second commonest cancer of women in the world as a whole, representing about 6% of all human cancers. Human papillomaviruses (HPV) 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 (see Figure 22–10). This can lead to cancer, with the viral genes acting as oncogenes (**Figure 20–40**).

The whole process, from initial infection to invasive cancer, is slow, taking many years. It involves a long intermediate stage when the affected patch of cervical epithelium is visibly disordered but the cells have not yet begun to invade the underlying connective tissue—a phenomenon called *intraepithelial neoplasia*. Many such lesions regress spontaneously. Moreover, 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 scraping off a sample of cells from the surface of the cervix and viewing it under the microscope (the “Pap smear” technique).

Better still, a vaccine has now been developed that protects against infection with the relevant strains of human papillomavirus. This vaccine, given to girls before puberty and thus before they become sexually active, has been shown to greatly reduce their risk of ever developing cervical cancer. Because the virus spreads through sexual activity, it is now recommended that both young males and young females be routinely vaccinated. Mass immunization programs have begun in several countries.

Figure 20–40 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 (red circles), whose replication is regulated so as to keep step with the chromosomes of the host. (A) Normally, the virus perturbs the host cell cycle only when the virus is programmed to produce infectious progeny, in the outer layers of an epithelium. This is relatively harmless. (B) 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 of an epithelium. This can disrupt the normal control of viral gene expression. The unregulated production of certain viral proteins (E6 and E7) interferes with the control of cell division in the basal cells, thereby helping to generate a cancer (bottom).



Infectious Agents Can Cause Cancer in a Variety of Ways

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 both 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; the other is p53. Other DNA tumor viruses use similar mechanisms to inhibit Rb and p53, underlining the central importance of inactivating both of these tumor suppressor pathways if a cell is to escape the normal constraints on proliferation.

In other cancers, viruses have indirect tumor-promoting actions. The hepatitis-B and C viruses, for example, favor the development of liver cancer by causing chronic inflammation (hepatitis), which stimulates an extensive cell division in the liver that promotes the eventual evolution of tumor cells. 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 herpesvirus (HHV-8) that has a direct carcinogenic action. By causing severe inflammation, 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; dramatic falls in the incidence of stomach cancer over the last half-century (see Figure 20–39) correlate with a decline in the incidence of *Helicobacter* infections.

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. Moreover, unlike normal cells, cancer cells can 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 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 peculiarity of cancer cells that distinguishes them from normal cells. One such property is genetic instability, reflecting deficiencies in chromosome maintenance, cell-cycle checkpoints, and/or DNA repair. Remarkably, the most widely used cancer therapies seem to work by exploiting these abnormalities, although this was not known by the scientists who first developed the treatments. Ionizing radiation and most anticancer drugs damage DNA or interfere with chromosome segregation at mitosis, and they preferentially kill cancer cells because cancer 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, thanks to the cell-cycle checkpoint responses discussed in Chapter 17. Because cancer cells generally have defects in their checkpoint responses, they may continue to divide after irradiation, only to die after a few days because the genetic damage remains unrepaired. More generally, most cancer cells are physiologically deranged to a stressful degree: they live dangerously. Even though the cells

in a tumor have evolved to be unusually tolerant of minor DNA damage, they are hypersensitive to the much greater amount of damage that can be created by radiation and by DNA-damaging drugs. A small increase of genetic damage can be enough to tip the balance between proliferation and death.

Unfortunately, while the molecular defects present in cancer cells often enhance their sensitivity to cytotoxic agents, they can also increase their resistance. For example, where a normal cell might die by apoptosis in response to DNA damage, thanks to the stress response mediated by p53, a cancer cell may escape apoptosis because its p53 is lacking. Cancers vary widely in their sensitivity to cytotoxic treatments, some responding to one drug, some to another, probably reflecting the particular kinds of defects that a particular cancer has in DNA repair, cell-cycle checkpoints, and the control of apoptosis.

New Drugs Can Kill Cancer Cells Selectively by Targeting Specific Mutations

Radiotherapy and traditional cytotoxic drugs are rather weakly selective: they hurt normal cells as well as the cancer cells, and the safety margin is narrow. The dose often cannot be raised high enough to kill all the cancer cells, because this would kill the patient, and curative treatments, where achievable, generally require a combination of several cytotoxic agents. The side effects can be harsh and hard to endure. How can we do better?

An ideal treatment is one that is cell-lethal in combination with some lesion that is present in the cancer cells, but harmless to cells where this lesion is absent. Such a treatment is said to be *synthetic-lethal* (from the original sense of the word *synthesis*, meaning “putting together”): it kills only in partnership with the cancer-specific mutation. As we become increasingly able to pinpoint the specific alterations in cancer cells that make them different from their normal neighbors, new opportunities for such precisely targeted treatments are coming into view. We end this chapter with some examples of new treatments of this type that are already being put into practice.

PARP Inhibitors Kill Cancer Cells That Have Defects in *Brca1* or *Brca2* Genes

As we have emphasized, the genetic instability of cancer cells makes the cells both dangerous and vulnerable—dangerous because of the enhancement in their ability to evolve and proliferate, and vulnerable because treatment that leads to still more extreme genetic disruption can take them over the brink and kill them. In some cancers, genetic instability results from an identified fault in one of the many devices on which normal cells depend for DNA repair and maintenance. In this case, a drug is tailored to block a complementary part of the DNA repair machinery can lead to such severe genetic damage that the cancer cells die.

Detailed studies of the mechanisms for DNA maintenance discussed in Chapter 5 reveal a surprising amount of apparent redundancy. Thus, knocking out a particular pathway for DNA repair is generally less disastrous than one might expect, because alternate repair pathways exist. For example, stalled DNA replication forks can arise when the fork encounters a single-strand break in a template strand, but cells can avoid the disaster that would otherwise result either by directly repairing these single-strand breaks, or, if that fails, repairing the broken fork that results by homologous recombination (see Figure 5–50). Suppose that the cells in a particular cancer have become genetically unstable by acquiring a mutation that reduces their ability to repair broken replication 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 break? The consequences of such drug treatment might be expected to be relatively harmless for normal cells, but lethal for the cancer.

This strategy appears to work to kill the cells in at least one class of cancers—those that have inactivated both copies of either their *Brca1* or their *Brca2* tumor

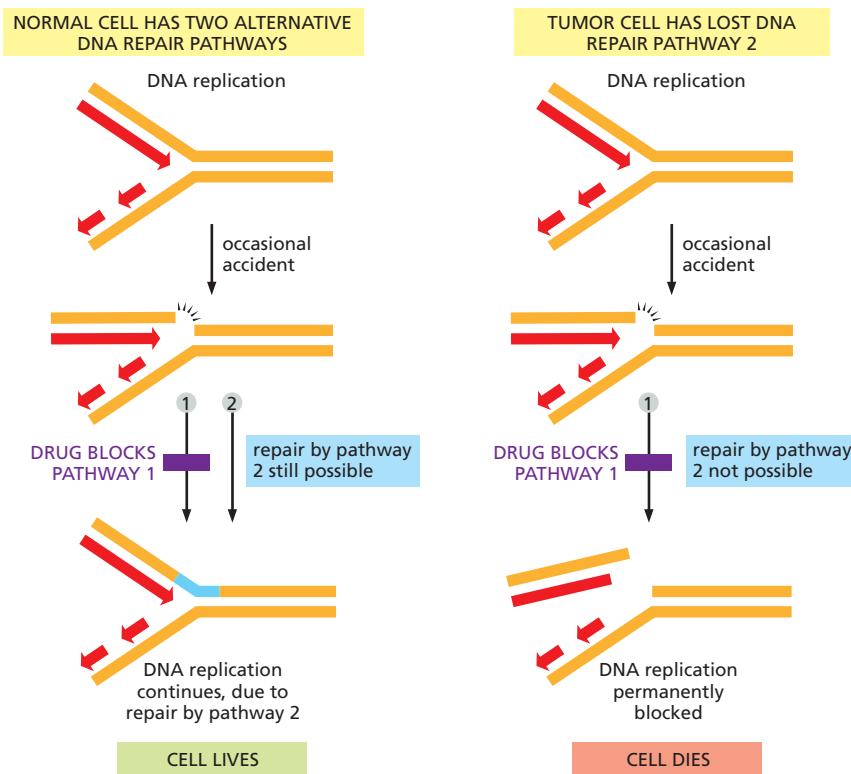


Figure 20–41 How a tumor's genetic instability can be exploited for cancer therapy. As explained in Chapter 5, the maintenance of DNA sequences is so critical for life that cells have evolved multiple pathways for repairing DNA damage and reducing DNA replication errors. As illustrated, a DNA replication fork will stall whenever it encounters a break in a DNA template strand. In this example, normal cells have two different repair pathways that help them to avoid the problem, pathways 1 and 2. They are therefore not harmed by treatment with a drug that blocks repair pathway 1. But, because the inactivation of repair pathway 2 was selected for during the evolution of the tumor cell, the tumor cells are killed by the same drug treatment.

In the actual case that underlies this example, the function of repair pathway 1 (requiring the PARP protein discussed in the text) is to remove persistent, accidental breaks in a DNA single strand before they are encountered by a moving replication fork. Pathway 2 is the recombination-dependent process (requiring the Brca2 and Brca1 proteins) for repairing stalled replication forks illustrated in Figure 5–50. PARP inhibitors have promise for treating cancers with defective *Brca2* or *Brca1* tumor suppressor genes.

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 repair of DNA double-strand breaks by homologous recombination. Brca1 is another protein that is also required for this repair process. Like *Rb*, the *Brca1* and *Brca2* genes were discovered as mutations that predispose humans to cancer—in this case, chiefly 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.

While Brca1 and Brca2 are needed for the repair of DNA double-strand breaks, single-strand breaks are repaired by other machinery, involving an enzyme called PARP (polyADP-ribose polymerase). This understanding of the basic mechanisms of DNA repair led to a striking discovery: drugs that block PARP activity kill *Brca*-deficient cells with extraordinary selectivity. At the same time, 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 persistent breaks in a DNA strand, these breaks can be repaired efficiently by a genetic recombination pathway in normal cells. 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 (Figure 20–41).

PARP inhibitors are still under clinical trial, but they have produced some striking results, causing tumors to regress in many *Brca*-deficient patients and delaying progression of their disease, with relatively few disagreeable side effects. These drugs also appear to be applicable to cancers with other mutations that cause defects in the cell's homologous recombination machinery—a small, though significant, proportion of cancer cases.

PARP inhibition provides an example of the type of rational, highly selective approach to cancer therapy that is beginning to be possible. Along with other new treatments to be discussed below, it raises high hopes for treating many other cancers.

Small Molecules Can Be Designed to Inhibit Specific Oncogenic Proteins

An obvious tactic for treating cancer is to attack a tumor expressing an oncogene with a drug designed to specifically block the function of the protein that the oncogene produces. But how can such a treatment avoid hurting the normal cells that depend on the function of the proto-oncogene from which the oncogene has evolved, and why should the drug kill the cancer cells, rather than simply calm them down? One answer may lie in the phenomenon of *oncogene dependence*. Once a cancer cell has undergone an oncogenic mutation, it will often undergo further mutations, epigenetic changes, or physiological adaptations that make it reliant on the hyperactivity of the initial oncogene, just as drug addicts become reliant on high doses of their drug. Blocking the activity of the oncogenic protein may then kill the cancer cell without significantly harming its normal neighbors. Some remarkable successes have been achieved in this way.

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 results from chromosome breakage and rejoining at the sites of two specific genes, *Abl* and *Bcr*. The fusion of these genes creates a hybrid gene, called *Bcr-Abl*, that codes for a chimeric protein consisting of the N-terminal fragment of *Bcr* fused to the C-terminal portion of *Abl* (Figure 20–42). *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 *imatinib* (trade name Gleevec®), that blocks *Bcr-Abl* (Figure 20–43). 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 imatinib-resistant cancers emerge with a probability of about 5% per year during the early years.

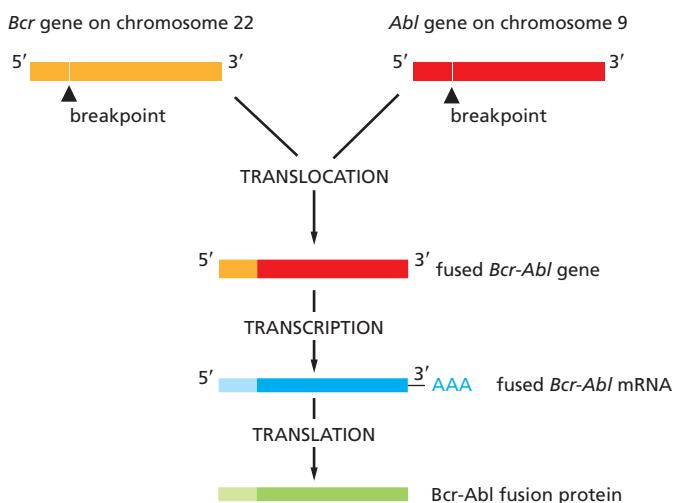


Figure 20–42 The conversion of the *Abi* proto-oncogene into an oncogene in patients with chronic myelogenous leukemia. The chromosome translocation responsible joins the *Bcr* gene on chromosome 22 to the *Abi* gene from chromosome 9, thereby generating a Philadelphia chromosome (see Figure 20–5). The resulting fusion protein has the N-terminus of the *Bcr* protein joined to the C-terminus of the *Abi* tyrosine protein kinase; in consequence, the *Abi* kinase domain becomes inappropriately active, driving excessive proliferation of a clone of hemopoietic cells in the bone marrow.

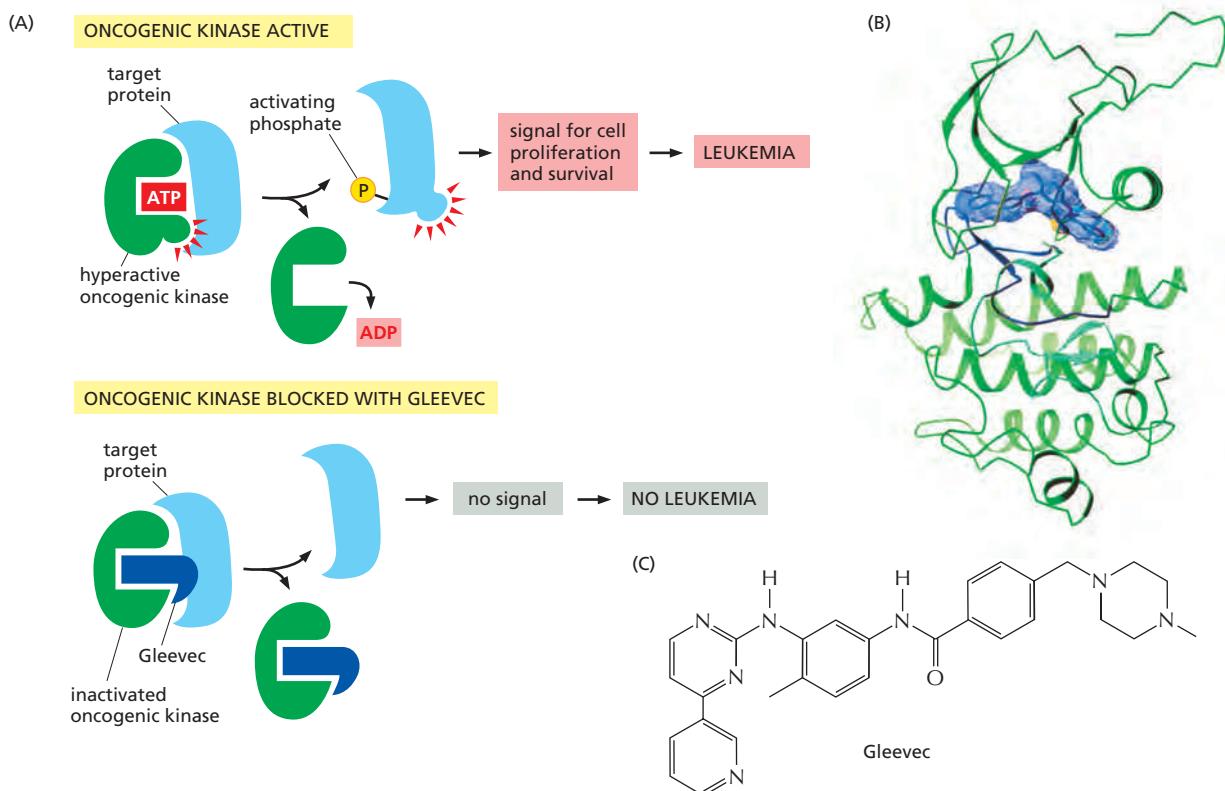


Figure 20-43 How imatinib (Gleevec) blocks the activity of Bcr-Abl protein and halts chronic myelogenous leukemia.
 (A) Imatinib sits in the ATP-binding pocket of the tyrosine kinase domain of Bcr-Abl and thereby prevents Bcr-Abl from transferring a phosphate group from ATP onto a tyrosine residue in a substrate protein. This blocks transmission of a signal for cell proliferation and survival. (B) The structure of the complex of imatinib (solid blue object) with the tyrosine kinase domain of the Abl protein (ribbon diagram), as determined by x-ray crystallography. (C) The chemical structure of the drug. It can be given by mouth; it has side effects, but they are usually quite tolerable. (B, from T. Schindler et al., *Science* 289:1938–1942, 2000. With permission from AAAS.)

Results are not so good for those patients who have already progressed to the more 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 imatinib. 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 imatinib to bind to Bcr-Abl kinase. Second-generation inhibitors that function effectively against a whole range of imatinib-resistant mutants have now been developed. By combining one or more of these new inhibitors with imatinib as the initial therapy (see below), it seems that CML—at least in the chronic (early) stage—may be on its way to becoming a curable disease.

Despite the complications with resistance, the extraordinary success of imatinib is enough to drive home an important 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 and to use them to attack the appropriate cancer cells. Increasing numbers are being developed. These include molecules that target the EGF receptor and are currently approved for the treatment of some lung cancers, as well as drugs that specifically target the B-Raf oncogene in melanomas.

Protein kinases have been relatively easy to inhibit with small molecules like imatinib, and many kinase inhibitors are being produced by pharmaceutical companies in the hope that they can be effective as drugs for some forms of cancer. Many cancers lack an oncogenic mutation in a protein kinase. But most tumors contain inappropriately activated signaling pathways, for which a target

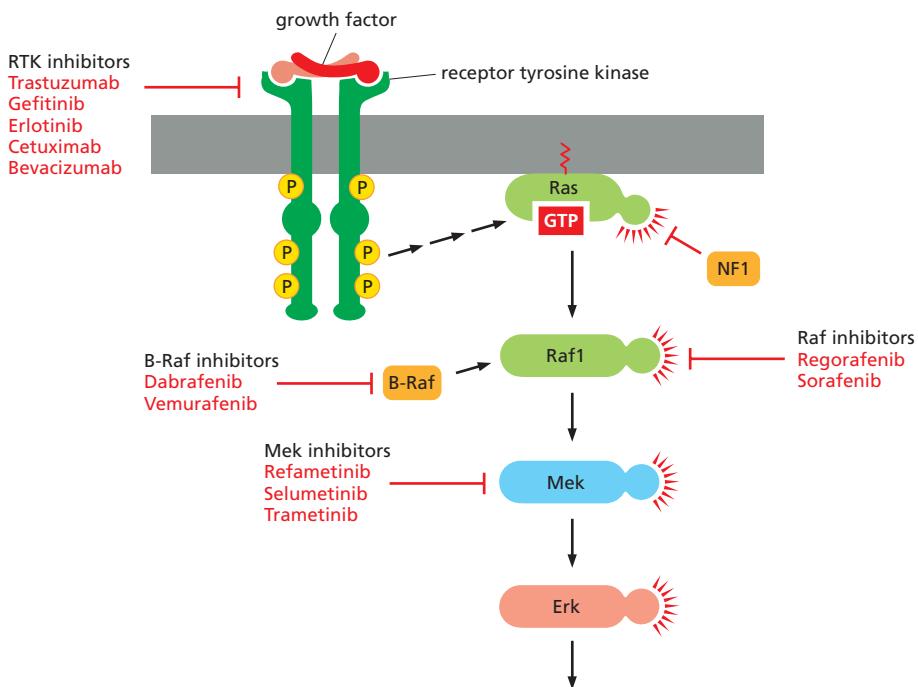


Figure 20–44 Some anticancer drugs and drug targets in the Ras-MAP-kinase signaling pathway. Each of the signaling proteins in this diagram has been identified as a product of a cancer-critical gene, with the exception of Raf1 and Erk. This Ras-MAP-kinase signaling pathway is triggered by a variety of receptor tyrosine kinases (RTKs), including the EGF receptor (see Figures 15–47 and 15–49). Those drugs that are antibodies end in “mab,” while those that are small molecules end in “nib.” (Adapted from B. Vogelstein et al, *Science* 339:1546–1558, 2013.)

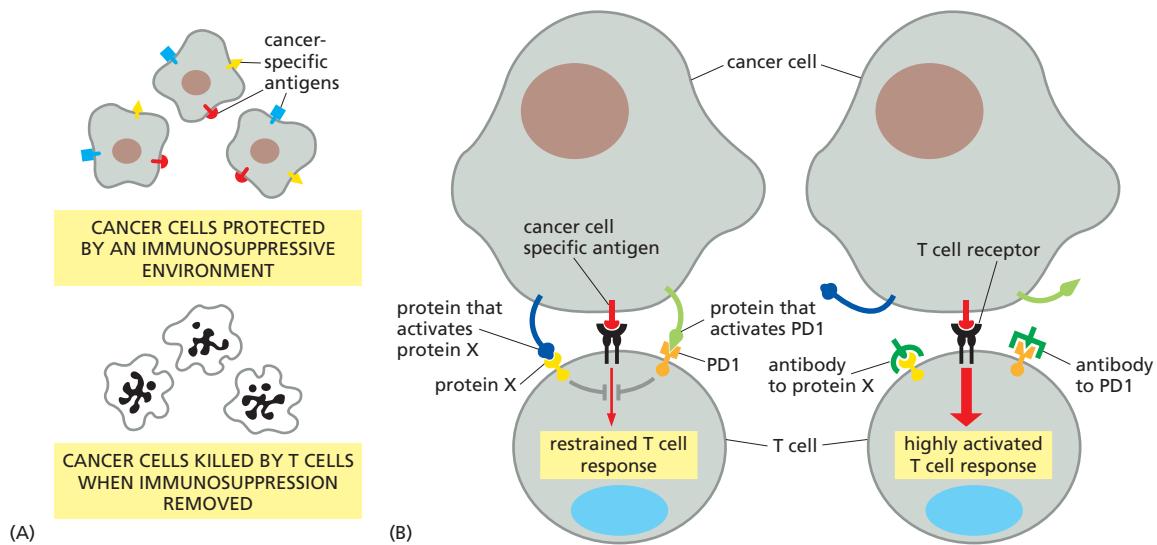
somewhere in the pathway can hopefully be found ([Movie 20.7](#)). As an example, [Figure 20–44](#) displays some of the anticancer drugs and drug targets that are currently being tested for a pathway frequently activated in cancers.

Many Cancers May Be Treatable by Enhancing the Immune Response Against the Specific Tumor

Cancers have complex interactions with the immune system, and its various components may sometimes help as well as hinder tumor progression. But for more than a century it has been a dream of cancer researchers to somehow harness the immune system in a controlled and efficient way to exterminate cancer cells, just as it exterminates infectious organisms. There are finally signs that this dream may one day be realized, at least for some forms of cancer.

The simplest type of immunological therapy, conceptually at least, is to inject the patient with antibodies that target the cancer cells. This approach has had some successes. 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. A monoclonal antibody called *trastuzumab* (trade name *Herceptin*®) that binds to Her2 and inhibits its function slows the growth of breast tumors in humans that overexpress Her2, and it is now an approved therapy for these cancers (see [Figure 20–44](#)). A related approach uses antibodies to deliver poisons to the cancer cells. Antibodies against proteins that are abundant on the surface of a particular type of cancer cell but rare on normal cells can be armed with a toxin that kills those cells that bind the antibody molecule.

A great deal of current excitement centers around a different type of approach, based on the relatively recent recognition that the microenvironment in a tumor is highly immunosuppressive. As a result, the cancer victim's immune system is prevented from destroying the tumor cells. Recall that, from the thousands of genome sequences thus far determined, we know that a typical cancer cell will contain on the order of 50 proteins with a mutation that alters an amino acid sequence, most of these being “passenger” mutations, as previously explained (see p. 1104). Many of these mutant proteins will be recognized by the patient's immune system as foreign, but—to allow the cancer cells to survive throughout the course of tumor progression—the cancer cells have evolved a set of anti-immune defenses. These



defenses include the expression on the cancer cell surface of one or more proteins that bind to inhibitory receptors on activated T cells.

The normal immune system is subject to complex controls that keep its activity within safe bounds and prevent autoimmunity from developing. The inhibitory receptors that are expressed on the surface of activated T cells have an important normal function: they control the immune response by down-regulating the T cell response under appropriate circumstances. But in the context of a tumor, the down-regulation is inappropriate, because it prevents the organism from killing the cancer cells that are threatening its survival.

In its attack on infectious organisms, the natural immune system usually eliminates every last trace of infection and maintains this immunity in the long term. The challenge is to find ways of recruiting the immune system to attack cancers with similar efficiency and specificity, hunting the cancer cells down by virtue of the tumor-specific antigens that they express. With this aim, a new type of anti-cancer therapy focuses on overcoming the immunosuppressive environment in a tumor through the use of specific antibodies that prevent the tumor cells from engaging with the inhibitory receptors on T cells. As illustrated in Figure 20-45A, blocking the action of the immune suppressors with such treatments should unleash an immune attack on the cancer cells. Importantly, multiple antigens are recognized as foreign; thus, the cancer cells cannot escape through the mutational loss of a single antigen, making it difficult for the tumor to escape from the T cell attack.

This is a potentially dangerous strategy. If one provokes the immune system to recognize the cancer cells as targets for destruction, there is a risk of autoimmune side effects with dire consequences for normal tissues of the body, since the cancer cells and the normal cells are close cousins and share most of their molecular features. Nevertheless, several recent successes seem to hold great promise for the future.

One of the many molecules involved in keeping the activity of the normal immune system within safe bounds is a protein called CTLA4 (cytotoxic T-lymphocyte-associated protein 4), which functions as an inhibitory receptor on the surface of T cells. If the function of CTLA4 is blocked, the T cells become more reactive and may mount an attack on cells that they would otherwise leave in peace. In particular, the T cells may attack tumor cells that are recognizably abnormal but whose presence was previously tolerated. With this in mind, cancer immunologists developed a monoclonal antibody, called *ipilimumab*, that binds to CTLA4 and blocks its action. Injected repeatedly into patients with metastatic melanoma, this antibody increases their median lifespan by several months and, in one large trial, enabled as many as a quarter of them to survive for five years

Figure 20-45 Therapies designed to remove the immunosuppressive microenvironment in tumors. (A) The cells in tumors will produce many mutant proteins. As described in Chapter 24, peptides from these proteins will be displayed on MHC complexes on the tumor-cell surface and would normally activate a T cell response that destroys the tumor (see Figure 24-42). However, as schematically illustrated, during the course of tumor progression, the cancer cells have evolved immunosuppressive mechanisms that protect them from such killing. (B) The cells in tumors often protect themselves from immune attack by expressing proteins on their surface that bind to and thereby activate the inhibitory receptors on T cells. As indicated, this makes the tumor susceptible to specific antibody therapies. In this diagram, two such inhibitory receptors are shown, PD1 and a hypothetical protein X. Different tumors are thought to protect themselves by activating different members of a large set of T cell inhibitory receptors, some of which are not yet well characterized.

or more—far beyond expectations for comparable patients without this treatment. Even more promising are recent clinical trials using a combination of two antibodies, one against CTLA4 and the other against PD1, a second cell-surface receptor on T cells that normally restrains their activity.

In clinical trials using such techniques, a substantial fraction of the patients can respond in a dramatic way, with their cancer being driven into remission for years, while the treatment fails to help others with the same type of cancer. One possible explanation is that, while most tumors express proteins that protect them from T-cell attack, these proteins are different for different tumors. Thus, while some tumors will respond dramatically when treated with an antibody that blocks a particular immunosuppressive agent, many others will not. If true, one can foresee an era of personalized immunotherapy, in which each patient's tumor is molecularly analyzed to determine its particular mechanisms of immunosuppression. The patient would then be treated with a specific cocktail of antibodies designed to remove these blocks (see Figure 20–45).

Cancers Evolve Resistance to Therapies

High hopes have to be tempered with sobering realities. We have seen that genetic instability can provide an Achilles heel that cancer therapies can exploit, but at the same time it can make eradicating the disease more difficult by allowing the cancer cells to evolve resistance to therapeutic drugs, often at an alarming rate. This applies even to the drugs that target genetic instability itself. Thus, PARP inhibitors give valuable remission of illness, but in the long term the disease generally comes back. For example, *Brcal*-deficient cancers can sometimes develop resistance to PARP inhibitors by undergoing a second mutation in an affected *Brcal* gene that restores its function. By then, the cancer is already out of control and it may be too late to affect the course of the disease with additional treatments.

There are many different strategies by which cancers can evolve resistance to anticancer drugs. Often, a cancer will 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. In such cases, the initial drug treatment has evidently failed to destroy some tiny fraction of cells in the original tumor-cell population. These cells have escaped death because they carry a protective mutation or epigenetic change, or perhaps simply because they were lurking in a protected environment. They eventually regenerate the cancer by continuing to proliferate, mutating and evolving still further as they do so.

In some cases, cells that are exposed to one anticancer drug evolve 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* or *Abcb1*. 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 (see Movie 11.5). 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.

In the to-and-fro struggle between advanced metastatic cancer and the therapist, as current practice stands, the cancer usually wins in the end. Does it have to be so? As we discuss below, there is reason to think that by attacking a cancer with many weapons at once—instead of using them one after another, each until it fails—it may be possible to do much better.

Combination Therapies May Succeed Where Treatments with One Drug at a Time Fail

Nowadays, cancers caught at an early stage can often be cured, by surgery, radiation, or drugs. For most cancers that have progressed and metastasized widely, however, cure is still beyond us. Treatments such as those described above can

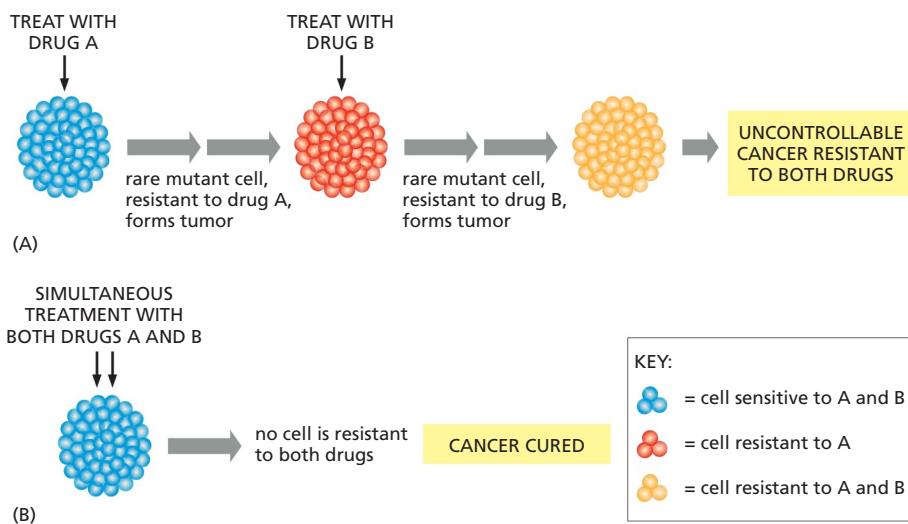


Figure 20-46 Why multidrug treatments can be more effective than sequential treatments for cancer therapy.

(A) Because tumor cells are hypermutable, two single-drug treatments that are given sequentially often allow for the selection of mutant cell clones that are resistant to both drugs. (B) Simultaneous treatment with both drugs can be more effective.

give valuable remissions, but sooner or later these are typically followed by relapse.

Nevertheless, for some relatively rare forms of advanced cancer, curative therapies have been developed. These generally involve a cocktail of several different anticancer agents: by trial and error, certain combinations of cytotoxic drugs have been found to wipe out the cancer completely. Discovering such combinations has hitherto involved a long, hard search. But now, armed with our new tools for identifying the specific genetic lesions that cancer cells contain, the prospects are better.

The logic of combination therapies 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 population carrying the rare mutations that confer resistance to any one drug treatment, there should be no cell carrying the whole set of rare mutations that would confer resistance to several different drugs delivered simultaneously. In contrast, sequential drug treatments will allow the few cells resistant to the first drug to multiply to large numbers. Within this large population of cells resistant to the first drug, a small number of cells are likely to have arisen that are resistant to the next drug also; and so on (Figure 20-46).

We Now Have the Tools to Devise Combination Therapies Tailored to the Individual Patient

Efficient, rational combination drug therapy requires three things. First, we have to identify multiple peculiarities of cancer cells that make them vulnerable in ways that normal cells are not. Second, we have to devise drugs (or other treatments) that target each of these vulnerabilities. Third, we have to match the combination of drugs to the specific set of peculiarities present in the cancer cells of the individual patient.

The first requirement is already partially met: we now have large catalogs of cancer-critical genes that are commonly mutated in cancer cells. The second requirement is harder, but attainable: we have described some remarkable recent successes, and for cancer researchers there is excitement in the air. It is becoming increasingly possible to use our growing knowledge of cell and molecular biology to design new drugs against designated targets. At the same time, efficient, high-throughput automated methods are available to screen large libraries of chemicals for any that may be effective against cells with a given cancer-related defect. In such searches, the goal is synthetic lethality: a cell death that occurs when and only when a particular drug is put together with a particular cancer cell abnormality. Through these and other approaches, the repertoire of precisely targeted anticancer drugs is rapidly increasing.

This brings us to the third requirement: the therapy—the choice of drugs to be given in combination—must be tailored to the individual patient. Here, too, the prospects are bright. Cancers evolve by a fundamentally random process, and each patient is different; but modern methods of genome analysis now let us characterize the cells from a tumor biopsy in exhaustive detail so as to discover which cancer-critical genes are affected in a particular case. Admittedly, this is not straightforward: the tumor cells in an individual patient are heterogeneous and do not all contain the same genetic lesions. With increased understandings of the pathways of cancer evolution, however, and with the experience gained from many different cases, it should become possible to make good guesses at the optimal therapies to use.

From the perspective of the patient, the pace of advance in cancer research can seem frustratingly slow. Each new drug has to be tested in the clinic, first for safety and then for efficacy, before it can be released for general use. And if the drug is to be used in combination with others, the combination therapy must then go through the same long process. Strict ethical rules constrain the conduct of trials, which means that they take time—typically several years. But slow and cautious steps, taken systematically in the right direction, can lead to great advances. There is still far to go, but the examples that we have discussed provide proof of principle and grounds for optimism.

From the cancer research effort, we have learned a great deal of what we know about the molecular biology of the normal cell. Now, more and more, we are discovering how to put that knowledge to use in the battle with cancer itself.

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 cancer cells from normal cells, including the cancer cells' dependence on oncogenic proteins and the defects they harbor in their DNA repair mechanisms. We now have good evidence that, by increasing our understanding of normal cell control mechanisms and exactly how they are subverted in specific cancers, we can eventually devise drugs to kill cancers precisely by attacking specific molecules critical for the growth and survival of the cancer cells. In addition, great progress has recently been made through sophisticated immunological approaches to cancer therapy. 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 The chemical carcinogen 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.

20–2 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.

20–3 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.

WHAT WE DON'T KNOW

- What is required to enable a cancer cell to metastasize?
- How can the molecular analysis of an individual tumor be more effectively used to design effective therapies to kill it?
- Can we identify general features common to all cancer cells—such as their production of misfolded, mutated proteins—that can be used for the targeted destruction of many different types of cancers?
- Can sensitive and reliable blood tests be devised to detect cancers very early, before they have grown to a size where treatment with a single drug will generally be defeated by the survival of a preexisting resistant variant?
- How can the observed environmental effects on cancer rates be exploited to reduce avoidable cancers?
- Can new technologies be devised to reveal exactly how a quiescent micrometastasis converts to a full-blown metastatic tumor?

20–4 The main environmental causes of cancer are the products of our highly industrialized way of life such as pollution and food additives.

Discuss the following problems.

20–5 In contrast to colon cancer, whose incidence increases dramatically with age, incidence of 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?

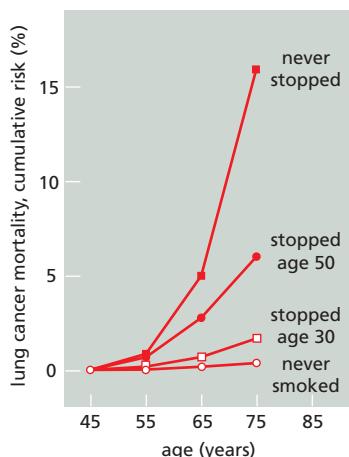


Figure Q20-1 Cumulative risk of lung cancer mortality for nonsmokers, smokers, and former smokers (Problem 20–6). Cumulative risk is the running total of deaths, as a percentage, for each group. Thus, for continuing smokers, 1% died of lung cancer between ages 45 and 55; an additional 4% died between 55 and 65 (giving a cumulative risk of 5%); and 11% more died between 65 and 75 (for a cumulative risk of 16%).

20–6 Mortality due to lung cancer was followed in groups of males in the United Kingdom for 50 years. **Figure Q20-1** 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 individuals can substantially reduce their cumulative risk of dying from lung cancer by stopping smoking. What do you suppose is the biological basis for this observation?

20–7 A small fraction—2 to 3%—of all cancers, across many subtypes, displays a quite remarkable phenomenon: tens to hundreds of rearrangements that primarily involve a single chromosome, or chromosomal region. The breakpoints can be tightly clustered, with several in a few kilobases; the junctions of the rearrangements often involve segments of DNA that were not originally close together on the chromosome. The copy number of various segments within the rearranged chromosome was found to be either zero, indicating deletion, or one, indicating retention.

You can imagine two ways in which such multiple, localized rearrangements might happen: a progressive rearrangements model with ongoing inversions, deletions, and duplications involving a localized area, or a catastrophic model in which the chromosome is shattered into fragments that are stitched back together in random order by nonhomologous end joining (**Figure Q20-2**).

A. Which of the two models in Figure Q20-2 accounts more readily for the features of these highly rearranged chromosomes? Explain your reasoning.

B. For whichever model you choose, suggest how such multiple rearrangements might arise. (The true mechanism is not known.)

C. Do you suppose such rearrangements are likely to be causative events in the cancers in which they are found, or are they probably just passenger events that are unrelated to the cancer? If you think they could be driver events, suggest how such rearrangements might activate an oncogene or inactivate a tumor suppressor gene.

20–8 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?

20–9 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 or the products of tumor suppressor genes? Explain why you would (or would not) select each type of gene.

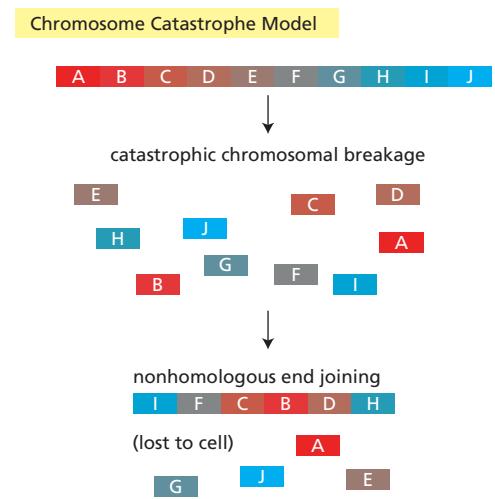
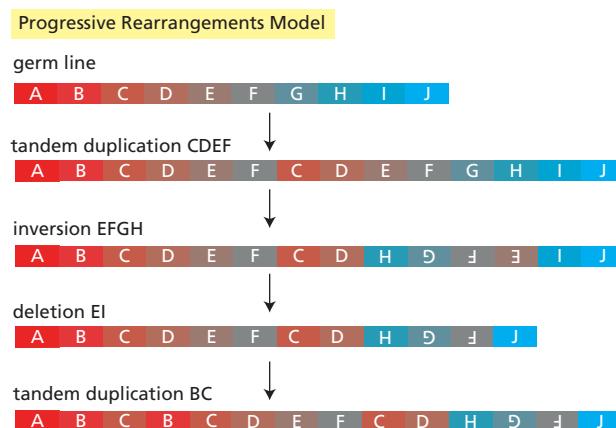


Figure Q20-2 Two models to explain the multiple, localized chromosome rearrangements found in some cancers (Problem 20–7). The progressive rearrangements model shows a sequence of rearrangements that disrupts the chromosome, generating increasingly complex chromosomal configurations. The chromosome catastrophe model shows the chromosome being shattered into fragments that are stitched back together in random order by nonhomologous end joining.

20–10 PolyADP-ribose polymerase (PARP) plays a key role in the repair of DNA single-strand breaks. In the presence of the PARP inhibitor olaparib, single-strand breaks accumulate. When a replication fork encounters a single-strand break, it converts it to a double-strand break, which in normal cells is then repaired by homologous recombination. In cells defective for homologous recombination, however, inhibition of PARP triggers cell death.

Patients who have only one functional copy of the *Brcal* gene, which is required for homologous recombination, are at much higher risk for cancer of the breast and ovary. Cancers that arise in these tissues in these patients can be treated successfully with olaparib. Explain how it is that treatment with olaparib kills the cancer cells in these patients, but does not harm their normal cells.

20–11 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.

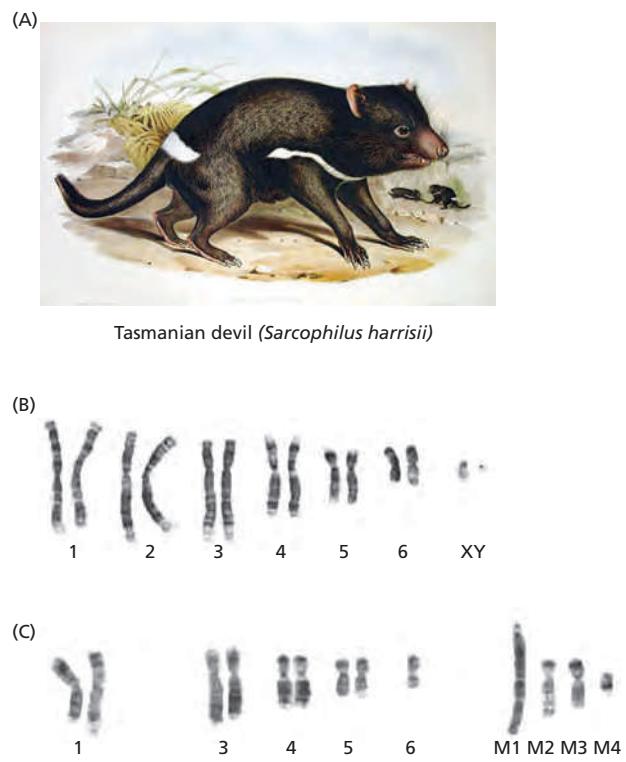


Figure Q20–3 Karyotypes of cells from Tasmanian devils (Problem 20–11). (A) A Tasmanian devil. (B) Normal karyotype for a male Tasmanian devil. The karyotype has 14 chromosomes, including XY. (C) Karyotype of cancer cells found in each of the 11 facial tumors studied. The karyotype has 13 chromosomes, 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). (A, reproduced courtesy of Museum Victoria; B and C, from A.M. Pearson 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. Cambridge, MA: Harvard University Press.
Hanahan D & Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144, 646–674.
Vogelstein B, Papadopoulos N, Velculescu VE et al. (2013) Cancer genome landscapes. *Science* 339, 1546–1558.
Weinberg RA (2013) The Biology of Cancer, 2nd ed. Garland Science: New York.

Cancer as a Microevolutionary Process

- Brown JM & Attardi LD (2005) The role of apoptosis in cancer development and treatment response. *Nat. 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.
Chi P, Allis CD & Wang GG (2010) Covalent histone modifications—miswritten, misinterpreted and mis-erased in human cancers. *Nat. Rev. Cancer* 10, 457–469.
Fidler IJ (2003) The pathogenesis of cancer metastasis: the ‘seed and soil’ hypothesis revisited. *Nat. Rev. Cancer* 3, 453–458.
Hoeijmakers JHH (2001) Genome maintenance mechanisms for preventing cancer. *Nature* 411, 366–374.
Joyce JA & Pollard JW (2009) Microenvironmental regulation of metastasis. *Nat. Rev. Cancer* 9, 239–252.
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.
Stephens PJ, McBride DJ, Lin M-L et al. (2009) Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* 462, 1005–1010.
Thiery JP (2002) Epithelial-mesenchymal transitions in tumour progression. *Nat. Rev. Cancer* 2, 442–454.
Vander Heiden MG, Cantley LC & Thompson CB (2009) Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* 324, 1029–1033.
Zink D, Fischer AH & Nickerson JA (2004) Nuclear structure in cancer cells. *Nat. Rev. Cancer* 4, 677–687.

Cancer-Critical Genes: How They Are Found and What They Do

- Berdasco M & Esteller M (2010) Aberrant epigenetic landscape in cancer: how cellular identity goes awry. *Dev. Cell* 19, 698–711.
- Brognard J & Hunter T (2011) Protein kinase signalling networks in cancer. *Curr. Opin. Genet. Dev.* 21, 4–11.
- Eilers M & Eisenman R (2008) Myc's broad reach. *Genes Dev.* 22, 2755–2766.
- Feinberg AP (2007) Phenotypic plasticity and the epigenetics of human disease. *Nature* 447, 433–440.
- Garraway LA & Lander ES (2013) Lessons from the cancer genome. *Cell* 153, 17–37.
- Greaves M & Maley CC (2012) Clonal evolution in cancer. *Nature* 481, 306–313.
- Junttila MR & Evan GI (2009) p53—a Jack of all trades but master of none. *Nat. Rev. Cancer* 9, 821–829.
- Levine AJ (2009) The common mechanisms of transformation by the small DNA tumor viruses: the inactivation of tumor suppressor gene products: *p53*. *Virology* 384, 285–293.
- Lu P, Weaver VM & Werb Z (2012) The extracellular matrix: a dynamic niche in cancer progression. *J. Cell Biol.* 196, 395–406.
- Mitelman F, Johansson B & Mertens F (2007) The impact of translocations and gene fusions on cancer causation. *Nat. Rev. Cancer* 7, 233–245.
- Negrini S, Gorgoulis VG & Halazonetis TD (2010) Genomic instability—an evolving hallmark of cancer. *Nat. Rev. Mol. Cell Biol.* 11, 220–228.
- Nguyen DX, Bos PD & Massagué J (2009) Metastasis: from dissemination to organ-specific colonization. *Nat. Rev. Cancer* 9, 274–284.
- Radtke F & Clevers H (2005) Self-renewal and cancer of the gut: two sides of a coin. *Science* 307, 1904–1909.
- Rowley JD (2001) Chromosome translocations: dangerous liaisons revisited. *Nat. Rev. Cancer* 1, 245–250.
- Shaw RJ & Cantley LC (2006) Ras, PI(3)K and mTOR signalling controls tumour cell growth. *Nature* 441, 424–430.
- Suvà ML, Riggi N & Bernstein BE (2013) Epigenetic reprogramming in cancer. *Science* 339, 1567–1570.
- Weinberg RA (1995) The retinoblastoma protein and cell cycle control. *Cell* 81, 323–330.

Cancer Prevention and Treatment: Present and Future

- Al-Hajj M, Becker MW, Wicha M et al. (2004) Therapeutic implications of cancer stem cells. *Curr. Opin. Genet. Dev.* 14, 43–47.
- 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.
- Bozic I, Reiter JG, Allen B et al. (2013) Evolutionary dynamics of cancer in response to targeted combination therapy. *eLife* 2, e00747.
- 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.
- Druker BJ & Lydon NB (2000) Lessons learned from the development of an Abi tyrosine kinase inhibitor for chronic myelogenous leukemia. *J. Clin. Invest.* 105, 3–7.
- Huang P & Oliff A (2001) Signaling pathways in apoptosis as potential targets for cancer therapy. *Trends Cell Biol.* 11, 343–348.
- Jain RK (2005) Normalization of tumor vasculature: an emerging concept in antiangiogenic therapy. *Science* 307, 58–62.
- Jonkers J & Berns A (2004) Oncogene addiction: sometimes a temporary slavery. *Cancer Cell* 6, 535–538.
- Kalos M & June CH (2013) Adoptive T cell transfer for cancer immunotherapy in the era of synthetic biology. *Immunity* 39, 49–60.
- Loeb LA (2011) Human cancers express mutator phenotypes: origin, consequences and targeting. *Nat. Rev. Cancer* 11, 450–457.
- Lord CJ & Ashworth A (2012) The DNA damage response and cancer therapy. *Nature* 481, 287–294.
- Pardoll DM (2012) The blockade of immune checkpoints in cancer immunotherapy. *Nat. Rev. Cancer* 12, 252–264.
- Peto J (2001) Cancer epidemiology in the last century and the next decade. *Nature* 411, 390–395.
- Sawyers C (2004) Targeted cancer therapy. *Nature* 432, 294–297.
- Schreiber RD, Old LJ & Smyth MJ (2011) Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. *Science* 331, 1565–1570.
- Sliwkowski MX & Mellman I (2013) Antibody therapeutics in cancer. *Science* 341, 1192–1198.
- Varmus H, Pao W, Politi K et al. (2005) Oncogenes come of age. *Cold Spring Harb. Symp. Quant. Biol.* 70, 1–9.
- Ward RJ & Dirks PB (2007) Cancer stem cells: at the headwaters of tumor development. *Annu. Rev. Pathol.* 2, 175–189.

Development of Multicellular Organisms

CHAPTER
21

An animal or plant starts its life as a single cell—a fertilized egg, or **zygote**. During development, this cell divides repeatedly to produce many different kinds of cells, arranged in a final pattern of spectacular complexity and precision. The goal of developmental cell biology is to understand the cellular and molecular mechanisms that direct this amazing transformation ([Movie 21.1](#)).

Plants and animals have very different ways of life, and they use different developmental strategies; in this chapter, we focus mainly on animals. Four processes are fundamental to animal development: (1) cell proliferation, which produces many cells from one; (2) cell-cell interactions, which coordinate the behavior of each cell with that of its neighbors; (3) cell specialization, or **differentiation**, which creates cells with different characteristics at different positions; and (4) cell movement, which rearranges the cells to form structured tissues and organs ([Figure 21–1](#)). It is on the fourth point that plant development differs radically: plant cells are unable to migrate or move independently through the embryo because each one is contained within a cell wall, through which it is cemented to its neighbors, as discussed in Chapter 19.

In a developing animal embryo, the four fundamental processes are happening in a kaleidoscopic variety of ways, as they give rise to different parts of the organism. Like the members of an orchestra, the cells in the embryo have to play their individual parts in a highly coordinated manner. In the embryo, however, there is no conductor—no central authority—to direct the performance. Instead, development is a self-assembly process in which the cells, as they grow and proliferate, organize themselves into increasingly complex structures. Each of the millions of cells has to choose for itself how to behave, selectively utilizing the genetic instructions in its chromosomes.

At each stage in its development, the cell is presented with a limited set of options, so that its developmental pathway branches repeatedly, reflecting a large set of sequential choices. Like the decisions we make in our own lives, the choices made by the cell are based on its internal state—which largely reflects its history—and on current influences from other cells, especially its close neighbors. To understand development, we need to know how each choice is controlled and how it depends on previous choices. Beyond that, we need to understand how the choices, once made, influence the cell's chemistry and behavior, and how cell behaviors act synergistically to determine the structure and function of the body.

IN THIS CHAPTER

- OVERVIEW OF DEVELOPMENT**
- MECHANISMS OF PATTERN FORMATION**
- DEVELOPMENTAL TIMING**
- MORPHOGENESIS**
- GROWTH**
- NEURAL DEVELOPMENT**

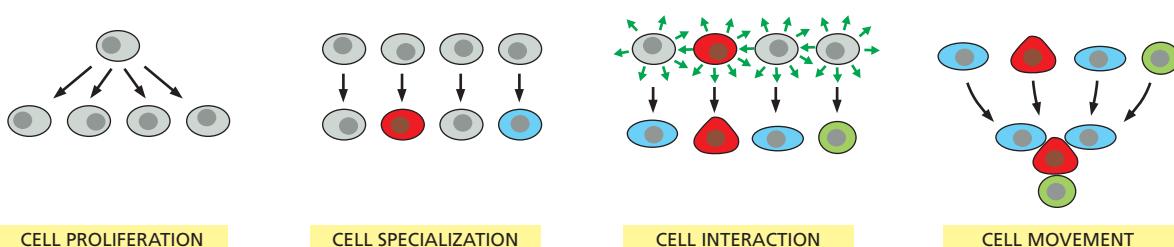


Figure 21–1 The four essential cell processes that allow a multicellular organism to be made.

As cells become specialized they change not only their chemistry but also their shape and their attachments to other cells and to the extracellular matrix. They move and rearrange themselves to create the complex architecture of the body, with all its tissues and organs, each structured precisely and defined in size. To understand this process of form generation, or *morphogenesis*, we will need to take account of the mechanical, as well as the biochemical, interactions between the cells.

At first glance, 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. Remarkably, however, research in the past 30 years has revealed that much of the basic machinery of development is essentially the same in all animals—not just in all vertebrates, but in all the major phyla of invertebrates too. Recognizably similar, evolutionarily related molecules define the specialized animal cell types, mark the differences between body regions, and help create the animal body pattern. Homologous proteins are often functionally interchangeable between very different species. Thus, a human protein produced artificially in a fly, for example, can perform the same function as the fly's own version of that protein (Figure 21–2). Thanks to an underlying unity of mechanisms, developmental biologists have been making great strides toward a coherent understanding of animal development.

We begin this chapter with an overview of some of the basic mechanisms that operate in animal development. We then discuss, in sequence, how cells in the embryo diversify to form patterns in space, how the timing of developmental events is controlled, how cell movements contribute to morphogenesis, and how the size of an animal is regulated. We end by considering the most challenging aspect of development—the mechanisms that enable a highly complex nervous system to form.

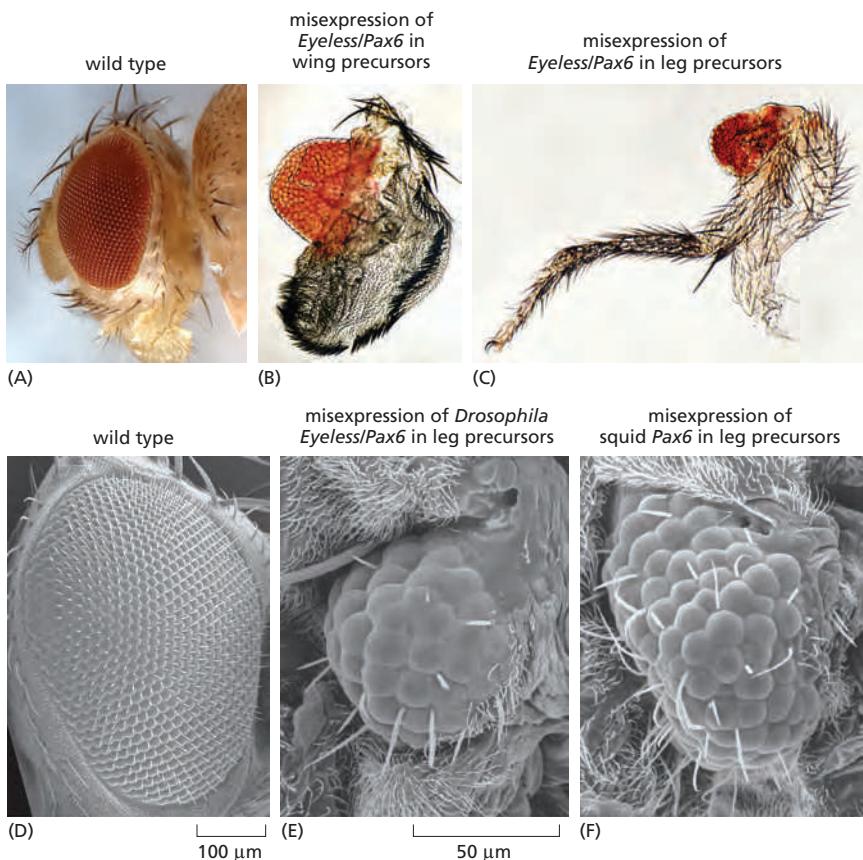


Figure 21–2 Homologous proteins can function interchangeably. (A–C) The Eyeless protein (also called Pax6) controls eye development in *Drosophila* and, when misexpressed during development, can cause an eye to form in an abnormal site, such as a wing (B) or a leg (C). The scanning electron micrographs show a patch of eye tissue on the leg of a fly resulting from misexpression of *Drosophila* Eyeless (E) and of squid Pax6 (F). The homologous protein from a human or practically any animal possessing eyes, when similarly misexpressed in a transgenic fly, has the same effect. The entire eye of a normal *Drosophila* is shown for comparison in (A) and (D). (B–C, courtesy of Georg Halder; D–F, from S. I. Tomarev, et al. *Proc. Natl Acad. Sci. USA* 94:2421–2426, 1997. With permission from National Academy of Sciences.)

OVERVIEW OF DEVELOPMENT

Animals live by eating other organisms. Thus, despite their remarkable diversity, animals as different as worms, mollusks, insects, and vertebrates share anatomical features that are fundamental to this way of life. Epidermal cells form a protective outer layer; gut cells absorb nutrients from ingested food; muscle cells allow movement; and neurons and sensory cells control behavior. These diverse cell types are organized into tissues and organs, forming a sheet of skin covering the exterior, a mouth for feeding, and an internal gut tube to digest food—with muscles, nerves, and other tissues arranged in the space between the skin and the gut tube. Many animals have clearly defined axes—an anteroposterior axis, with mouth and brain anterior and anus posterior; a dorsoventral axis, with back dorsal and belly ventral; and a left-right axis. In this section, we discuss some fundamental mechanisms underlying animal development, beginning with how the basic animal body plan is established.

Conserved Mechanisms Establish the Basic Animal Body Plan

The shared anatomical features of animals develop through conserved mechanisms. After fertilization, the zygote usually divides rapidly, or **cleaves**, to form many smaller cells; during this cleavage, the embryo, which cannot yet feed, does not grow. This phase of development is initially driven and controlled entirely by the material deposited in the egg by the mother. The embryonic genome remains inactive until a point is reached when maternal mRNAs and proteins rather abruptly begin to be degraded. The embryo's genome is activated, and the cells cohere to form a **blastula**—typically a solid or a hollow fluid-filled ball of cells. Complex cell rearrangements called **gastrulation** (from the Greek “gaster,” meaning “belly”) then transform the blastula into a multilayered structure containing a rudimentary internal gut (Figure 21–3). Some cells of the blastula remain external, constituting the **ectoderm**, which will give rise to the epidermis and the nervous system; other cells invaginate, forming the **endoderm**, which will give rise to the gut tube and its appendages, such as lung, pancreas, and liver. Another group of cells moves into the space between ectoderm and endoderm and forms the **mesoderm**, which will give rise to muscles, connective tissues, blood, kidney, and various other components. Further cell movements and accompanying cell differentiations create and refine the embryo's architecture.

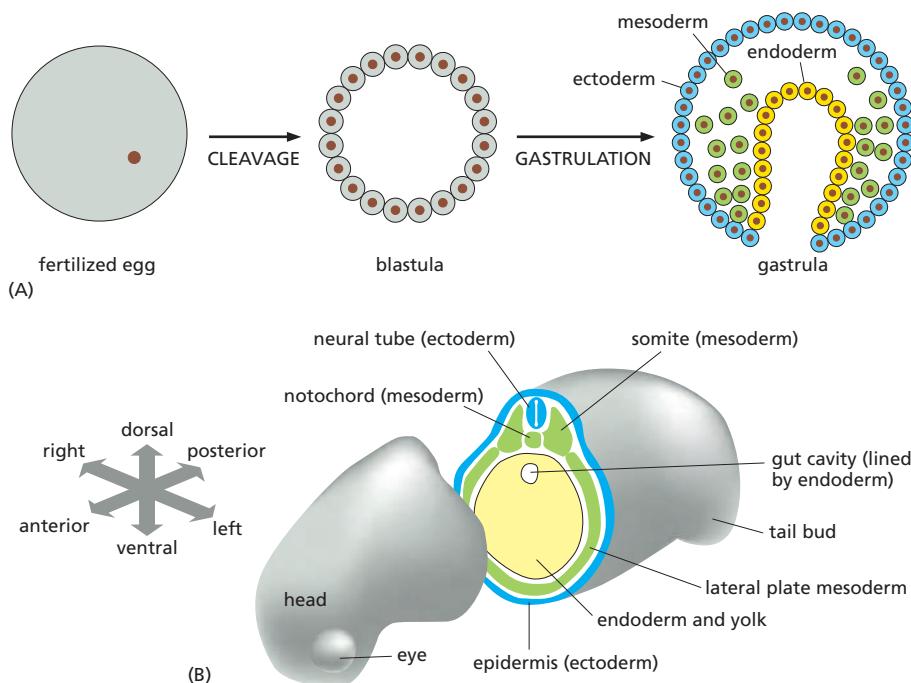


Figure 21–3 The early stages of development, as exemplified by a frog.

(A) A fertilized egg divides to produce a blastula—a sheet of epithelial cells often surrounding a cavity. During gastrulation, some of the cells tuck into the interior to form the mesoderm (green) and endoderm (yellow). Ectodermal cells (blue) remain on the outside. (B) A cross section through the trunk of an amphibian embryo shows 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. The endoderm forms 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 structures. The salivary glands, liver, pancreas, trachea, and lungs, for example, all develop from the wall of the 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 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, blood vessels, and blood cells—and the tubules, ducts, and supporting tissues of the kidneys and gonads. The notochord forms from the mesoderm and serves as the core of the future backbone and the source of signals that coordinate the development of surrounding tissues.

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. (B, after T. Mohun et al., *Cell* 22:9–15, 1980. With permission from Elsevier.)

The ectoderm, mesoderm, and endoderm formed during gastrulation constitute the three **germ layers** of the early embryo. Many later developmental transformations will produce the elaborately structured organs. But the basic body plan and axes set up in miniature during gastrulation are preserved into adult life, when the organism may be billions of times larger (**Movie 21.2**).

The Developmental Potential of Cells Becomes Progressively Restricted

Concomitant with the refinement of the body plan, the individual cells become more and more restricted in their developmental potential. During the blastula stages, cells are often **totipotent** or **pluripotent**—they have the potential to give rise to all or almost all of the cell types of the adult body. The pluripotency is lost as gastrulation proceeds: a cell located in the endodermal germ layer, for example, can give rise to the cell types that will line the gut or form gut-derived organs such as the liver or pancreas, but it no longer has the potential to form mesoderm-derived structures such as skeleton, heart, or kidney. Such a cell is said to be *determined* for an endodermal fate. Thus, **cell determination** starts early and progressively narrows the options as the cell steps through a programmed series of intermediate states—guided at each step by its genome, its history, and its interactions with neighbors. The process reaches its limit when a cell undergoes **terminal differentiation** to form one of the highly specialized cell types of the adult body (**Figure 21–4**). Although there are cell types in the adult that retain some degree of pluripotency, their range of options is generally narrow (discussed in Chapter 22).

Cell Memory Underlies Cell Decision-Making

Underlying the richness and astonishingly complex outcomes of development is **cell memory** (see p. 404). Both the genes a cell expresses and the way it behaves depend on the cell's past, as well as on its present circumstances. The cells of our body—the muscle cells, the neurons, the skin cells, the gut cells, and so on—maintain their specialized characters largely because they retain a record of the extracellular signals their ancestors received during development, rather than because they continually receive such instructions from their surroundings. Despite their radically different phenotypes, they retain the same complete genome that was present in the zygote; their differences arise instead from differential gene expression. We have discussed the molecular mechanisms of gene regulation, cell memory, cell division, cell signaling, and cell movement in previous chapters. In this chapter, we shall see how these basic processes are collectively deployed to create an animal.

Several Model Organisms Have Been Crucial for Understanding Development

The anatomical features that animals share have undergone many extreme modifications in the course of evolution. As a result, the differences between species are usually more striking to our human eye than the similarities. But at the level of the underlying molecular mechanisms and the macromolecules that mediate them, the reverse is true: the similarities among all animals are profound and extensive. Through more than half a billion years of evolutionary divergence, all animals have retained unmistakably similar sets of genes and proteins that are responsible for generating their body plans and for forming their specialized cells and organs.

This astonishing degree of evolutionary conservation was discovered not by broad surveys of animal diversity, but through intensive study of a small number of representative species—the model organisms discussed in Chapter 1. For animal developmental biology, the most important have been the fly *Drosophila melanogaster*, the frog *Xenopus laevis*, the roundworm *Caenorhabditis elegans*, the mouse *Mus musculus*, and the zebrafish *Danio rerio*. In discussing the mechanisms of development, we shall draw our examples mainly from these few species.

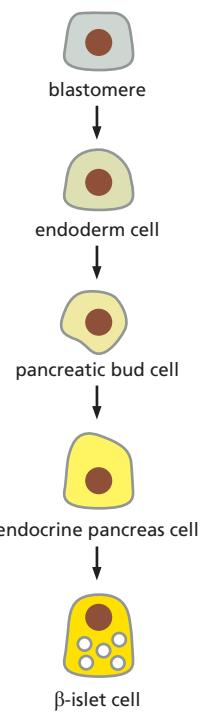


Figure 21–4 The lineage from blastomere to differentiated cell type.

As development proceeds, cells become more and more specialized. Blastomeres have the potential to give rise to most or all cell types. Under the influence of signaling molecules and gene regulatory factors, cells acquire more restricted fates until they differentiate into highly specialized cell types, such as the pancreatic β -islet cells that secrete the hormone insulin.

Genes Involved in Cell–Cell Communication and Transcriptional Control Are Especially Important for Animal Development

What are the genes that animals share with one another but not with other kingdoms of life? These would be expected to include genes required specifically for animal development but not needed for unicellular existence. Comparison of animal genomes with the genome of budding yeast—a unicellular eukaryote—suggests that three classes of genes are especially important for multicellular organization. The first class includes genes that encode proteins used for cell–cell adhesion and cell signaling; hundreds of human genes encode signal proteins, cell-surface receptors, cell adhesion proteins, or ion channels that are either not present in yeast or present in much smaller numbers. The second class includes genes encoding proteins that regulate transcription and chromatin structure: more than 1000 human genes encode transcription regulators, but only about 250 yeast genes do so. As we shall see, the development of animals is dominated by cell–cell interactions and by differential gene expression. The third class of non-coding RNAs has a more uncertain status: it includes genes that encode microRNAs (miRNAs); there are at least 500 of these in humans. Along with the regulatory proteins, they play a significant part in controlling gene expression during animal development, but the full extent of their importance is still unclear. The loss of individual miRNA genes in *C. elegans*, where their functions have been well studied, rarely leads to obvious phenotypes, suggesting that the roles of miRNAs during animal development are often subtle, serving to fine-tune the developmental machinery rather than to form its core structures.

Regulatory DNA Seems Largely Responsible for the Differences Between Animal Species

As discussed in Chapter 7, each gene in a multicellular organism is associated with many thousands of nucleotides of noncoding DNA that contains regulatory elements. These regulatory elements determine when, where, and how strongly the gene is to be expressed, according to the transcription regulators and chromatin structures that are present in the particular cell (Figure 21–5). Consequently, a change in the regulatory DNA, even without any change in the coding DNA, can alter the logic of the gene-regulatory network and change the outcome of development.

As discussed in Chapter 4, when we compare the genomes of different animal species, we find that evolution has altered the coding and regulatory DNA to different extents. The coding DNA, for the most part, has been highly conserved, the noncoding regulatory DNA much less so. It seems that changes in regulatory DNA are largely responsible for the dramatic differences between one class of animals and another (see p. 227). We can view the protein products of the coding sequences as a conserved kit of common molecular parts, and the regulatory DNA as instructions for assembly: with different instructions, the same kit of parts can be used to make a whole variety of different body structures. We will return to this important concept later.

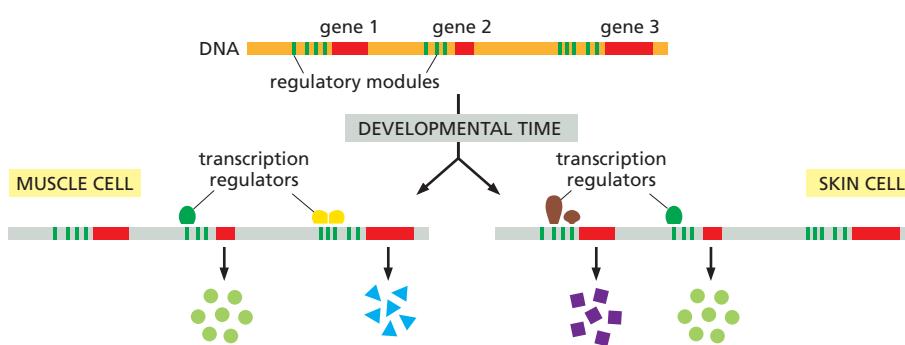


Figure 21–5 Regulatory DNA defines the gene expression patterns in development. The genome is the same in a muscle cell as in a skin cell, but different genes are active because these cells express different transcription regulators that bind to gene regulatory elements. For example, transcription regulators in skin cells recognize a regulatory element in gene 1, leading to its activation, whereas a different set of regulators is present in muscle cells, binding to and activating gene 3. Transcriptional regulators that activate the expression of gene 2 are present in both cell types.

Small Numbers of Conserved Cell–Cell Signaling Pathways Coordinate Spatial Patterning

Spatial patterning of a developing animal requires that cells become different according to their positions in the embryo, which means that cells must respond to extracellular signals produced by other cells, especially their neighbors. In what is probably the commonest mode of spatial patterning, a group of cells starts out with the same developmental potential, and a signal from cells outside the group then induces one or more members of the group to change their character. This process is called *inductive signaling*. Generally, the inductive signal is limited in time and space so that only a subset of the cells capable of responding—the cells close to the source of the signal—take on the induced character (Figure 21–6). Some inductive signals depend on cell-cell contact; others act over a longer range and are mediated by molecules that diffuse through the extracellular medium or are transported in the bloodstream (see Figure 15–2).

Most of the known inductive events in animal development are governed by a small number of highly conserved signaling pathways, including transforming growth factor- β (TGF β), Wnt, Hedgehog, Notch, and receptor tyrosine kinase (RTK) pathways (discussed in Chapter 15). The discovery of the limited vocabulary that developing cells use for intercellular communication has emerged over the past 25 years as one of the great simplifying features of developmental biology.

Through Combinatorial Control and Cell Memory, Simple Signals Can Generate Complex Patterns

But how can this small number of signaling pathways generate the huge diversity of cells and patterns? Three kinds of mechanisms are responsible. First, through gene duplication, the basic components of a pathway often come to be encoded by small families of closely related homologous genes. This allows for diversity in the operation of the pathway, according to which family member is employed in a given situation. Notch signaling, for example, may be mediated by Notch1 in one tissue, but by its homolog Notch4 in another. Second, the response of a cell to a given signal protein depends on the other signals that the cell is receiving concurrently (Figure 21–7A). As a result, different combinations of signals can generate a large variety of different responses. Third, and most fundamental, the effect of activating a signaling pathway depends on the previous experiences of the responding cell: past influences leave a lasting mark, registered in the state of the cell's chromatin and the selection of transcription regulatory proteins and RNA molecules that the cell contains. This cell memory enables cells with different

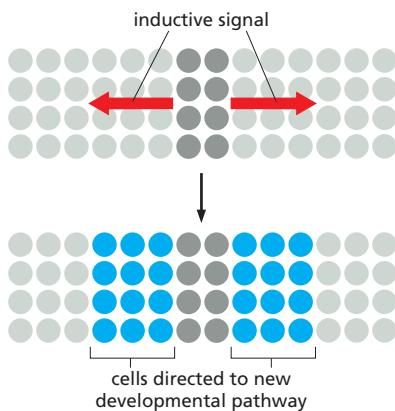


Figure 21–6 Inductive signaling.

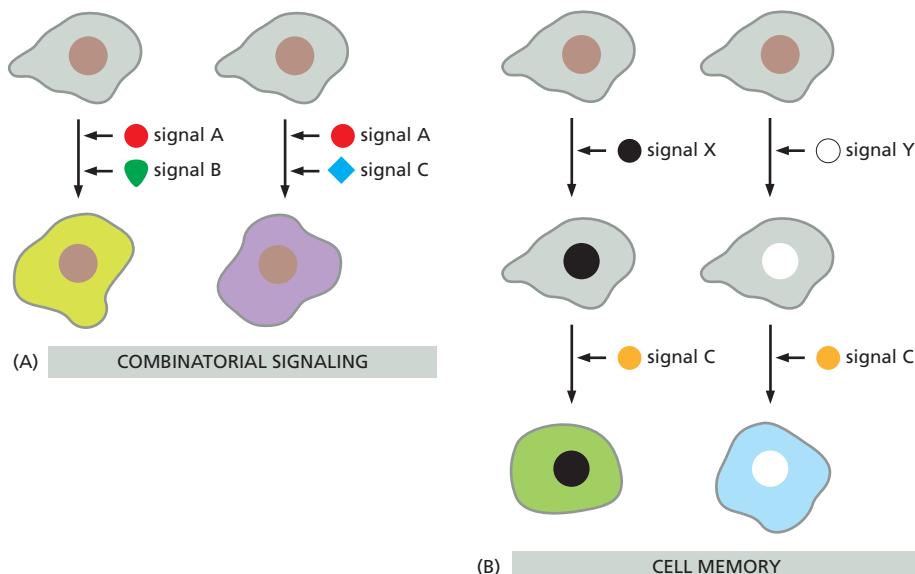


Figure 21–7 Two mechanisms for generating different responses to the same inductive signal. (A) In combinatorial signaling, the effect of a signal depends on the presence of other signals received at the same time. (B) Through cell memory, previous signals (or other events) can leave a lasting trace that alters the response to the current signal (see Figure 7–54). The memory trace is represented here in the coloring of the cell nucleus.

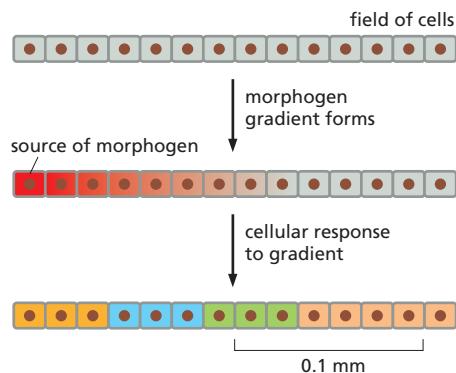


Figure 21–8 Gradient formation and interpretation. A gradient forms by localized production of an inducer—a morphogen—that diffuses away from its source. Different concentrations of morphogen (or different durations of exposure) induce different gene expression patterns and cell fates in responding cells. Diffusive transport can generate gradients only over short distances, and morphogens generally act over distances of 1 mm or less.

histories to respond to the same signals differently (Figure 21–7B). Thus, the same few signaling pathways can be used repeatedly at different times and places with different outcomes, so as to generate patterns of unlimited complexity.

Morphogens Are Long-Range Inductive Signals That Exert Graded Effects

Signal molecules often govern simple yes-no choices—one outcome when their concentration is high, another when it is low. In many cases, however, the responses are more finely graded: a high concentration of a signal molecule may, for example, direct cells into one developmental pathway, an intermediate concentration into another, and a low concentration into yet another.

One common way to generate such different concentrations of a signal molecule is for the molecule to diffuse out from a localized signaling source, creating a 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 (Figure 21–8). A signal molecule that imposes a pattern on a whole field of cells in this way is called a **morphogen**. In the simplest case, a specialized group of cells produces a morphogen at a steady rate, and the morphogen is then degraded as it diffuses away from this source. The speed of diffusion and the half-life of the morphogen will together determine the range and steepness of its resulting gradient (Figure 21–9).

This simple mechanism can be modified in various ways. Receptors on the surface of cells along the way, for example, may trap the diffusing morphogen and cause it to be endocytosed and degraded, shortening its effective half-life. Alternatively, the morphogen may bind to molecules in the extracellular matrix such as heparan sulfate proteoglycan (discussed in Chapter 19), thereby greatly reducing its diffusion rate.

Lateral Inhibition Can Generate Patterns of Different Cell Types

Morphogen gradients, and other kinds of inductive signal, exploit an existing asymmetry in the embryo to create further asymmetries and differences between cells: already, at the outset, some cells are specialized to produce the morphogen and thereby impose a pattern on another class of cells that are sensitive to it. But

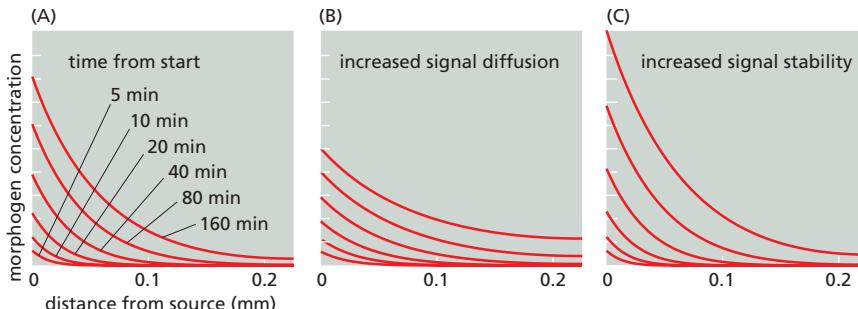


Figure 21–9 Setting up a signal gradient by diffusion. (A–C) Each graph shows six successive stages in the buildup of the concentration of a signal molecule that is produced at a steady rate at the origin, with production starting at time 0. In all cases, the molecule undergoes degradation as it diffuses away from the source, and the graphs are calculated on the assumption that diffusion is occurring along two axes in space (for example, radially from a source in an epithelial sheet). (A) The pattern of the morphogen assuming that the molecule has a half-life of 170 minutes, and that it diffuses with an effective diffusion constant of $D = 1 \mu\text{m}^2 \text{ sec}^{-1}$, typical of a small protein molecule in extracellular tissues. Note that the gradient is already close to its steady-state form within an hour and that the concentration at steady state falls off exponentially with distance. (B) A threefold increase in the diffusion constant of the morphogen extends its range but lowers its concentration next to the source, whereas (C) a threefold increase in morphogen half-life increases its concentration throughout the tissue. Effects of the morphogen will depend not just on its concentration at some critical moment, but also on how each target cell integrates its response over time. (Courtesy of Patrick Müller.)

Figure 21–10 Genesis of asymmetry through lateral inhibition and 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.

what if there is no clear initial asymmetry? Can a regular pattern arise spontaneously within a set of cells that are initially all alike?

The answer is yes. The fundamental principle underlying such *de novo* pattern formation is positive feedback: cells can exchange signals in such a way that any small initial discrepancy between cells at different sites becomes self-amplifying, driving the cells toward different fates. This is most clearly illustrated in the phenomenon of *lateral inhibition*, a form of cell-cell interaction that forces close neighbors to become different and thereby generates fine-grained patterns of different cell types.

Consider a pair of adjacent cells that start off in a similar state. Each of these cells can both produce and respond to a certain signal molecule X, with the added rule that the stronger the signal a cell receives, the weaker the signal it generates (Figure 21–10). If one cell produces more X, the other is forced to produce less. This gives rise to a positive feedback loop that tends to amplify any initial difference between the two adjacent cells. Such a difference may arise from a bias imposed by some present or past external factor, or it may simply originate from spontaneous random fluctuations, or “noise”—an inevitable feature of the genetic control circuitry in cells (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 production of X in cell 1 to rise higher still; and so on, until a steady state is reached where cell 1 produces a lot of X and cell 2 produces very little. In the standard case, the signal molecule X acts in the receiving cell by regulating gene transcription, and the result is that the two cells are driven along different pathways of differentiation.

In almost all tissues, a balanced mixture of different cell types is required. Lateral inhibition provides a common way to generate the mixture. As we shall see, lateral inhibition is very often mediated by exchange of signals at cell-cell contacts via the Notch signaling pathway, driving cell diversification by enabling individual cells that express one set of genes to direct their immediate neighbors to express a different set, in exactly the way we have described (see also Figure 15–58).

Short-Range Activation and Long-Range Inhibition Can Generate Complex Cellular Patterns

Lateral inhibition mediated by the Notch pathway is not the only example of pattern generation through **positive feedback**: there are other ways in which, through the same basic principle, a system that starts off homogeneous and symmetrical can pattern itself spontaneously, even in the absence of an external morphogen. Positive feedback processes mediated by diffusible signal molecules can operate over broad arrays of cells to create many types of spatial patterns. Mechanisms of this sort are called *reaction-diffusion systems*. For example, a substance A (a short-range activator) may stimulate its own production in the cells that contain it and in their immediate neighbors, while also causing these cells to produce a signal I (a long-range inhibitor) that diffuses widely and inhibits the production of A in cells farther away. If the cells all start the same, but one group gains a slight advantage by making a little more A than the rest, the asymmetry can be self-amplifying

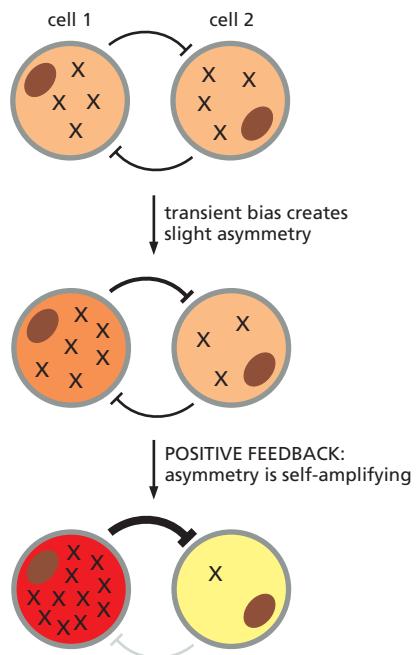


Figure 21–11 Pattern generation by a reaction-diffusion system. From (A) a uniform field of cells, (B) local positive feedback and (C) long-range inhibition can (D) generate patterns within the initially uniform field. The patterns can be complex, resembling the spots of a leopard (as shown) or the stripes of a zebra; or they can be simple, with creation of a single cluster of specialized cells that can, for example, go on to serve as the source of a morphogen gradient.

(**Figure 21–11**). Such short-range activation combined with long-range inhibition can account for the formation of clusters of cells within an initially homogeneous tissue that become specialized as localized **signaling centers**.

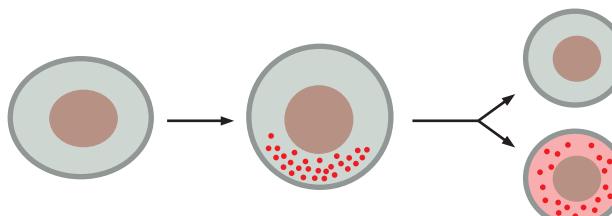
Asymmetric Cell Division Can Also Generate Diversity

Cell diversification does not always depend on extracellular signals: in some cases, daughter cells are born different as a result of an **asymmetric cell division**, in which some important molecule or molecules are distributed unequally between the two daughters. This asymmetric inheritance ensures that the two daughter cells develop differently (**Figure 21–12**). Asymmetric division is a common feature of early development, where the fertilized egg already has an internal pattern and cleavage of this large cell segregates different determinants into separate blastomeres. We shall see that asymmetric division also plays a part in some later developmental processes.

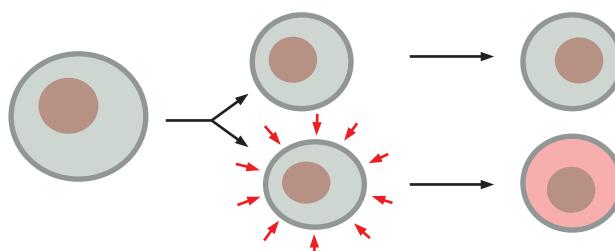
Initial Patterns Are Established in Small Fields of Cells and Refined by Sequential Induction as the Embryo Grows

The signals that organize the spatial pattern of cells in 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—and directs choices between several developmental options for the cells on which it acts. Yet 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 plus other interactions that add successive levels of detail on an initially simple sketch. For example, as soon as two types of cells are present in a developing tissue, one of them can produce a signal that induces a subset of the neighboring cells to specialize in a third way. The third cell type can in turn signal



1. asymmetric division: sister cells born different



2. symmetric division: sister cells become different as a result of influences acting on them after their birth

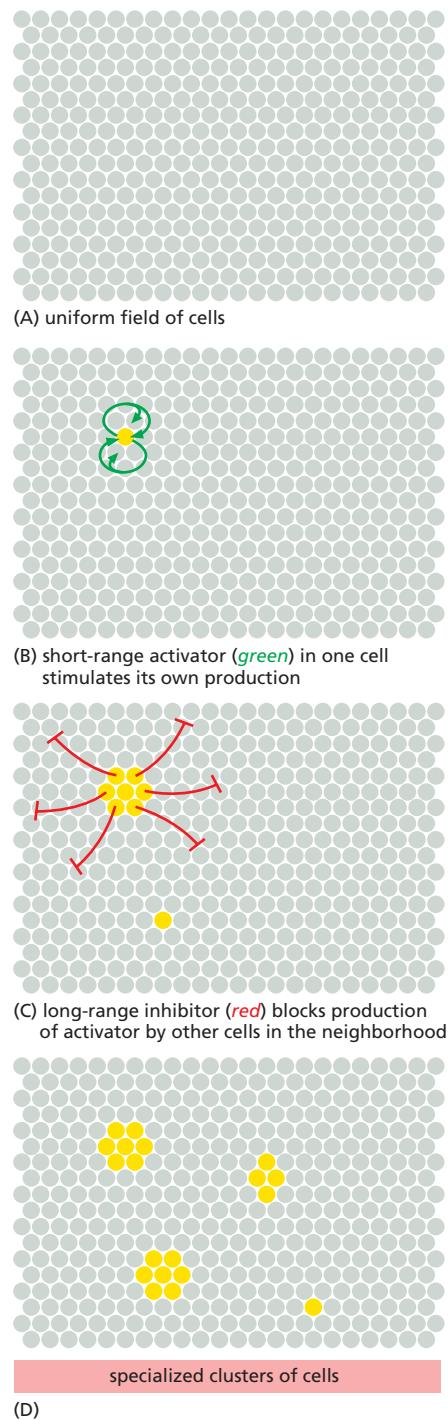


Figure 21–12 Two ways of making sister cells different.

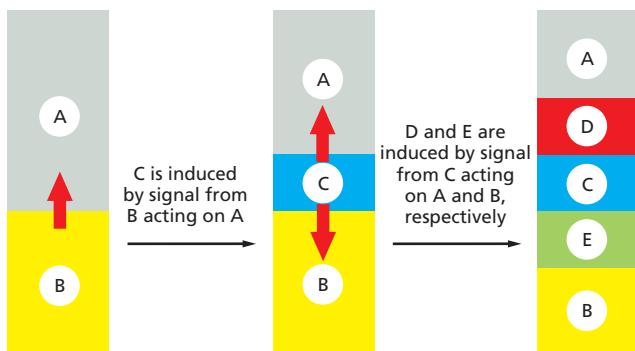


Figure 21–13 Patterning by sequential induction. A series of inductive interactions can generate many types of cells, starting from only a few.

back to the other two cell types nearby, generating a fourth and a fifth cell type, and so on (Figure 21–13).

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.

Developmental Biology Provides Insights into Disease and Tissue Maintenance

The rapid progress in understanding animal development has been one of the great success stories in biology over the last few decades, and it has important practical implications. Some 2 to 5% of all human babies are born with anatomical abnormalities, such as heart malformations, truncated limbs, cleft palate, or spina bifida. Advances in developmental biology help us understand how these defects arise, even if we cannot yet prevent or cure most of them.

Less obvious, but even more important from a practical point of view, is that developmental biology provides insights into the workings of cells and tissues in the adult body. Developmental processes do not halt at birth; they continue throughout life, as tissues are maintained and repaired. The fundamental mechanisms of cell growth and division, cell-cell signaling, cell memory, cell adhesion, and cell movement are involved in adult tissue maintenance and repair—just as they are in embryo development.

Embryos are simpler than adults, and they allow us to analyze such basic processes more easily. Studies of the early *Drosophila* embryo, for example, were crucial to the discovery of several conserved signaling pathways, including the Wnt, Hedgehog, and Notch pathways. They also provided the key to understanding the central role of these pathways in the maintenance of normal adult human tissues and in diseases such as cancer.

In Chapter 22, we shall consider how other developmental mechanisms operate in the normal adult body, especially in tissues that are continually renewed by means of stem cells—including the gut, skin, and the hematopoietic system. But now, we must look more closely at the way in which an early embryo generates its spatial pattern of specialized cells, beginning with the transformations that create the adult body plan.

Summary

Animal development is a self-assembly process, in which the cells of the embryo become different from one another and organize themselves into increasingly complex structures. The process begins with a single large cell—the fertilized egg. This cell cleaves to form many smaller cells, producing a blastula. The blastula undergoes gastrulation to generate the three germ layers of the embryo—ectoderm, mesoderm, and endoderm—consisting of cells determined for different fates. As development continues, the cells become more and more narrowly specialized according to their locations and their interactions with one another. Through cell memory, these cell-cell interactions, even though transient, can have lasting effects on each

cell's internal state. Thus, a succession of simple cues that a cell receives at different times can direct it along a complex developmental pathway. At each step, the cell becomes further restricted in the range of final states open to it. The process reaches its limit when the cell differentiates to form one of the specialized cell types of the adult body.

Differences between developing cells arise in various ways and have to be properly coordinated in space. In one common strategy, initially similar cells within a group become different by exposure to different levels of an inductive signal or morphogen emanating from a source outside the group. Neighboring cells can also become different by lateral inhibition, in which a cell signals to its neighbors not to follow the same fate. These cell-cell interactions are mediated by a small number of highly conserved signaling pathways, which are used repeatedly in different organisms and at different times during development. Not all cell diversification arises by cell-cell interactions, however: daughter cells can be born different as a result of asymmetric cell division.

Regulators of transcription and chromatin structure bind to regulatory DNA and determine the fate of each cell. 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.

Development has been most thoroughly studied in a handful of model organisms. But most of the genes and mechanisms thereby identified are used in all animals and repeatedly at different stages of development. Thus, insights from worms, flies, fish, frogs, and mice deeply inform our understanding of embryology, birth defects, and adult tissue maintenance in humans.

MECHANISMS OF PATTERN FORMATION

A developing multicellular organism has to create a pattern in fields of cells where there was little or none before. Some of the early microscopists imagined the entire shape and structure of the human body to be already present in the sperm as a "homunculus," a miniature human; after fertilization, the homunculus would simply grow and generate a full-sized human. We now know that this view is incorrect and that development is a progression from simple to complex, through a gradual refinement of an animal's anatomy. To see how the whole sequence of events of spatial patterning and cell determination is set in train, we must return to the egg and the early embryo.

Different Animals Use Different Mechanisms to Establish Their Primary Axes of Polarization

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 specializations. Gastrulation occurs in all animal embryos, but the details of its timing, of the associated pattern of cell movements, and of the shape and size of the embryo as gastrulation proceeds are highly variable. Likewise, there is great variation in the time and manner in which the primary axes of the body become marked out. However, this *polarization* of the embryo usually becomes discernible very early, before gastrulation begins: it is the first step of spatial patterning.

Three axes generally have to be established. The *animal-vegetal* (*A-V*) axis, in most species, defines which parts are to become internal (through the movements of gastrulation) and which are to remain external. (The bizarre name dates from a century ago and has nothing to do with vegetables.) The *anteroposterior* (*A-P*) axis specifies the locations of future head and tail. The *dorsoventral* (*D-V*) axis specifies the future back and belly.

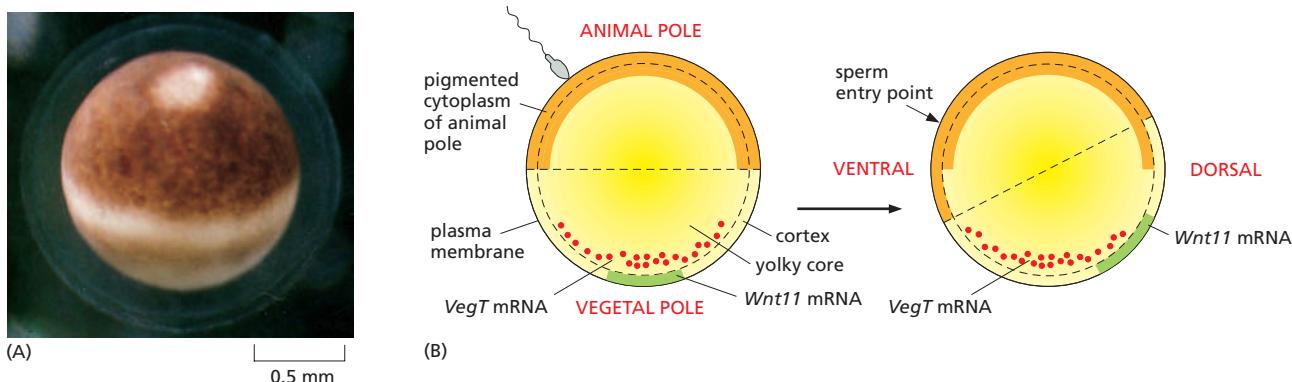


Figure 21–14 The frog egg and its asymmetries. (A) Side view of a *Xenopus* egg photographed just before fertilization. (B) The asymmetric distribution of molecules inside the egg, and how this changes following fertilization so as to define a dorsoventral as well as an animal-vegetal asymmetry. Fertilization, 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; the direction of rotation 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. Vegetally localized *VegT* defines the vegetal source of signals that will induce endoderm and mesoderm. (A, courtesy of Tony Mills.)

At one extreme, the egg is spherically symmetrical, and the axes only become defined during embryogenesis. The mouse comes close to being an example, with little obvious sign of polarity in the egg. Correspondingly, the **blastomeres** produced by the first few cell divisions seem to be all alike and are remarkably adaptable. If the early mouse 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 two-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.

At the opposite extreme, the structure of the egg defines the future axes of the body. This is the case for most species, including insects such as *Drosophila*, as we shall see shortly. Many other organisms lie between the two extremes. The egg of the frog *Xenopus*, for example, has a clearly defined A-V axis even before fertilization: the nucleus near the top defines the animal pole, while the mass of yolk (the embryo’s food supply, destined to be incorporated in the gut) toward the bottom defines the vegetal pole. Several types of mRNA molecules are already localized in the vegetal cytoplasm of the egg, where they produce their protein products. After fertilization, these mRNAs and proteins act in and on the cells in the lower and middle part of the embryo, giving the cells there specialized characters, both by direct effects and by stimulating the production of secreted signal proteins. For example, mRNA encoding the transcription regulator *VegT* is deposited at the vegetal pole during oogenesis. After fertilization, this mRNA is translated, and the resulting *VegT* protein activates a set of genes that code for signal proteins that induce mesoderm and endoderm, as discussed later.

The D-V axis of the *Xenopus* embryo, by contrast, is defined through the act of fertilization. Following entry of the sperm, the outer cortex of the egg cytoplasm rotates relative to the central core of the egg, so that the animal pole of the cortex becomes slightly shifted to one side (Figure 21–14). Treatments that block the rotation allow cleavage to occur normally but produce an embryo with a central gut and no dorsal structures or D-V asymmetry. Thus, this cortical rotation is required to define the D-V axis of the future body by creating the D-V axis of the egg.

The site of sperm entry that biases the direction of the cortical rotation in *Xenopus*, perhaps through the centrosome that the sperm brings into the egg—inasmuch as the rotation is associated with a reorganization of the microtubules nucleated from the centrosome in the egg cytoplasm. The reorganization leads to a microtubule-based transport of several cytoplasmic components, including the mRNA coding for *Wnt11*, a member of the Wnt family of signal proteins, moving

it toward the future dorsal side (see Figure 21–14). This mRNA is soon translated and the Wnt11 protein secreted from cells that form in that region of the embryo activates the Wnt signaling pathway (see Figure 15–60). This activation is crucial for triggering the cascade of subsequent events that will organize the dorsoventral axis of the body. (The A-P axis of the embryo will only become clear later, in the process of gastrulation.)

Although different animal species use a variety of different mechanisms to specify their axes, the outcome has been relatively well conserved in evolution: head is distinguished from tail, back from belly, and gut from skin. It seems that it does not much matter what tricks the embryo uses to break the initial symmetry and set up this basic body plan.

Studies in *Drosophila* Have Revealed the Genetic Control Mechanisms Underlying Development

It is the fly *Drosophila*, more than any other organism, that has provided the key to our present understanding of how genes govern development. Decades of genetic study culminated in a large-scale genetic screen, focusing especially on the early embryo and searching for mutations that disrupt its pattern. This revealed that the key developmental genes fall into a relatively small set of functional classes defined by their mutant phenotypes. The discovery of these genes and the subsequent analysis of their functions was a famous *tour de force* and had a revolutionary impact on all of developmental biology, earning its discoverers a Nobel Prize. Some parts of the developmental machinery revealed in this way are conserved between flies and vertebrates, some parts not. But the logic of the experimental approach and the general strategies of genetic control that it revealed have transformed our understanding of multicellular development in general.

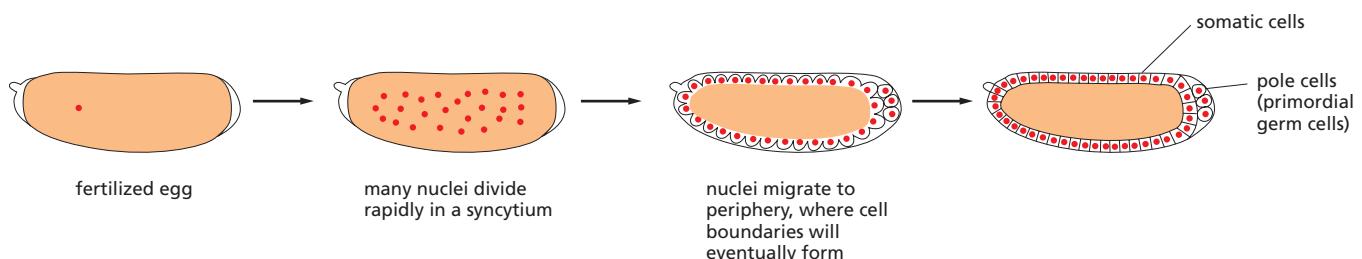
To understand how the early developmental machinery operates in *Drosophila*, it is important to note a peculiarity of fly development. Like the eggs of other insects, but unlike most vertebrates, the *Drosophila* egg—shaped like a cucumber—begins its development with an extraordinarily rapid series of nuclear divisions without cell division, producing multiple nuclei in a common cytoplasm—a **syncytium**. The nuclei then migrate to the cell cortex, forming a structure called the *syncytial blastoderm*. After about 6000 nuclei have been produced, the plasma membrane folds inward between them and partitions them into separate cells, converting the syncytial blastoderm into the *cellular blastoderm* (Figure 21–15).

We shall see that the initial patterning of the *Drosophila* embryo depends on signals that diffuse through the cytoplasm at the syncytial stage and exert their actions on genes in the rapidly dividing nuclei, before the partitioning of the egg into separate cells. Here, there is no need for the usual forms of cell-cell signaling; neighboring regions of the syncytial blastoderm can communicate by means of transcription regulatory proteins that move through the cytoplasm of the giant multinuclear cell.

Egg-Polarity Genes Encode Macromolecules Deposited in the Egg to Organize the Axes of the Early *Drosophila* Embryo

As in most insects, the main axes of the future body of *Drosophila* are defined before fertilization by a complex exchange of signals between the developing egg,

Figure 21–15 Development of the *Drosophila* egg from fertilization to the cellular blastoderm stage.



or *oocyte*, and the *follicle cells* that surround it in the ovary. In the stages before fertilization, the anteroposterior and dorsoventral axes of the future embryo become defined by four systems of **egg-polarity genes** that create landmarks—either mRNA or protein—in the developing oocyte. Following fertilization, each landmark serves as a beacon, providing a signal that organizes the developmental process in its neighborhood.

The nature of the genes emerged from studies of mutants in which the patterning of the embryo was altered. One class of mutations gave embryos with disrupted polarity—for example, tail-end structures at both ends of the body, with no head-end structures. This class of mutations identified the set of egg-polarity genes. The egg-polarity gene responsible for the signal that organizes the anterior end of the embryo is called ***Bicoid***. A deposit of *Bicoid* mRNA molecules is localized, before fertilization, at the anterior end of the egg. Upon fertilization, the mRNA is translated to produce Bicoid protein. This protein is an intracellular morphogen and transcription regulator that diffuses away from its source to form a concentration gradient within the syncytial cytoplasm, with its maximum at the head end of the embryo (**Figure 21–16**). The different concentrations of Bicoid along the A-P axis help determine different cell fates by regulating the transcription of genes in the nuclei of the syncytial blastoderm (discussed in Chapter 7).

Of the three other egg-polarity gene systems, two contribute to patterning the syncytial nuclei along the A-P axis and one to patterning them along the D-V axis. Together with the *Bicoid* group of genes, and acting in a broadly similar way, their gene products mark out three fundamental partitions of body regions—head versus rear, dorsal versus ventral, and endoderm versus mesoderm and ectoderm—as well as a fourth partition, no less fundamental to the body plan of animals: the distinction between germ cells and somatic cells (**Figure 21–17**).

The egg-polarity genes have a further special feature: they are all **maternal-effect genes**, in that it is the mother's genome rather than the zygote's genome that is critical. For example, 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 offspring is a female, she cannot deposit any functional *Bicoid* mRNA into her own eggs, which will therefore develop into headless embryos, regardless of the father's genotype.

The egg-polarity genes act first in a hierarchy of gene systems that define a progressively more detailed pattern of body parts. In the next few pages, we begin with the molecular mechanisms that pattern the developing *Drosophila* embryo and larva along the A-P axis, before considering the patterning along the D-V axis.

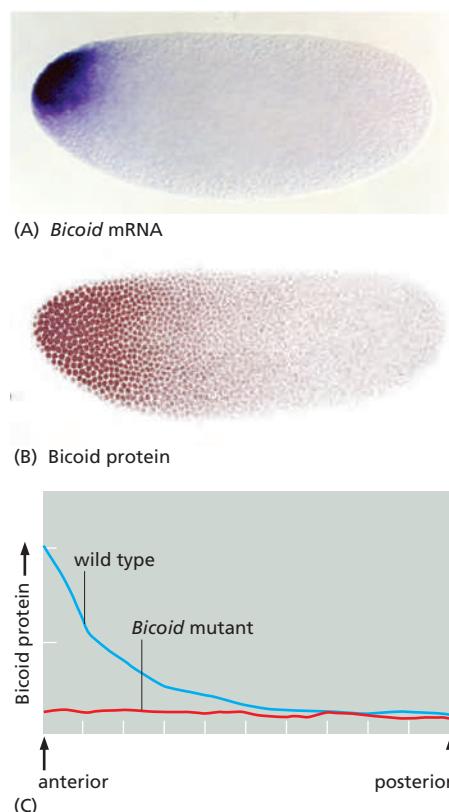


Figure 21–16 The Bicoid protein gradient. (A) *Bicoid* mRNA is deposited at the anterior pole during oogenesis. (B) Local translation followed by diffusion generates the Bicoid protein gradient. (C) Absence of the Bicoid protein gradient in embryos from *Bicoid* homozygous mutant mothers. (A and B, courtesy of Stephen Small.)

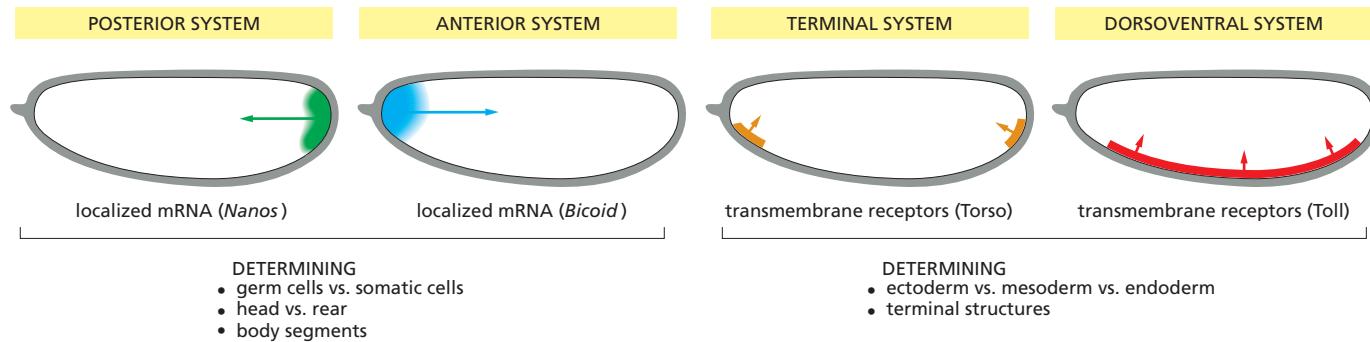
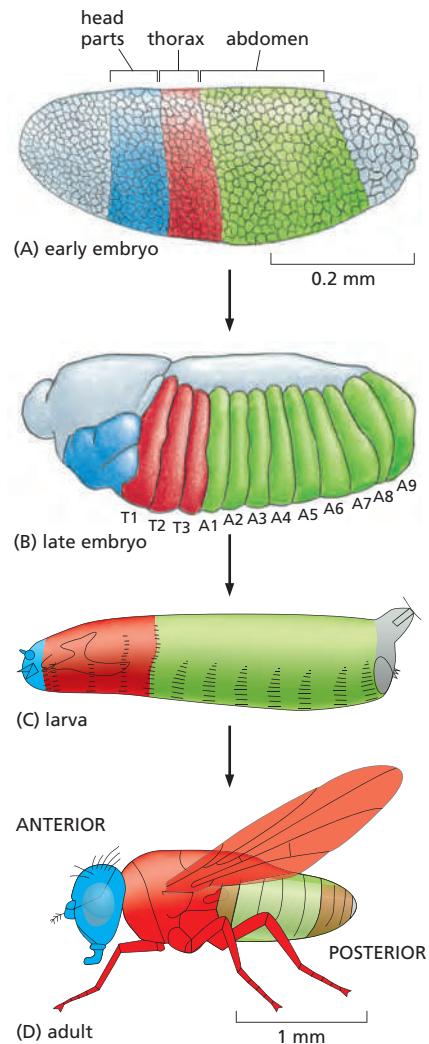


Figure 21–17 The organization of the four egg-polarity gradient systems in *Drosophila*. *Nanos* is a translational repressor that governs the formation of the abdomen. Localized *Nanos* mRNA is also incorporated into the germ cells as they form at the posterior of the embryo, and *Nanos* protein is necessary for germ-line development. *Bicoid* protein is a transcriptional activator that determines the head and thoracic regions. *Toll* and *Torso* are receptor proteins that are distributed all over the membrane but are activated only at the sites indicated by the coloring, through localized exposure to the extracellular ligands *Spaetzle* (the ligand for *Toll*) and *Trunk* (the ligand for *Torso*). *Toll* activity determines the mesoderm and *Torso* activity determines the formation of terminal structures.

Figure 21–18 The origins of the *Drosophila* body segments. (A) At 3 hours, the embryo (shown in side view) is at the blastoderm stage and no segmentation is visible, although a fate map can be drawn showing the future segmented regions (color). (B) At 10 hours, all the segments are clearly defined (T1: first thoracic segment; A1: first abdominal segment). See Movie 21.3. (C) The segments of the *Drosophila* larva and their correspondence with regions in the embryo. (D) The segments of the *Drosophila* adult and their correspondence with regions in the embryo.



Three Groups of Genes Control *Drosophila* Segmentation Along the A-P Axis

The body of an insect is divided along its A-P axis into a series of **segments**. The segments are repetitions of a theme with variations: each segment forms highly specialized structures, but all built according to a similar fundamental plan (Figure 21–18). The gradients of transcription regulators set up along the A-P axis in the early embryo by the egg-polarity genes are the prelude to creation of the segments. These regulators initiate the orderly transcription of *segmentation genes*, which refine the pattern of gene expression to define the boundaries and ground plan of the individual segments. Segmentation genes are expressed by subsets of cells in the embryo, and their products are the first components that the embryo's own genome contributes to embryonic development; they are therefore called *zygotic-effect genes*, to distinguish them from the earlier-acting maternal-effect genes. Mutations in segmentation genes can alter either the number of segments or their basic internal organization.

The **segmentation genes** fall into three groups according to their mutant phenotypes (Figure 21–19). It is convenient to think of the three groups as acting in sequence, although in reality their functions overlap in time. First to be expressed is a set of at least six **gap genes**, whose products mark out coarse A-P subdivisions of the embryo. Mutations in a gap gene eliminate one or more groups of adjacent segments: in the mutant *Krüppel*, for example, the larva lacks eight segments. Next comes a set of eight **pair-rule genes**. Mutations in these genes cause a series of deletions affecting alternate segments, leaving the embryo with only half as many segments as usual; although all the mutants display this two-segment periodicity, they differ in the precise pattern. Finally, there are at least 10 **segment-polarity genes**, in which mutations produce 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 parallel with the segmentation process, a further set of genes—the *homeotic selector*, or *Hox*, genes—serves to define and preserve the differences between one segment and the next, as we describe shortly.

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 A-P axis, each distinguished by a different pattern of gene expression. Molecular genetics has helped to reveal how this system works.

A Hierarchy of Gene Regulatory Interactions Subdivides the *Drosophila* Embryo

Like *Bicoid*, most of the segmentation genes encode transcription regulator proteins. Their control by the egg-polarity genes and their actions on one another and on still other genes can be deciphered by comparing gene expression in normal and mutant embryos. By using appropriate probes to detect RNA transcripts or their protein products, one can observe genes switch on and off in changing patterns. By comparing these patterns in different mutants, one can begin to discern the logic of the entire gene control system.

The products of the egg-polarity genes provide the global positional signals in the early embryo (see Figure 21–17). The Bicoid protein, as we have seen, acts as

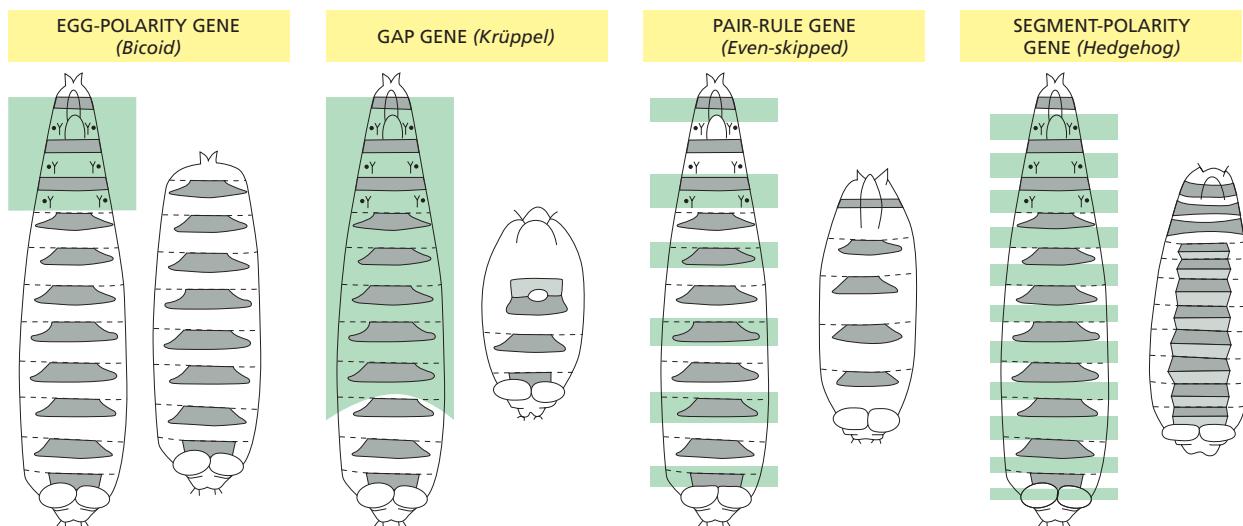


Figure 21-19 Examples of the phenotypes of mutations affecting egg-polarity genes and the three types of segmentation genes. In each case, the areas shaded in green on the normal larva (left) are deleted in the mutant or are replaced by mirror-image duplicates of the unaffected regions. (Modified from C. Nüsslein-Volhard and E. Wieschaus, *Nature* 287:795–801, 1980. With permission from Macmillan Publishers Ltd.)

a morphogen and activates different sets of genes at different positions along the A-P axis: some gap genes are only activated in regions with high levels of Bicoid, others only where levels of Bicoid are lower. After the gap gene products refine their positions by mutual repression, they provide a second tier of positional signals that act more locally to regulate finer details of patterning. Gap genes act by controlling the expression of yet other genes, including the pair-rule genes. 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 the segment-polarity genes, which collaborate with one another to define the internal pattern of each individual segment (**Figure 21-20**).

The initial steps in creation of the segmental pattern occur before cellularization of the syncytial blastoderm and are governed by the combinatorial effects of transcription regulators, as discussed in detail in Chapter 7 for the regulation of the expression of the pair-rule gene *Even-skipped* (see pp. 394–396). After cellularization, the segment-polarity genes further subdivide each segment into smaller domains. A large subset of the segment-polarity genes codes for components of two signaling pathways—the Wnt pathway and the Hedgehog pathway, including the secreted signal proteins Wingless (the first-named member of the Wnt family) and Hedgehog. (The Hedgehog pathway was first discovered through study of *Drosophila* segmentation, and it takes its name from the prickly appearance of the surface of the *Hedgehog* mutant embryo.) Wingless and Hedgehog are synthesized in different bands of cells that serve as signaling centers within each segment. The two proteins mutually maintain each other's expression, while regulating the expression of genes such as *Engrailed* in neighboring cells (**Figure 21-21**). In such a manner, a series of sequential inductions creates a fine-grained pattern of gene expression within each segment.

Egg-Polarity, Gap, and Pair-Rule Genes Create a Transient Pattern That Is Remembered by Segment-Polarity and Hox Genes

The gap genes and pair-rule genes are activated within the first few hours after fertilization. Their mRNA products initially appear in patterns that only approximate the final picture; then, within a short time, this fuzzy initial pattern resolves itself into a regular, crisply defined system of stripes. But this pattern itself is unstable

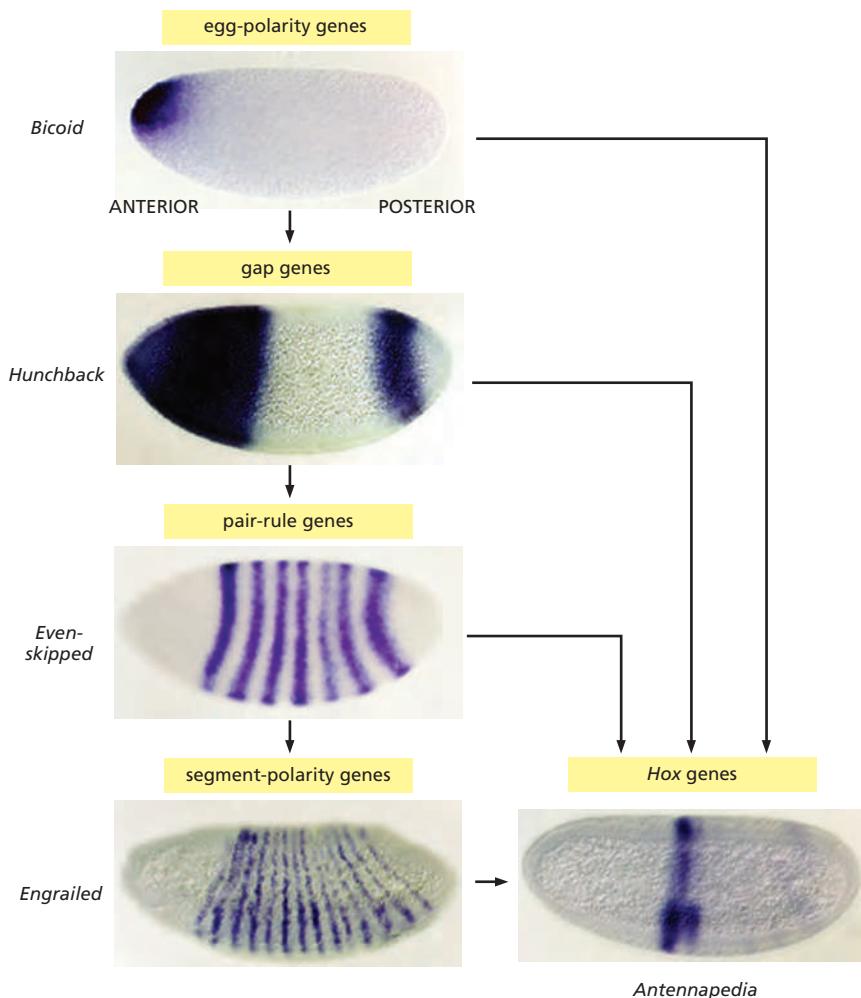


Figure 21–20 An example of the regulatory hierarchy of egg-polarity, segmentation, and *Hox* genes. As discussed in the text, there are three groups of segmentation genes. The photographs show mRNA expression patterns of representative examples of genes of each type. (Courtesy of Stephen Small.)

and transient: as the embryo proceeds through gastrulation and beyond, the pattern disintegrates. The genes' actions, however, have passed on an enduring memory of their patterns of expression by inducing the expression of certain segment polarity genes along with *Hox* genes (discussed shortly). After a period of pattern refinement mediated by cell-cell interactions, the expression patterns of these new groups of patterning genes is stabilized to provide *positional labels* that serve to maintain the segmental organization of the larva and adult fly.

The segment-polarity gene *Engrailed* provides a good example. Its RNA transcripts form a series of 14 bands in the cellular blastoderm, each approximately one-cell wide. These stripes lie immediately anterior to similar stripes of expression of another segment polarity gene, *Wingless*. As the cells in the developing embryo continue to grow, divide, and move, a mutually reinforcing signal between the *Wingless* expressing cells and the *Engrailed* expressing cells maintains narrow stripes of their expression (see Figure 21–21). After three cell cycles, newly expressed regulators stabilize an *Engrailed* expression pattern that will last

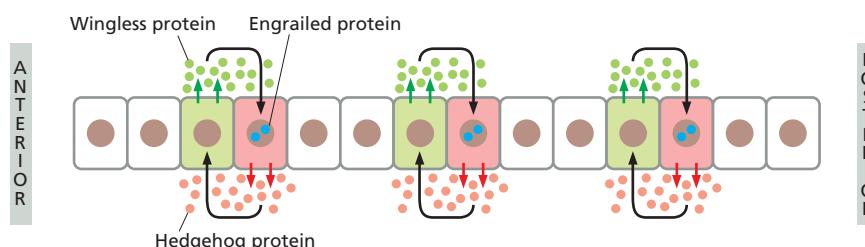


Figure 21–21 Mutual maintenance of *Hedgehog* and *Wingless* expression. *Engrailed* is a transcription regulator (blue) that drives the expression of *Hedgehog*. *Hedgehog* encodes a secreted protein (red) that activates its signaling pathway in neighboring cells and thereby drives them to express the *Wingless* gene. In turn, *Wingless* encodes a secreted protein (green) that acts back on neighbors of the *Wingless*-expressing cell to maintain their expression of *Engrailed* and *Hedgehog*. As indicated, the same control loop repeats along the A-P axis of the fly. (Based on S. Dinardo et al., *Curr. Opin. Genet. Dev.* 4:529–534, 1994.)

throughout the life of the fly, long after the signals that induced and refined it have disappeared. The segment borders will form at the posterior edge of each such *Engrailed* stripe (Figure 21–22).

In addition to regulating the segment-polarity genes, the products of pair-rule genes collaborate with those of gap genes to induce the precisely localized activation of a further set of genes—originally called *homeotic selector genes* and now often called *Hox genes*, for reasons that will become clear shortly. It is the *Hox* genes that permanently distinguish one segment from another. In the next section, we examine these important genes in detail and consider their role in cell memory; we shall see that this role is critical in a wide range of animals, including ourselves.

Hox Genes Permanently Pattern the A-P Axis

As animal development proceeds, the body becomes more and more complex. But 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 questions: 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 in different animals? The segments of the insect body provide a good example. We have thus far sketched the way in which the rudiment of a single body segment is constructed and how cells within each segment become different from one another. We now consider how one segment becomes determined, or *specified*, to be different from another.

The first glimpse of the answer to this problem came over 80 years ago, with the discovery of a set of mutations in *Drosophila* that cause bizarre disturbances in the organization of the adult fly. In the *Antennapedia* mutant, for example, legs sprout from the head in place of antennae, whereas in the *Bithorax* mutant, portions of an extra pair of wings appear where normally there should be the much smaller appendages called halteres (Figure 21–23). These mutations transform parts of the body into structures appropriate to other positions, and they are called *homeotic* mutations (from the Greek “*homoios*,” meaning similar) because the transformation is between structures of a recognizably similar general type, changing one kind of limb, or one kind of segment, into another. It was eventually discovered that a whole set of genes, the **homeotic selector genes**, or **Hox genes**, serve to permanently specify the A-P characters of the whole set of animal segments. These genes are all related to one another as members of a multigene family.

There are eight *Hox* genes in the fly, and they all lie in one or the other of two gene clusters known as the ***Bithorax complex*** and the ***Antennapedia complex***.

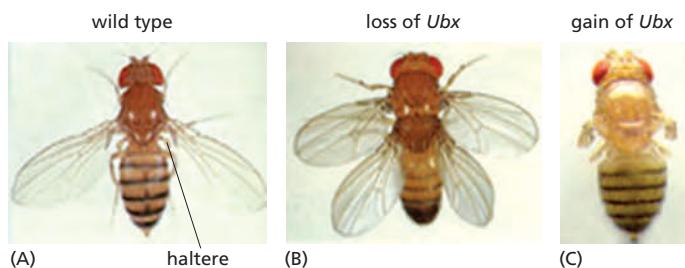


Figure 21–22 The pattern of expression of *Engrailed*, a segment-polarity gene. The *Engrailed* pattern is shown in a 10-hour embryo and an adult (whose wings have been removed in this preparation). The pattern is revealed by constructing a strain of *Drosophila* containing the control sequences of the *Engrailed* gene coupled to the coding sequence of the reporter *LacZ*, whose product is detected histochemically through the brown product generated by immunohistochemistry against *LacZ* (10-hour embryo) or through the blue product generated by a reaction that *LacZ* catalyzes (adult). Note that the *Engrailed* pattern, once established, is preserved throughout the animal's life. (Courtesy of Tom Kornberg.)

Figure 21–23 Homeotic mutations. *Ultrabithorax*, or *Ubx*, is one of three genes in the *Bithorax* gene complex (a *Hox* gene cluster). *Ubx* is responsible for all of the differences between the second and third thoracic segments. (A, B) *Ubx* loss-of-function mutations transform the haltere-bearing segment (A) into a wing-bearing segment, resulting in four-winged flies (B). (C) *Ubx* gain-of-function in the second thoracic segment transforms this wing-bearing segment into a haltere-bearing segment, resulting in wingless flies. (Courtesy of Richard Mann.)

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.

The products of the *Hox* genes, the **Hox proteins**, are transcription regulators, all of which possess a highly conserved, 60-amino-acid-long DNA-binding *homeodomain* (see p. 376). The corresponding motif in the DNA sequence is called a "homeobox," from which, by abbreviation, the *Hox* complex takes its name. There are many homeobox-containing genes, but only those located in a *Hox* complex are *Hox* genes.

Hox Proteins Give Each Segment Its Individuality

The Hox proteins can be viewed as molecular address labels possessed by the cells of each segment: these labels give the cells in each region a **positional value**—that is, an intrinsic character that differs according to a cell's location. If the address labels in a developing *Drosophila* segment are changed, the segment behaves as though it were located somewhere else; if all the *Hox* genes in an embryo are deleted, the body segments in the larva will all be alike.

To a first approximation, each *Hox* gene is normally expressed in those regions that develop abnormally when that gene is mutated or absent. How does each Hox protein give a segment its permanent identity? All the Hox proteins are similar in their DNA-binding regions, but they are very different in the regions that interact with the other proteins with which the Hox proteins form transcriptional regulatory complexes. The different protein partners act together with the Hox proteins to dictate which DNA binding sites will be recognized, as well as whether the effect on transcription at those sites will be activation or repression. Acting in this way, the Hox proteins modulate the actions of many other transcription regulators. Hundreds of genes are under this type of Hox-modulated control, including genes for cell-cell signaling, transcriptional regulation, cell polarity, cell adhesion, cytoskeletal function, cell growth, and cell death, all conspiring (in ways that are not yet understood) to give each segment its distinctive Hox-dependent character.

Hox Genes Are Expressed According to Their Order in the Hox Complex

How, then, is the expression of the *Hox* genes themselves regulated? The coding sequences of the eight *Hox* genes in the *Antennapedia* and *Bithorax* complexes in *Drosophila* are interspersed amid a much larger quantity of regulatory DNA. This DNA includes binding sites for the products of the egg-polarity and segmentation genes, thereby serving as an interpreter of the multiple items of spatial information supplied to it by all these transcription regulators. The net result is that the particular set of *Hox* genes transcribed is appropriate for each location along the A-P body axis.

The pattern of *Hox* gene expression exhibits a remarkable regularity that suggests an additional form of control. 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 A-P axis of the body (Figure 21–24). This hints at some process of gene activation, perhaps dependent on chromatin structures that propagate along the *Hox* complexes, switching on one *Hox* gene after another according to their order along the chromosome. The most "posterior" of the *Hox* genes that are expressed in a cell generally dominates, driving down expression and activity of the "anterior" genes and dictating the character of the segment. The gene regulatory mechanisms

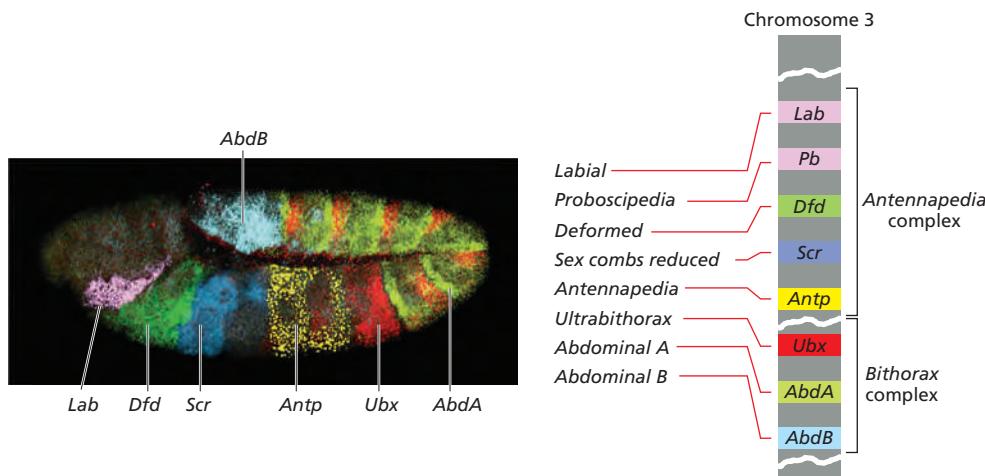


Figure 21–24 The patterns of expression compared to the chromosomal locations of the genes of the *Hox* complex. The diagram shows the sequence of genes in each of the two subdivisions of the chromosomal complex. This corresponds, with minor deviations, to the spatial sequence in which the genes are expressed, shown in the photograph of a *Drosophila* embryo at the so-called germ band retraction stage, about 10 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.)

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 animal evolution.

Triphorax and Polycomb Group Proteins Enable the *Hox* Complexes to Maintain 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* complexes serve to stamp each cell and its progeny with a permanent record of the A-P 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 A-P axis of the body. This memory trace is somehow imprinted on the *Hox* complexes, and it governs the segment-specific identity not only of the larval segments, but also of the structures of the adult fly.

The molecular mechanism of this memory of positional information relies on two types of regulation. One is from the *Hox* genes themselves: many of the *Hox* proteins autoactivate the transcription of their own genes, thereby helping to keep the genes on indefinitely. Another crucial input is from two large, complementary sets of proteins, called the **Triphorax group** and the **Polycomb group**, which stamp the chromatin of the *Hox* complex with a heritable record of its embryonic state of activation or repression. These are key general regulators of chromatin structure that can be shown to be critical for cell memory: if genes of the *Triphorax* or *Polycomb* group are defective, the pattern of expression of the *Hox* genes is set up correctly at first, but it is not correctly maintained as the embryo grows older.

The two sets of regulators act in opposite ways. Triphorax 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 21–25). How such changes in chromatin can store developmental cell memory is discussed in Chapters 4 and 7.

The D-V Signaling Genes Create a Gradient of the Transcription Regulator Dorsal

As with the patterning along the *Drosophila* A-P axis just discussed, the patterning along the dorsoventral (D-V) axis begins with maternal gene products that define

Figure 21–25 The role of genes of the *Polycomb* group. (A) Photograph of a wild-type *Drosophila* embryo. (B) 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. 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. (From G. Struhl, *Nature* 293:36–41, 1981. With permission from Macmillan Publishers Ltd.)

this axis in the egg (see Figure 21–17), and it then progresses through zygotic gene products that further subdivide the D-V axis in the embryo.

Initially, a protein that is produced by follicle cells underneath the future ventral region of the embryo leads to the localized activation of a transmembrane receptor, called **Toll**, on the ventral side of the egg membrane. The various maternal genes required for this process are called *D-V egg-polarity genes*. (Curiously, *Drosophila* Toll and vertebrate Toll-like proteins also operate in innate immune responses, as discussed in Chapter 24). The localized activation of Toll controls the distribution of **Dorsal**, a transcription regulator of the NF κ B family discussed in Chapter 15. The Toll-regulated activity of Dorsal, like that of NF κ B, depends on the translocation of Dorsal from the cytosol, where it is held in an inactive form, to the nucleus, where it regulates gene expression (see Figure 15–62). In the newly laid egg, both Dorsal mRNA and protein are distributed uniformly in the cytosol. After the nuclei in the syncytial blastoderm have migrated to the surface of the embryo, but before cellularization (see Figure 21–15), Toll receptor activation on the ventral side induces a remarkable redistribution of the Dorsal protein. On the dorsal side, the protein remains in the cytosol, but ventrally it becomes concentrated in the nuclei, with a smooth gradient of nuclear localization between these two extremes (Figure 21–26).

Once inside the nucleus, the Dorsal protein acts as a morphogen and turns on or off the expression of different sets of genes depending on Dorsal's 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 transcription regulators. In this way, the regulatory DNA interprets the positional signal provided by the nuclear Dorsal protein gradient, so as to define a D-V series of territories—distinctive bands of cells that run the length of the embryo. Most ventrally—where the nuclear concentration of Dorsal protein is highest—it switches on, for example, the expression of a gene called *Twist*, which is specific for mesoderm. Most dorsally, where the nuclear concentration of Dorsal protein is lowest, the cells switch on a gene called *Decapentaplegic* (*Dpp*). And in an intermediate region, where the nuclear 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 21–27A).

Products of the genes directly regulated by the Dorsal protein generate in turn more local signals, which define finer subdivisions along the D-V axis. These signals act during cellularization and take the form of conventional extracellular

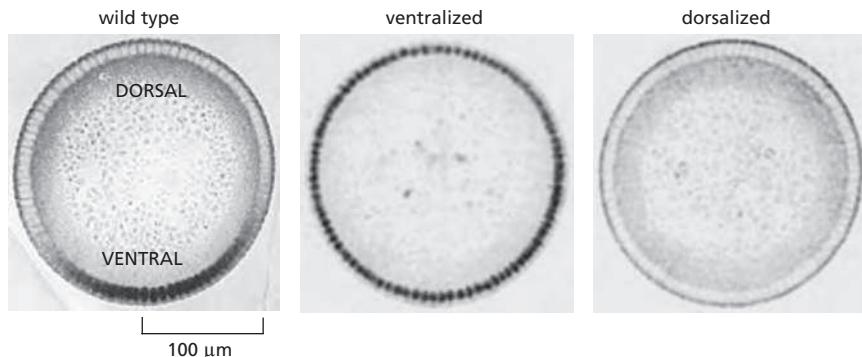
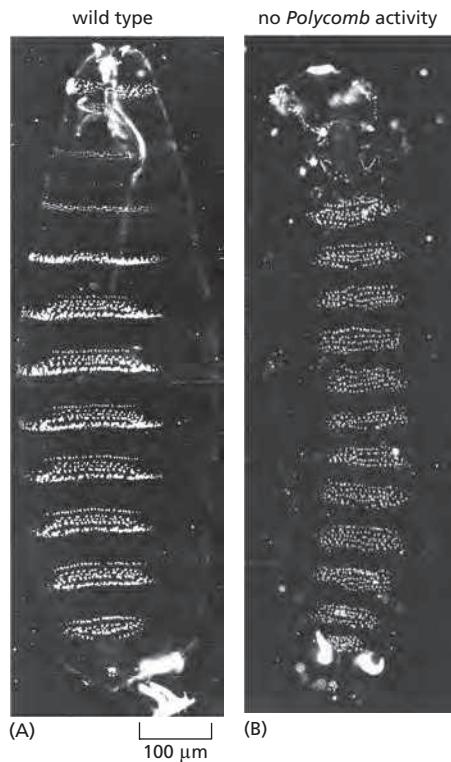


Figure 21–26 The concentration gradient of Dorsal protein in the nuclei of the blastoderm. In wild-type *Drosophila* embryos, the protein is present in the dorsal cytoplasm and absent from the dorsal nuclei; ventrally, it is depleted in the cytoplasm and concentrated in the nuclei. In a mutant in which the Toll pathway is activated everywhere and not just ventrally, Dorsal protein is everywhere concentrated in the nuclei; the result is a ventralized embryo. Conversely, in a mutant in which the Toll signaling pathway is inactivated, Dorsal protein everywhere remains in the cytoplasm and is absent from the nuclei; the result is a dorsalized embryo. (From S. Roth, D. Stein and C. Nüsslein-Volhard, *Cell* 59:1189–1202, 1989. With permission from Elsevier.)

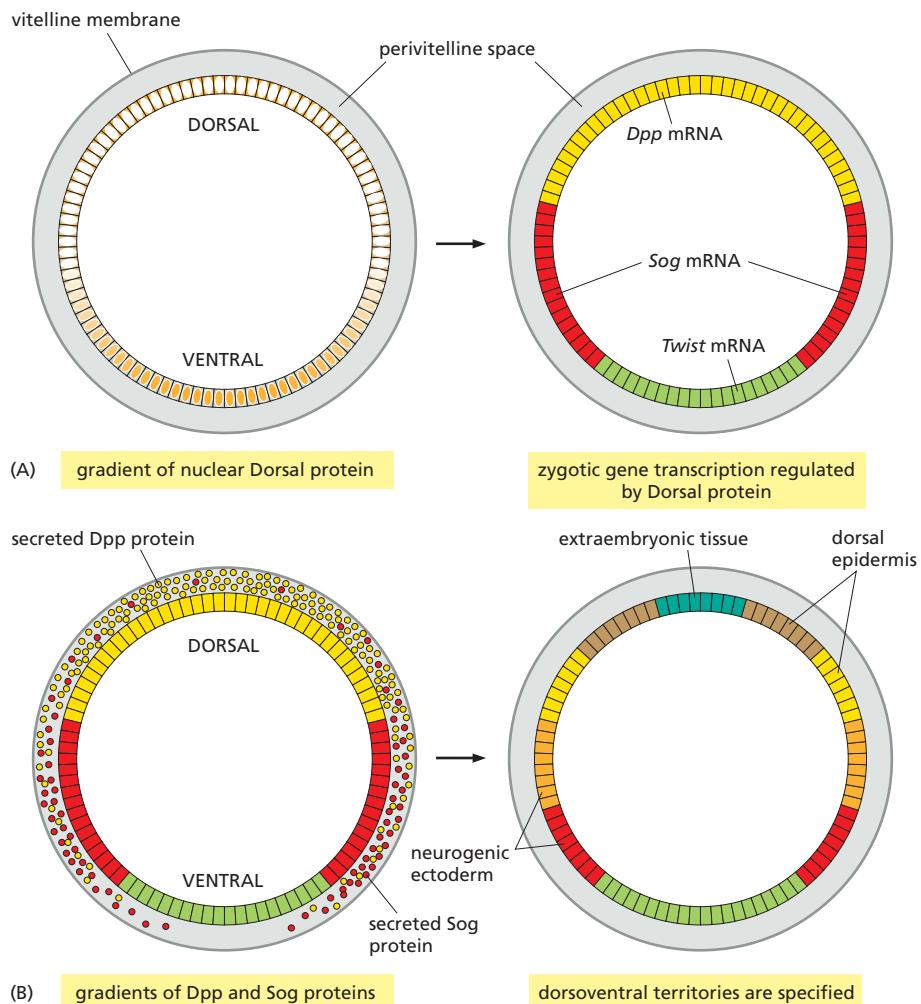


Figure 21–27 How morphogen gradients guide a patterning process along the dorsoventral axis of the *Drosophila* embryo. (A) Initially, a gradient of Dorsal protein defines three broad territories of gene expression, marked here by the expression of three representative genes—*Dpp*, *Sog*, and *Twist*. (B) Slightly later, the cells expressing *Dpp* and *Sog* secrete, respectively, the signal proteins Dpp (a TGF β family member) and Sog (an antagonist of Dpp). These two proteins then diffuse and interact with one another (and with certain other factors) to create the dorsoventral (D–V) territories shown.

signal proteins. In particular, *Dpp* codes for a secreted TGF β -family protein, which forms a local morphogen gradient in the dorsal part of the embryo. *Sog* encodes another secreted protein that is produced by the *neurogenic ectoderm* (which gives rise to the nervous system) and acts as an antagonist of Dpp protein. The opposing diffusion gradients of these two signal 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—an extraembryonic membrane. Intermediate levels cause development of dorsal epidermis; and the absence of Dpp activity allows the development of neurogenic ectoderm (Figure 21–27B).

A Hierarchy of Inductive Interactions Subdivides the Vertebrate Embryo

The molecular genetic analysis of *Drosophila* development has uncovered how a cascade of transcription regulators and signaling pathways subdivides the embryo. The same principle of progressive pattern refinement is used during the development of all animal embryos, including vertebrates. Remarkably, conservation is not restricted to the general strategy of pattern formation, but also extends to many of the molecules involved.

As mentioned previously, the earliest phases of vertebrate development are surprisingly variable, even between closely related species, and it is even hard to say precisely how the axes of an early fly embryo correspond to those of an early frog or mouse embryo. Nevertheless, we shall see that amid this display of evolutionary plasticity, some features of early development turn out to be highly

conserved. The same is true of later developmental stages also, often to an astonishing degree. From our own anatomy, it is obvious that we are cousins to birds and fish. But looking at molecular mechanisms, we see that we are cousins to flies and worms too.

In the following pages, we discuss how vertebrate embryos are patterned by the interplay of signaling molecules and transcription regulators. We begin by discussing the formation and patterning of the embryonic axes in amphibians, taking the frog *Xenopus* as our example. We have already broached this topic earlier in the chapter. Here, we pick up the thread and draw comparisons with the fly.

As noted earlier, the origins of the embryonic axes and the three germ layers in the frog can be traced back to the blastula (see Figure 21–3A). By labeling individual blastomeres, we can track cells through all their divisions, transformations, and migrations and see what they become and where they come from. The precursors of ectoderm, mesoderm, and endoderm are arranged in order along the animal-vegetal axis of the blastula: the endoderm derives from the most vegetal blastomeres, the ectoderm from the most animal, and the mesoderm from a middle set. Within each of these territories, the cells have diverse fates according to their positions along the D-V axis of the later embryo. For ectoderm, epidermal precursors are located ventrally, and future neurons are found dorsally; for mesoderm, precursors for notochord, muscle, kidney, and blood are arranged from dorsal to ventral. All this can be represented by a **fate map** that shows which cell types derive from which regions of the blastula (Figure 21–28). The fate map confronts us with the central question: how are the cells in different positions driven toward their different fates? We have already explained how maternal factors deposited in the developing frog egg define its animal-vegetal axis, and how cortical rotation triggered by fertilization defines the orientation of the dorsoventral axis (see Figure 21–14). But how does the establishment of axes lead on to the subdivision of the embryo into the future body parts?

The maternal gene products lead to the formation of signaling centers on the vegetal and dorsal sides of the embryo. The dorsal signaling center in particular has a special place in the history of developmental biology. Experiments in the early twentieth century identified it as a small cluster of cells, located on the dorsal side of the amphibian embryo, with an extraordinary property: when the cells were transplanted to an opposite site, they could trigger a radical reorganization of the neighboring tissue, causing it to form a second whole-body axis (Figure 21–29). The discovery of this signaling center, called the **Organizer**, led the way to a pioneering analysis of the chain of inductive interactions that establish the framework of the vertebrate body.

In contrast to the *Drosophila* syncytial embryo, the fertilized frog egg undergoes rapid cleavage divisions that result in an embryo consisting of thousands of cells. Patterning must therefore be mediated by extracellular signal molecules

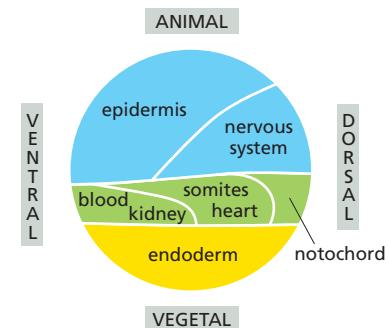


Figure 21–28 Blastula fate map in a frog embryo. The endoderm derives from the most vegetal blastomeres (yellow), the ectoderm from the most animal (blue), and the mesoderm from a middle set (green) that contributes also to endoderm and ectoderm. Different cell types derive from different positions along the dorsoventral axis.

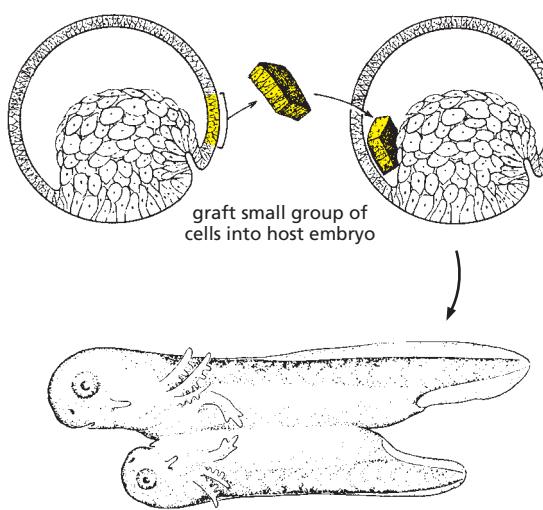


Figure 21–29 Induction of a secondary axis by the Organizer. An amphibian embryo receives a graft of a small cluster of cells taken from a specific site, called the Organizer region, on the dorsal side of another embryo at the same stage. Signals from the graft organize the behavior of neighboring cells of the host embryo, causing development of a pair of conjoined (Siamese) twins. See Movie 21.4. [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.]

that diffuse through the embryo from cell to cell, not by transcription regulators that move through the cytoplasm of a syncytium. Not surprisingly, the Organizer is now known to be a major source of secreted protein signals.

A Competition Between Secreted Signaling Proteins Patterns the Vertebrate Embryo

The signal molecules that pattern the frog embryo along the animal-vegetal (A-V) axis belong to the TGF β family: they are secreted by a signaling center at the vegetal pole and form concentration gradients along the A-V axis. The *Nodal* protein acts over a relatively short range: cells near the vegetal pole are exposed to high levels of it and respond by switching on genes that promote the development of endoderm; cells further away are exposed to lower levels and activate genes that promote the formation of mesoderm. The cells at the vegetal pole that produce Nodal also produce a more rapidly diffusing TGF β -like protein called *Lefty*, which antagonizes Nodal. The result is a high ratio of Lefty to Nodal at the animal pole, where Lefty predominates and Nodal signaling is blocked; this causes the cells there to develop as ectoderm (**Figure 21–30A**). Thus, a mid-range activation by Nodal, combined with a long-range inhibition by Lefty, sets up the pattern of progenitors along the A-V axis for the three germ layers—endoderm, mesoderm, and ectoderm.

The frog's dorsal signaling system uses a different set of secreted signals from that of the vegetal signaling system to subdivide the germ-layer territories according to location along the D-V axis of the embryo. It exerts its influence by secreting two inhibitory signal proteins, called *Chordin* and *Noggin*. These antagonize the action of *bone morphogenetic proteins* (BMPs; members of yet another subclass of the TGF β family), which themselves are secreted throughout the embryo. In this way, Chordin and Noggin form a dorsal-to-ventral gradient that blocks BMP signaling on the dorsal side but allows it to remain high on the ventral side (Figure 21–30B). Ectodermal cells that experience high levels of BMP signaling are driven to epidermal fates, whereas cells that experience little or no BMP signaling remain neural.

Knowing the signals that specify the three germ layers and various tissue types of the vertebrate body, one can reproduce this specification in a culture dish. Frog cells taken from the animal-pole region of the embryo, for example,

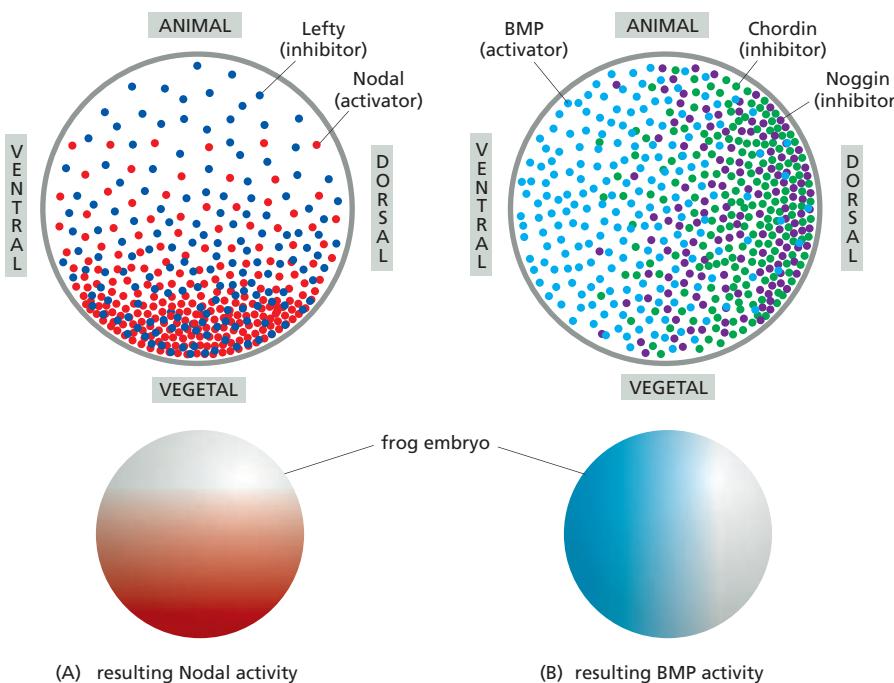


Figure 21–30 How Nodal and bone morphogenic protein (BMP) signaling pattern the embryonic axes. Nodal and its antagonist Lefty pattern the animal-vegetal axis, while BMP and its antagonists Chordin and Noggin pattern the dorsoventral axis. (A) In the animal pole region, where Nodal levels are low relative to Lefty, Lefty blocks Nodal from binding to its receptors. In the vegetal region, there is an excess of Nodal, resulting in Nodal pathway activation. (B) Along the dorsoventral axis, BMP is widely present but Chordin and Noggin are concentrated at the dorsal side: there, they bind to BMP and block its binding to receptors. The resulting patterns of Nodal and BMP activity are illustrated at the bottom of the figure.

will differentiate into blood (a ventral mesodermal tissue) when diverted from their original fate by exposure to intermediate concentrations of Nodal and high concentrations of BMP. Similarly, mouse or human embryonic stem cells can be coaxed to generate specific cell types by exposing them in culture to appropriate combinations of signal molecules. In this way, the insights gained through studies of animal development can be used to generate the cell types needed for regenerative medicine, as we discuss in the next chapter.

The Insect Dorsoventral Axis Corresponds to the Vertebrate Ventral-Dorsal Axis

The signaling systems that pattern the D-V axis in *Drosophila* and in vertebrates are similar. In *Drosophila*, as we saw, Dpp and its inhibitor Sog are responsible, whereas in vertebrates, BMP and its inhibitors Chordin and Noggin do the job. Dpp is a member of the BMP family, while Sog is a homolog of Chordin. Both in flies and frogs, high levels of the inhibitors define the region that is neurogenic, and high levels of BMP/Dpp activity define the region that is not. These and other molecular parallels strongly suggest that this aspect of body patterning has been conserved in evolution from insects to vertebrates. Curiously, however, the axis is inverted: dorsal in the fly corresponds to ventral in the vertebrate (Figure 21–31). At some point in evolution, it seems that the ancestor of one of these classes of animals took to living life upside-down.

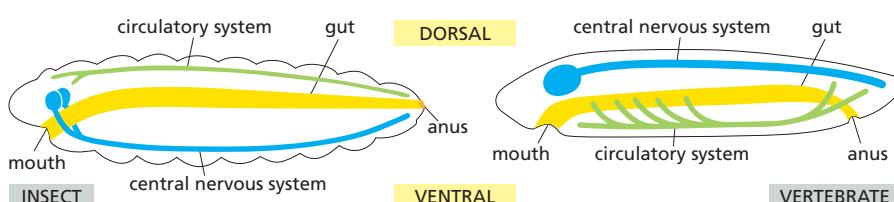
Hox Genes Control the Vertebrate A-P Axis

The conservation of developmental mechanisms between *Drosophila* and vertebrates extends beyond the D-V signaling system. *Hox* genes are found in almost every animal species studied, where they are often grouped in complexes similar to the insect *Hox* complex. In mice and humans, for example, 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 one complete insect *Hox* complex (that is, an *Antennapedia* complex plus a *Bithorax* complex) (Figure 21–32).

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 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, 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*. As in *Drosophila*, vertebrate *Hox* gene expression patterns are often aligned with vertebrate segments. This alignment is especially clear in the hindbrain (see Figure 21–32), where the segments are called *rhombomeres*.

The products of the vertebrate *Hox* genes, the *Hox* proteins, specify positional values that control the A-P 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

Figure 21–31 The vertebrate body plan as a dorsoventral inversion of the insect body plan. 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. Unlike vertebrates, insects have 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.)



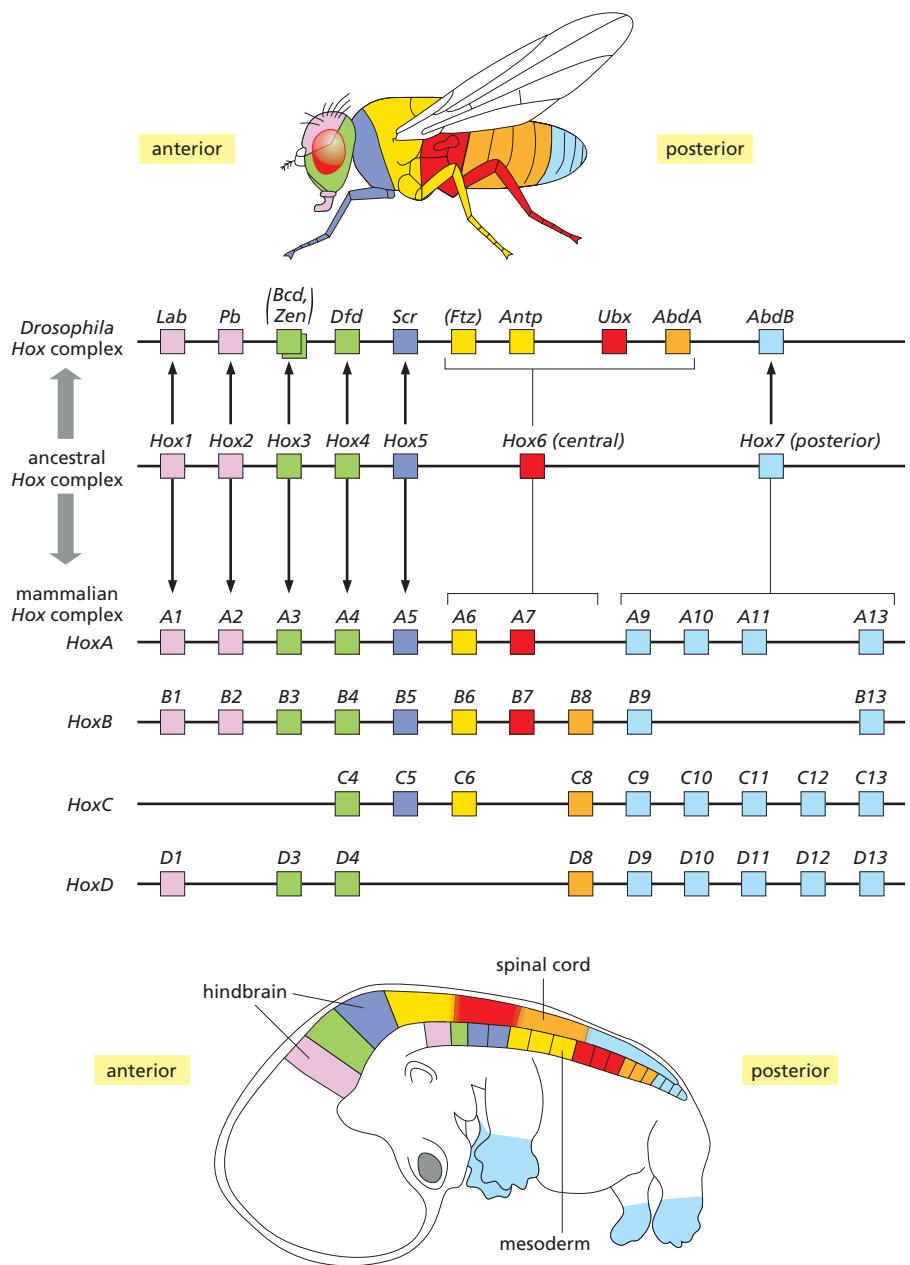


Figure 21–32 The *Hox* complexes of an insect and a mammal, compared and related to body regions. The genes of the *Antennapedia* and *Bithorax* complexes of *Drosophila* are shown in their chromosomal order in the top line. The corresponding genes of the four mammalian *Hox* complexes are shown below, 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. There is a remarkable parallelism. 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.

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 and others lost. Still others have been co-opted for different purposes (genes in parentheses in the top line) over the time that has elapsed since the complexes diverged. (Based on a diagram courtesy of William McGinnis.)

a posterior character. Conversely, loss of posterior *Hox* genes allows the posterior tissue where they are normally expressed to adopt an anterior character (Figure 21–33). Because of a redundancy between genes in the four *Hox* gene clusters, the transformations observed in mouse *Hox* mutants are not always so straightforward as those in the fly, and they are often incomplete. Nonetheless, it seems clear that the fly and the mouse use essentially the same molecular machinery to impart individual characteristics to successive regions along at least a part of the A-P axis.

Some Transcription Regulators Can Activate a Program That Defines a Cell Type or Creates an Entire Organ

Just as there are genes that regulate pattern formation and segmental identity, there are genes whose products act as triggers for the development of a specific cell type or even a specific organ, initiating and coordinating the whole complex program of gene expression that is required. An example is the *MyoD/myogenin*

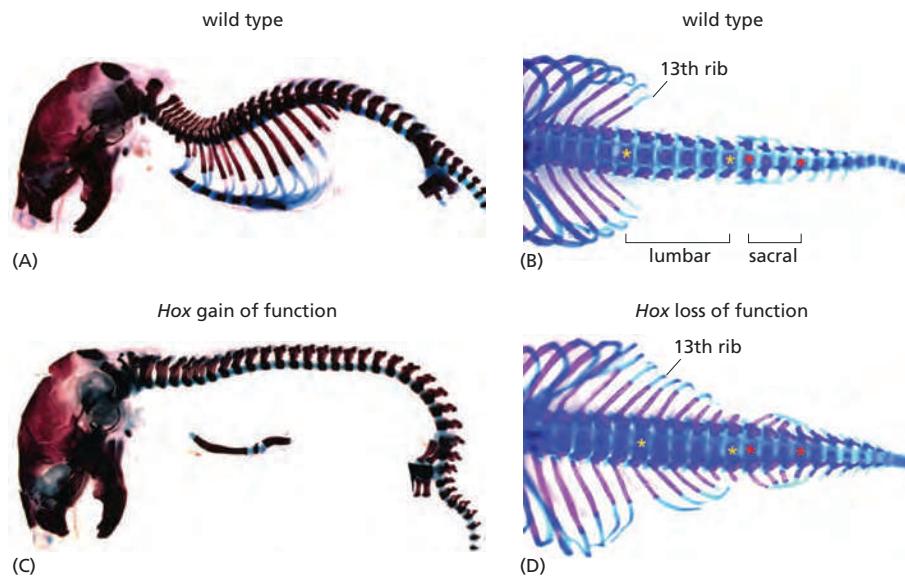


Figure 21–33 Control of anteroposterior pattern by *Hox* genes in the mouse.

(A,B) A normal mouse (wild type) 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 and (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 removed 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.)

family of transcription regulators that we encountered in Chapter 7. These proteins drive cells to differentiate into muscle, expressing muscle-specific actins and myosins and all the other specialized cytoskeletal, metabolic, and membrane proteins that a muscle cell needs. Analogously, members of the Achaete/Scute family of transcription regulators drive cells to become neural progenitors. In both these examples, the proteins belong to the basic helix-loop-helix (bHLH) class of transcription regulators (see p. 377), and the same is true for many of the other proteins that induce the differentiation of particular cell types. These *master transcription regulators* exert their powerful differentiation-inducing activity by binding to many different regulatory sites in the genome and thereby controlling the expression of large numbers of downstream target genes. In one well-studied case, that of an Achaete/Scute family member called Atonal homolog 1 (*Atoh1*), the number of direct target genes in the mouse genome is more than 600. It is important to note, however, that even such powerful drivers of cell differentiation can have radically different effects according to the context and history of the cells in which they act: *Atoh1*, for example, drives differentiation of certain classes of neurons in the brain, of sensory hair cells in the inner ear, and of secretory cells in the lining of the gut.

Other genes encoding transcription regulators can drive the formation and assembly of the multiple cell types that constitute an entire organ. A famous example is the transcription regulator *Eyeless*. When it is artificially expressed in a patch of cells in the leg precursors of *Drosophila*, a well-organized eye-like organ develops on the leg, with the various eye cell types correctly arranged (see Figure 7–35B); conversely, loss of the *Eyeless* gene results in flies that lack eyes. Moreover, loss of the *Eyeless* homolog *Pax6* in vertebrates likewise leads to loss of eye structures. Similar organ-selector proteins are known for foregut, heart, pancreas, and other organs. They are all master transcription regulators that directly regulate hundreds of target genes, the products of which then specify and construct the different elements of the appropriate organ. However, as in the example of *Atoh1*, they usually exert their specific effect only in combination with the right partners, which are only expressed in cells that were appropriately primed during their earlier development.

Notch-Mediated Lateral Inhibition Refines Cellular Spacing Patterns

After the establishment of the basic body plan and the generation of organ precursors, many further steps of pattern refinement are required to achieve the adult pattern of terminally differentiated cells in tissues and organs. As we discussed

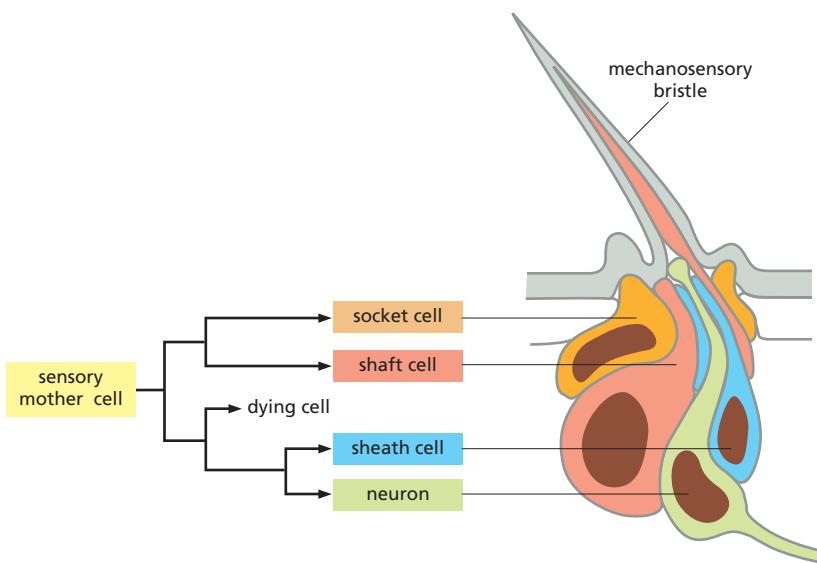


Figure 21–34 The basic structure of a mechanosensory bristle. The lineage of the four cells of the bristle—all descendants of a single sensory mother cell—is shown on the left. The sensory mother cell, once it is specified, generates this set of cells through a short program of division cycles. In each generation of the progeny, lateral inhibition operates again to drive the newborn cells toward different fates: one of the ultimate progeny will become the neuron; another, the shaft of the bristle; others, supporting cells of various sorts. As the sensory mother cell and its progeny divide, certain proteins are allocated preferentially to one of each pair of newborn sister cells, biasing the outcome of the lateral-inhibition competition mediated by Notch signaling.

earlier, lateral inhibition mediated by Notch signaling is crucial for both cell diversification and fine-grained patterning in an enormous variety of tissues in all animals.

One example is the development of **sensory bristles** in *Drosophila*, most easily seen on the fly's back, but also present on most of its other exposed surfaces. Each of these is a miniature sense organ, consisting of a sensory neuron and a small set of supporting cells. Some bristles respond to chemical stimuli, others to mechanical stimuli, but they are all constructed in a similar way (Figure 21–34). The proneural genes *Achaete* and *Scute* mentioned earlier mark the patches of epidermis 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 abnormal sites cause bristles to develop there.

The initial cells expressing the proneural genes are called proneural cells, and they are primed to take the neurosensory pathway of differentiation, but which of the cells will actually do so depends on competitive interactions among them. In the first round of these interactions, a single cell within each small group of proneural cells is picked to serve as the progenitor of the bristle. This single cell is called the *sensory mother cell*. It becomes distinct from the other cells of the cluster through lateral inhibition mediated by the Notch signaling pathway. This operates in the way we discussed earlier. The cells in the proneural cluster initially all express both the transmembrane receptor Notch and its transmembrane ligand *Delta*, along with proteins that regulate the signaling activity of Delta. Wherever Delta activates Notch, an inhibitory signal is transmitted that diminishes the tendency of the Notch-activated cell to specialize as a sensory mother cell. At first, all the cells in the cluster inhibit one another. However, receipt of the signal in a given cell diminishes that cell's ability to fight back by delivering the inhibitory Delta signal in return. This creates a competitive situation, from which a single cell in each cluster—the future sensory mother cell—eventually emerges as winner, sending a strong inhibitory signal to its immediate neighbors but receiving no such signal in return (Figure 21–35). If a cell that would normally become a sensory mother cell is genetically disabled from doing so, a neighboring proneural cell, freed from lateral inhibition, will become a sensory mother cell instead.

The sensory mother cell goes through a short program of further divisions to generate the set of cells that form the final bristle. Notch signaling acts repeatedly at successive stages in this program to drive the descendants of the sensory mother cell along different pathways and assign them to their various specialized fates. However, it does so in conjunction with additional mechanisms that bias the outcome of the competition mediated by lateral inhibition. Determinants that

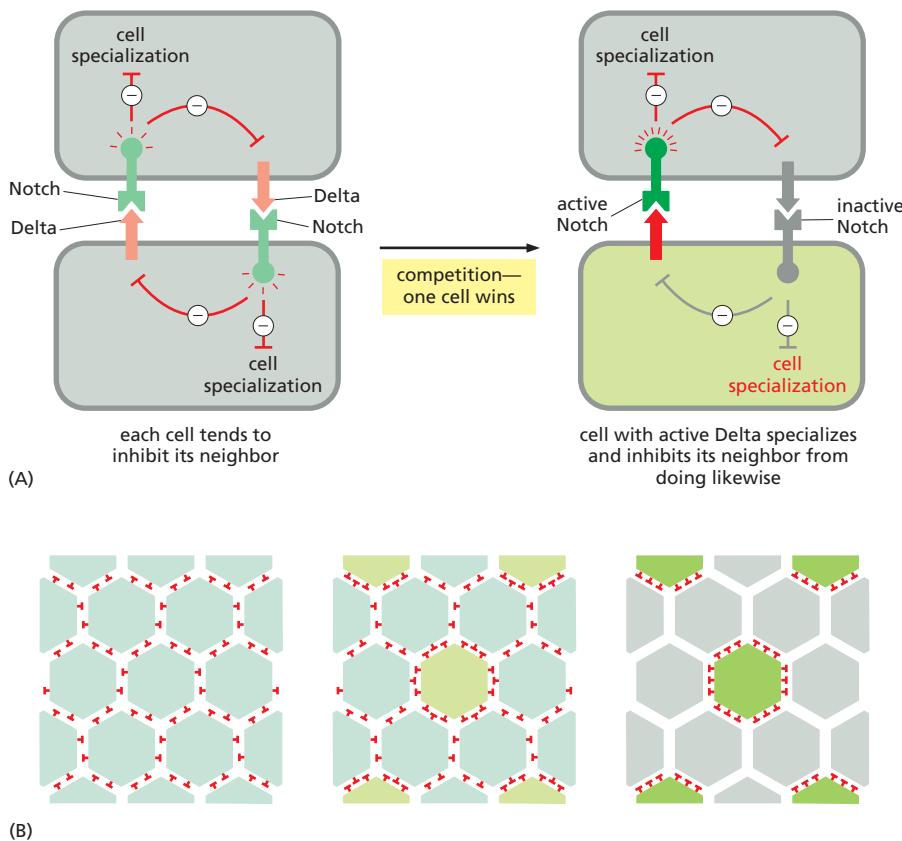


Figure 21–35 Lateral inhibition. (A) The basic mechanism of Notch-mediated competitive lateral inhibition, illustrated for just two interacting cells. In this diagram, the absence of color on proteins or effector lines indicates inactivity. (B) The outcome of the same process operating in a larger patch 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 cell, 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.

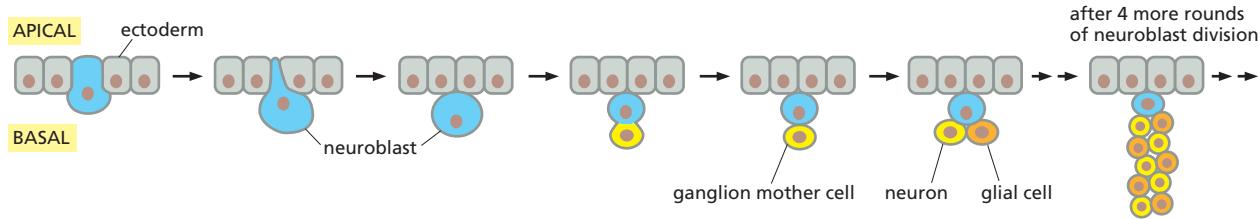
are asymmetrically localized inside the dividing cells have this role in sensory bristle development. They are also important in other contexts, as we now discuss.

Asymmetric Cell Divisions Make Sister Cells Different

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, during which some significant set of molecules is divided unequally between them. 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 (see Figure 21–12). We have already encountered the asymmetric segregation of molecules in the context of the early frog embryo: *VegT* RNA is localized in the vegetal region of the fertilized egg. Following cell division, only vegetal daughter cells will inherit *VegT* RNA.

Asymmetric divisions often occur at the beginning of development, but they are also encountered at some later stages. As mentioned for the sensory bristle, they can set the scene for an exchange of Notch signals between the daughter cells, with the signaling occurring after the cells have become separate and reinforcing the differences between them. In the central nervous system, asymmetric divisions have a key role in generating the very large numbers of neurons and glial cells that are needed. A special class of cells becomes 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 process is best understood in *Drosophila*, although there are many hints that something similar occurs also in vertebrate neurogenesis.

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 21–36). At each division, one daughter remains as a neuroblast, while the other, which is much



smaller, becomes specialized as a *ganglion mother cell*. Each 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, with Notch-mediated interactions helping to drive the daughters along different paths. The neuroblast itself 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, but now they are accompanied by cell growth, permitting the process to continue indefinitely and to generate the much larger numbers of neurons and glial cells required in the adult fly.

Differences in Regulatory DNA Explain Morphological Differences

In the preceding sections, we have seen that animals contain the same essential cell types, have a similar collection of genes, and share many of the molecular mechanisms of pattern formation. But how can we square this with the radical differences that we see in the body structures of animals as diverse as a worm, a fly, a frog, and a mouse? We asserted earlier, in a general way, that these differences usually seem to reflect differences in the regulatory DNA that calls into play the components of the conserved basic kit of parts. We must now examine the evidence a little more closely.

When we compare animal species with similar basic body plans—different vertebrates, for example, such as fish, birds, and mammals—we find that corresponding genes usually have similar sets of regulatory elements: the regulatory DNA sequences have been well conserved and are recognizably homologous in the different animals. The same is true if we compare different species of nematode worms or insects. But, when we compare vertebrate regulatory regions with those of worms or flies, it is hard to see any such resemblance. The protein-coding sequences are unmistakably similar, but the corresponding regulatory DNA sequences appear mostly very different, suggesting that the differences in body plans mainly reflect differences in regulatory DNA. Although variations in the proteins themselves also contribute, differences in regulatory DNA would be enough to generate radically different tissues and body structures even if the proteins were the same.

It is not yet possible to trace the genetic steps that have led to all the spectacular diversity of animals. Their lineages have diverged over hundreds of millions of years, and in most cases too many changes have occurred for us to be able to say that this or that feature results from this or that mutation. The picture is clearer, however, for more recent evolutionary events. Studies of both closely related animal populations and plant populations whose members have different morphologies have revealed that dramatic developmental effects can result from subtle changes in regulatory DNA.

A well-studied example is the morphological diversity found in stickleback fish. After the last ice age ended about 10,000 years ago, marine sticklebacks colonized many newly formed freshwater streams and lakes. Marine sticklebacks extend sharp spines from their pelvic skeleton. These spines are thought to help protect the fish from soft-mouthed fish predators. In contrast, several populations of freshwater sticklebacks have lost these spines, usually in lakes that lack such predators. The different morphologies reflect differences in control of the expression of a transcription regulator called *Pitx1*. Whereas marine sticklebacks express the *Pitx1* gene in the pelvic bone precursor cells that will form the spikes,

Figure 21–36 Neuroblasts and asymmetric cell division in the central nervous system of a fly embryo. 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).

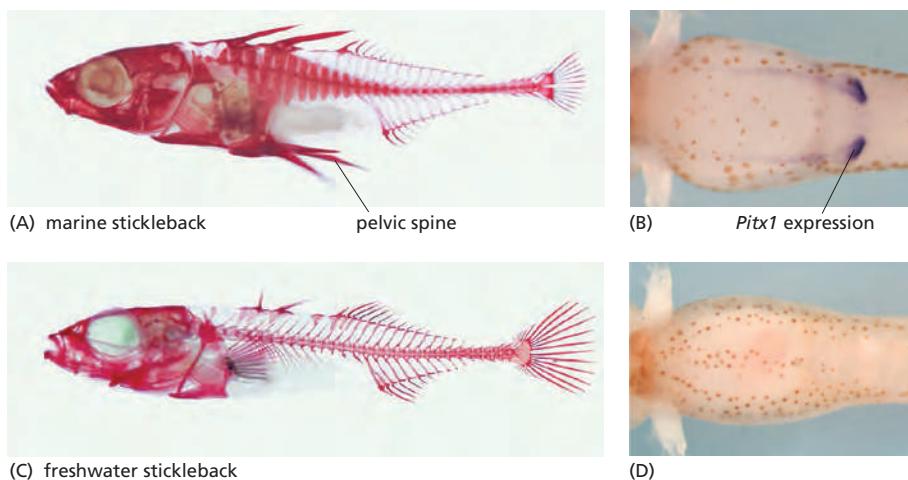


Figure 21–37 Morphological diversity in stickleback fish is caused by changes in regulatory elements.

(A–D) Pelvic spines are present in marine (A) but not in freshwater (C) populations. Correspondingly, *Pitx1* is expressed in the pelvic area in marine (B) but not in freshwater (D) fish. The lack of expression in the pelvic area of freshwater populations is caused by mutations in an enhancer element. Other enhancers and sites of expression for *Pitx1* are the same in marine and freshwater sticklebacks. (Courtesy of Michael D. Shapiro.)

freshwater sticklebacks have lost this expression as a result of a change at the *Pitx1* locus. These changes do not lie in the coding sequence. Instead, each is a small deletion of a block of adjacent regulatory DNA that controls *Pitx1* expression specifically in the pelvic cells (Figure 21–37).

The *Pitx1* protein has important functions elsewhere in the body, so that the DNA sequences that encode this protein must be retained. The regulatory DNA responsible for *Pitx1* expression at these other sites is also unchanged in the two populations of sticklebacks. The evolution of pelvis development in sticklebacks shows how the modular nature of regulatory DNA elements that we encountered in Chapter 7 (see Figure 7–29) allows independent modification of the different parts of the body, even when formation of those body parts depends on the same proteins.

In the recent evolution of plants, changes of body structure can be traced in a similar way to changes in regulatory DNA. For example, these account for a large part of the dramatic difference between the wild teosinte plant and its modern descendant, maize, through some 10,000 years of mutation and selection by Native Americans.

Summary

Drosophila has been the foremost model organism for the study of the genetics of animal development. Its embryonic pattern is initiated by the products of maternal-effect genes called egg-polarity genes, which operate by setting up graded distributions of transcription regulators in the egg and early embryo. The gradient of Bicoid protein along the A-P axis, for example, helps initiate the orderly expression of gap genes, pair-rule genes, and segment-polarity genes. These three classes of segmentation genes, through a hierarchy of interactions, become expressed in some regions of the embryo and not others, progressively subdividing the embryo along the A-P axis into a regular series of repeating modular units called segments.

Superimposed on the pattern of gene expression that repeats itself in every segment, there is a serial pattern of expression of Hox genes that confer on each segment a different identity. These genes are grouped in complexes and are arranged in a sequence that matches their sequence of expression along the A-P axis of the body.

Although Hox gene expression is initiated in the embryo, it is subsequently maintained by the action of chromatin-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 repression or activation, respectively. Hox complexes homologous to that of Drosophila are found in virtually every type of animal, where they help pattern the A-P axis of the body.

Signaling gradients are also set up along the dorsoventral (D-V) axis. Initially, Toll signaling generates a nuclear gradient of Dorsal protein, which induces an extracellular signaling gradient of the TGF β -family protein Dpp and its antagonist,

Sog. This creates a gradient of *Dpp* activity that helps refine the assignment of different characters to cells at different positions along the D-V axis.

In *Xenopus*, the polarity of the egg and the site of sperm entry set up the embryonic axes. A gradient generated by the TGF β -family protein *Nodal* induces different fates along the animal-vegetal axis, whereas *BMP* and *Chordin*—proteins homologous to *Drosophila Dpp* and *Sog*, respectively—control the patterning of the D-V axis. This axis is inverted, so that dorsal in the fly corresponds to ventral in the frog.

Transcription regulators control the formation of specific cell types. Members of the *MyoD/myogenin* family drive the process of muscle cell determination, coordinating the many components required, whereas *Achaete/Scute* transcription regulators control neural fate. Other genes encoding such master transcriptional regulators can regulate the formation of entire organs. *Eyeless*, for example, is both necessary and sufficient to generate eye structures in *Drosophila*.

To refine the anatomical pattern within such an organ, the cells interact locally, both by diffusible inductive signals and by short-range mechanisms. Often, the cells compete with one another by lateral inhibition. This process results in activation of the Notch signaling pathway in one cell and inhibition in its neighbors, generating two different cell types. Asymmetric cell divisions, in which daughter cells inherit different molecular determinants from the mother cell, provide an additional way to organize a fine-grained diversity of cell types.

Evidence from recent evolutionary events indicates that anatomical changes are mostly driven by changes in regulatory DNA sequences that determine when and where developmental genes are expressed. How the striking diversity in body structures has evolved over longer times remains largely unknown, although it seems likely that similar principles apply.

DEVELOPMENTAL TIMING

Developmental events unfold over minutes, hours, days, weeks, months, or even years, with each organism following its own strict timetable. The cascades of inductive interactions and transcriptional regulatory events described earlier take time, as signals are transmitted and transcription regulators are synthesized and then bind to DNA to activate or repress their target genes. At the beginning of this chapter, we compared development with an orchestral performance. There are many players, and each must do the right thing at the right time; yet there is no leader or conductor to set the tempo and coordinate the timing of all the different events. Each developmental process must thus occur at an appropriate rate, tuned by evolution to fit with the timing of other processes in the embryo or in the environment. The control of timing is one of the most important problems in developmental biology, but also one of the least understood.

Molecular Lifetimes Play a Critical Part in Developmental Timing

Developmental processes are complex, but they are built up from simple steps. A first challenge is to understand the timing of these steps. How long does it take, for example, to switch the expression of a gene on or off? This is not like throwing a light switch: it involves delays. First, it takes time to make an mRNA molecule: the RNA polymerase must travel the length of the gene, the primary RNA transcript must be spliced and otherwise processed, and the resulting mRNA must be exported from the nucleus and delivered to the site where it will be translated. This adds up to what one might call the *gestation time* of the individual molecule. Second, it takes time for the individual mRNA molecules to accumulate to their fully effective concentration; as explained in Chapter 15, this *accumulation time* is dictated by the average lifetime of the molecules—the longer they last, the higher their ultimate concentration, and the longer the time taken to attain it. Similar delays occur at the next step, where the mRNA is translated into protein: synthesis of each individual protein molecule involves a gestation delay, and attainment of an effective concentration of protein molecules involves an accumulation delay that depends on the protein’s lifetime. The time for the whole gene switching process is just the sum of the gestation delays and the accumulation delays (basically,

the molecular lifetimes) for both the mRNA and the protein molecules. Somewhat counterintuitively, it is the combined length of these delays, rather than the rate of molecular synthesis (the number of molecules synthesized per second), that chiefly determines the switching time.

The same additive principle applies to long cascades of gene switching, where gene A activates gene B, and gene B activates gene C, and so on. It also applies in other circumstances, such as in signaling pathways where one protein directly regulates the activation of the next. In all these cases, molecular lifetimes, along with gestation delays, play a key part in determining the pace of development. The lifetimes of mRNA and protein molecules are enormously variable, from a few minutes or hours to days or more, explaining much of the variation we see in the tempo of developmental events.

Gene switching delays, however, are not the be-all and end-all of developmental timing. Development involves many other kinds of delay that contribute to timing. Chromatin structure takes time to remodel. Inductive signals take time to diffuse across a field of cells (see Figure 21–9). Cells take time to move and rearrange themselves in space. Nevertheless, the timing of gene switching plays a fundamental part in developmental timing, as illustrated in an especially clear and striking way by a gene-expression oscillator that controls the segmentation of the vertebrate body axis, as we now explain.

A Gene-Expression Oscillator Acts as a Clock to Control Vertebrate Segmentation

The main body axis of all vertebrates has a repetitive, periodic structure, seen in the series of vertebrae, ribs, and segmental muscles of the neck, trunk, and tail. These segmental structures originate from the mesoderm that lies as a long slab on either side of the embryonic midline. This slab becomes broken up into a regular repetitive series of separate blocks, or **somites**—cohesive groups of cells, separated by clefts (Figure 21–38A). The somites form (as bilateral pairs) one after another, in a regular rhythm, starting in the region of the head and ending in the tail. Depending on the species, the final number of somites ranges from less than 40 (in a frog or a zebrafish) to more than 300 (in a snake).

The posterior, most immature part of the mesodermal slab, called the *presomitic mesoderm*, supplies the required cells: as the cells proliferate, this mesoderm

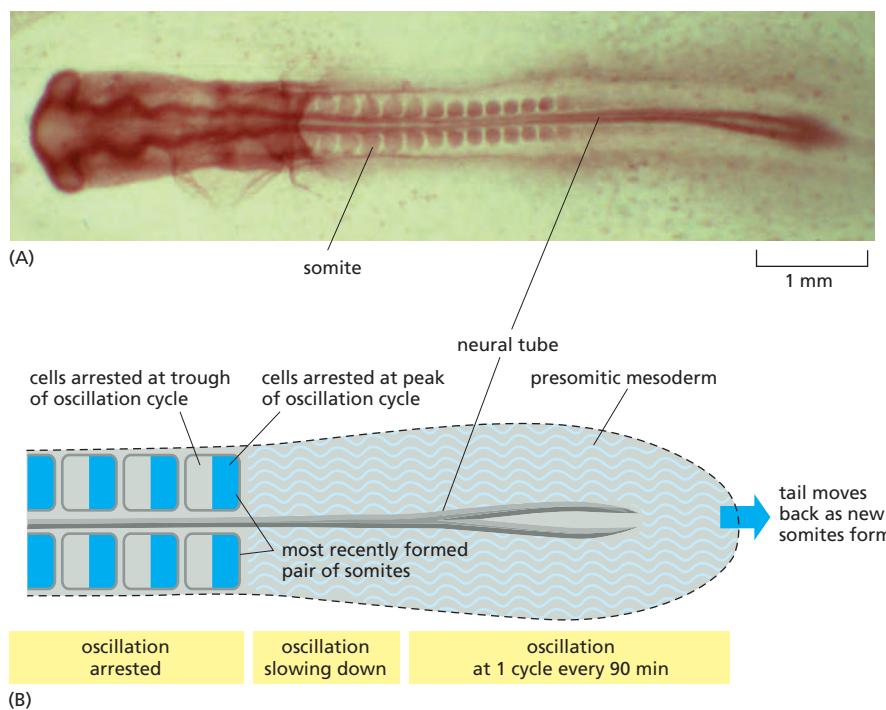


Figure 21–38 Somite formation in the chick embryo. (A) A chick embryo at 40 hours of incubation. (B) How the temporal oscillation of gene expression in the presomitic mesoderm becomes converted into a spatial alternating pattern of gene expression in the formed somites. In the posterior part of the presomitic 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.)

retreats tailward, extending the embryo (Figure 21–38B). In the process, it deposits a trail of somites formed from cells that group together into blocks as they emerge from the anterior end of the presomitic region. The special character of the presomitic mesoderm is maintained by a combination of fibroblast growth factor (FGF) and Wnt signals, produced by a signaling center at the tail end of the embryo, and the range of these signals seems to define the length of the presomitic mesoderm. The somites emerge with clocklike timing, but what determines the rhythm of the process?

In the posterior part of the presomitic mesoderm, the expression of certain genes oscillates in time. Snapshots of gene expression taken by fixing embryos for analysis at different times in the oscillation cycle reveal what is happening, and the oscillations can now also be observed in time-lapse movies of embryos containing fluorescent reporters of individual oscillating genes. One new somite pair is formed in each oscillation cycle, and, in mutants where the oscillations fail to occur, somite segmentation is disrupted: the cells may still break up, belatedly, into separate clusters, but they do so in a haphazard, irregular way. The gene-expression oscillator controlling regular segmentation is called the **segmentation clock**. The length of one complete oscillation cycle depends on the species: it is 30 minutes in a zebrafish, 90 minutes in a chick, 120 minutes in a mouse.

As cells emerge from the presomitic mesoderm to form somites—in other words, as they escape from the influence of the FGF and Wnt signals—their oscillation stops. Some become arrested in one state, some in another, according to the phase of the oscillation cycle at the time they leave the presomitic region. In this way, 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; this in turn dictates how the tissue will break up into physically separate blocks, through effects on the pattern of cell-cell adhesion (see Figure 21–38B).

How does the segmentation clock work? The first somite oscillator genes to be discovered were *Hes* genes, which are key components of the Notch signaling pathway. They are directly regulated by the activated form of Notch, and they code for inhibitory transcription regulators that inhibit the expression of other genes, including *Delta*. As well as regulating other genes, the products of *Hes* genes can directly regulate their own expression, creating a remarkably simple negative feedback loop. Autoregulation of certain specific *Hes* genes (depending on species) is thought to be the basic generator of the oscillations of the somite clock. Although the machinery has been modified in various ways in different species, the underlying principle seems to be conserved. When the key *Hes* gene is transcribed, the amount of *Hes* protein product builds up until it is sufficient to block *Hes* gene transcription; synthesis of the protein ceases; the protein then decays, permitting transcription to begin again; and so on, cyclically (Figure 21–39). The period of oscillation, which determines the size of each somite, depends on the delay in the feedback loop. This equals the sum of the gestation delays and accumulation delays (that is, the molecular lifetimes) of the *Hes* mRNA and protein molecules, according to the additive principle discussed earlier. Mathematical modeling (see Chapter 8) allows us to relate these basic molecular parameters to the cycle time of the segmentation clock: to a first approximation, the cycle period is simply equal to twice the total delay in the negative feedback loop, and thus twice the sum of the delays occurring at each step of the loop.

The feedback loop just described is intracellular, and each cell in the presomitic mesoderm can generate oscillations on its own. But these oscillations at the single-cell level are somewhat erratic and imprecise, reflecting the fundamentally noisy, stochastic nature of the control of gene expression, as discussed in Chapter 7. A mechanism is needed to keep all the cells in the presomitic mesoderm that will form a particular somite oscillating in synchrony. This is achieved through cell-cell communication via the Notch signaling pathway, to which the *Hes* genes are coupled. The gene regulatory circuitry is such that in this context Notch signaling does not drive neighboring cells to be different, as in lateral inhibition, but does just the opposite: it keeps them in unison. In mutants where Notch signaling

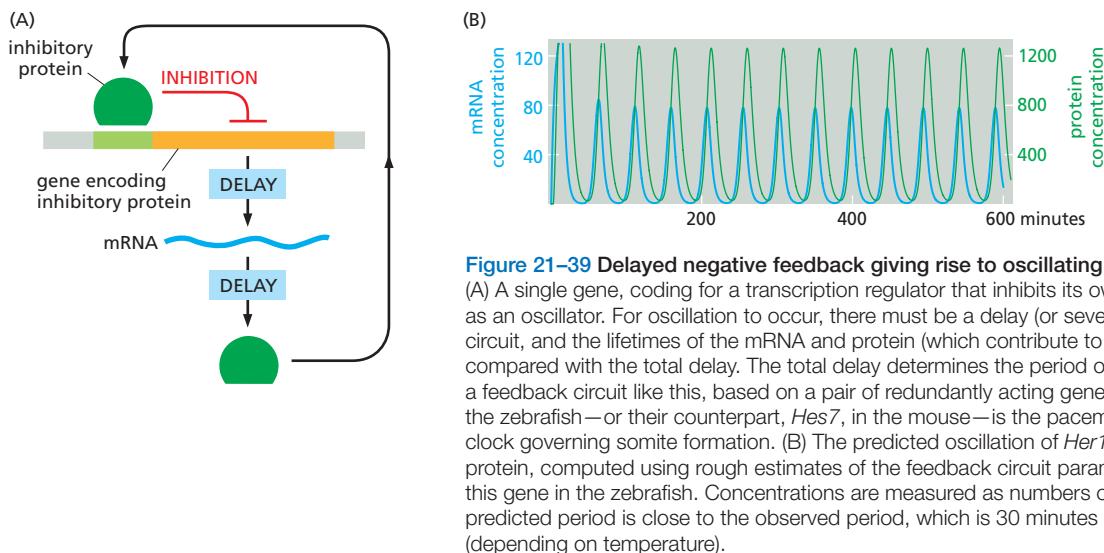


Figure 21-39 Delayed negative feedback giving rise to oscillating gene expression.

(A) A single gene, coding for a transcription regulator 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 (which contribute to the delay) must be short compared with the total delay. The total delay determines the period of oscillation. It is thought that a feedback circuit like this, based on a pair of redundantly acting genes called *Her1* and *Her7* in the zebrafish—or their counterpart, *Hes7*, in the mouse—is the pacemaker of the segmentation clock governing somite formation. (B) The predicted oscillation of *Her1* and *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 (depending on temperature).

fails, including mutants defective in Delta or Notch itself, the cells drift out of synchrony and somite segmentation is again disrupted. This leads to gross deformity of the vertebral column—an extraordinary display of the consequences of the noisy temporal control of gene expression at the single-cell level, writ large in the structure of the vertebrate body as a whole.

Intracellular Developmental Programs Can Help Determine the Time-Course of a Cell's Development

Although signaling between cells plays an essential part in driving the progress of development, this does not mean that cells always need signals from other cells to prod them into changing their character as development proceeds. Some of these changes are intrinsic to the cell (like the ticking of the segmentation clock) and depend on *intracellular developmental programs* that can operate even when the cell is removed from its normal environment.

The best-understood example is in the development of neural precursor cells, or neuroblasts, in the embryonic *Drosophila* central nervous system. These cells, as we saw, are initially singled out from the neurogenic ectoderm of the embryo by a typical lateral-inhibition mechanism that depends on Notch, and they then proceed through an entirely predictable series of asymmetric cell divisions to generate ganglion mother cells that divide to form neurons and glial cells (see Figure 21-36). The neuroblast changes its internal state as it goes through its set program of divisions, generating different cell types with a reproducible sequence and timing. These successive changes in neuroblast specification occur through the sequential expression of specific transcription regulators. For example, most embryonic neuroblasts sequentially express the transcription regulators Hunchback, Krüppel, Pdm, and Cas in a fixed order (Figure 21-40). When a neuroblast divides, the set of transcription regulators expressed at that time is inherited by the ganglion mother cell and its neural progeny; thus, the differentiated neural cells are endowed with different characters according to their time of birth.

Remarkably, when neuroblasts are taken from an embryo and maintained in culture, isolated from their normal surroundings, they step through much the same stereotyped developmental program as if they had been left in the embryo. Moreover, many of the neuroblast transitions occur even when cell division is blocked. The neuroblasts seem to have a built-in timer that determines when each of the transcription regulators is expressed, and this timer can continue to run in the absence of cell division. The molecular basis of the timing is largely unknown; in part, at least, it must depend on the time taken for gene switching, as described above; but it may well also depend on slow progressive changes in chromatin structure. These too can serve to measure the passage of time in the embryo.

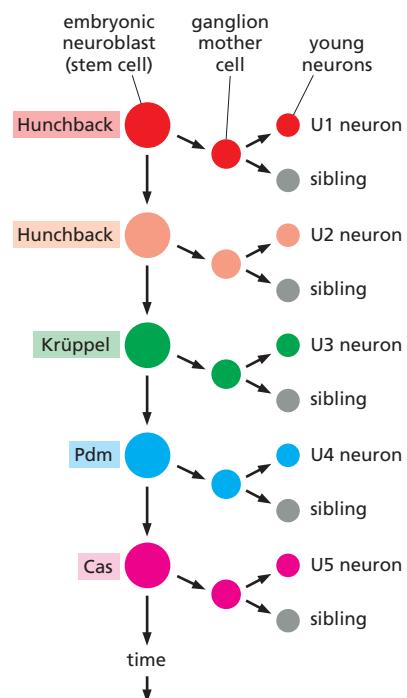


Figure 21-40 Temporal patterning of neuroblast fate in *Drosophila*. Hunchback, Krüppel, Pdm, and Cas are transcription regulators that are expressed consecutively in the cell lineage of neuroblasts during development of the *Drosophila* nervous system. At successive time steps, correlated with cell division, the neuroblast switches its pattern of gene expression. Each neuroblast division produces one daughter that remains a neuroblast and expresses the updated set of genes, and one ganglion mother cell that maintains the expression of this gene set and differentiates into specific cell types accordingly. (After B.J. Pearson and C.Q. Doe, *Nature* 425:624–628, 2003. With permission from Macmillan Publishers.)

Cells Rarely Count Cell Divisions to Time Their Development

Many specialized cells in animals develop from proliferating progenitor cells that stop dividing and terminally differentiate after a limited number of cell divisions. In these cases, differentiation is coordinated with withdrawal from the cell cycle, but it is usually not known how the coordination is achieved. It has often been suggested that the cell-division cycle might serve as an intracellular timer to control the timing of cell differentiation. The cell cycle would be the ticking clock that sets the tempo of other developmental processes, with maturational changes in gene expression being dependent on cell-cycle progression. Most of the evidence, however, indicates that this tempting idea is wrong. Although there are examples where cells change their maturation state with each division and the change depends on cell division, this is not the general rule. As we just saw for neuroblasts in the *Drosophila* embryo, cells in developing animals often carry on with their normal timetable of maturation and differentiation even when cell division is artificially blocked; necessarily, some abnormalities occur, if only because a single undivided cell cannot differentiate in two ways at once. But it seems that most developing cells can change their state without a requirement for cell division. Developmental control genes can switch the cell-division-cycle machinery on or off, and it is the dynamics of these genes, rather than the cell cycle, that sets the tempo of development.

MicroRNAs Often Regulate Developmental Transitions

Genetic screens are useful for tracking down the genes involved in almost any biological process, and they have been used to search for mutations that alter developmental timing. Such screens were performed in the nematode *Caenorhabditis elegans* (Figure 21–41). This worm is small, relatively simple, and precisely structured. The anatomy of its development is highly predictable and has been described in extraordinary detail, so that one can map out the exact lineage of every cell in the body and see exactly how the developmental program is altered in a mutant. Genetic screens in *C. elegans* revealed mutations that disrupt developmental timing in a particularly striking way: in these so-called **heterochronic** mutants, certain cells in a larva at one stage of development behave as though they were in a larva at a different stage of development, or cells in the adult carry on dividing as though they belonged to a larva (Figure 21–42).

Genetic analyses showed that the products of the heterochronic genes act in series, forming regulatory cascades. Unexpectedly, two genes at the top of their respective cascades, called *Lin4* and *Let7*, were found to code not for protein but instead for **microRNAs (miRNAs)**—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 repressing their translation and promoting their degradation, as discussed in Chapter 7. Increasing levels of *Lin4* miRNA govern the progression from first-stage larva cell behaviors to third-stage larva cell behaviors.

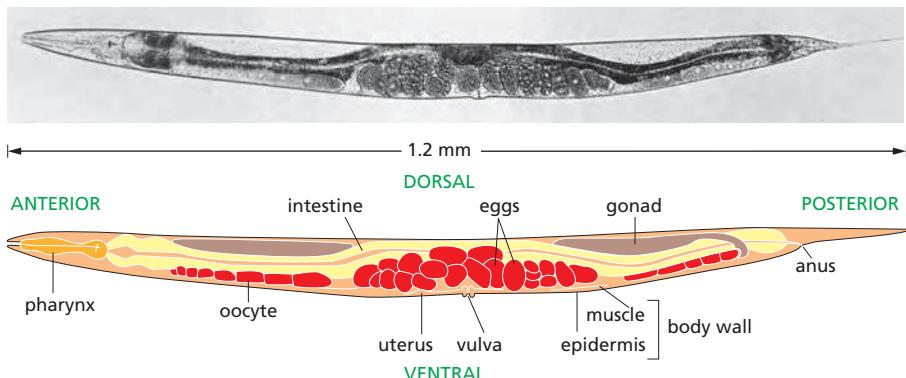


Figure 21–41 *Caenorhabditis elegans*.
A side view of an adult hermaphrodite is shown. (From J.E. Sulston and H.R. Horvitz, *Dev. Biol.* 56:110–156, 1977. With permission from Academic Press.)

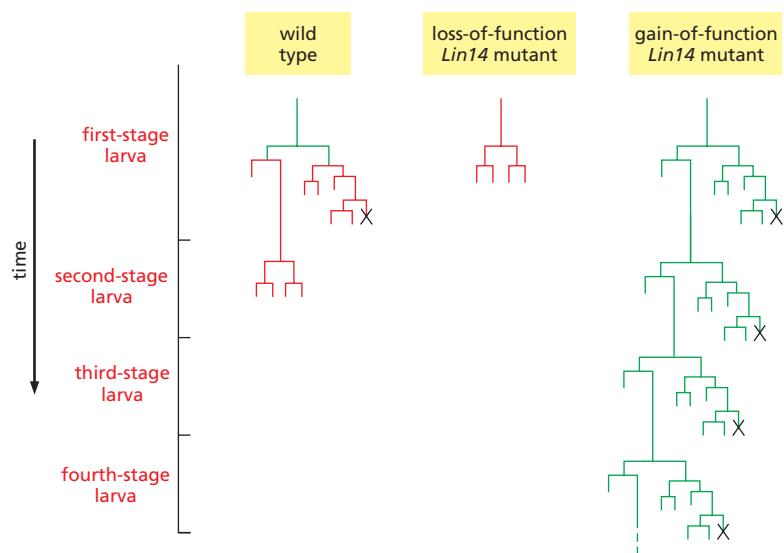


Figure 21–42 Heterochronic mutations in the *Lin14* gene of *C. elegans*. Only the effects on one of the many altered lineages are shown. A loss-of-function (recessive) mutation in *Lin14* causes premature occurrence of the pattern of cell division and differentiation characteristic of a late larva, 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 patterns of cell divisions characteristic of the first larval stage, continuing through as many as five or six molt cycles. 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. (Adapted from V. Ambros and H.R. Horvitz, *Science* 226:409–416, 1984. With permission from the authors; and P. Arasu, B. Wightman and G. Ruvkun, *Genes Dev.* 5:1825–1833, 1991. With permission from the authors.)

Increasing levels of *Let7* miRNA govern the progression from late larva to adult. In fact, *Lin4* and *Let7* were the first miRNAs 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.

More generally, in many animals, miRNAs help regulate the transitions between different stages of development. For example, in flies, fish, and frogs, the maternal mRNAs that are loaded into the egg in the mother are removed during early development when the genome of the embryo begins to be transcribed; at this stage, the embryo begins to express specific miRNAs that target many maternal mRNAs for translational repression and degradation.

Thus, miRNAs can sharpen developmental transitions by blocking and removing mRNAs that define an earlier developmental stage. But how is the timing of miRNA expression itself controlled? In the case of the miRNAs that disable maternal mRNAs in frogs and fish, expression is activated at the end of the series of rapid, synchronous divisions that cleave the fertilized egg into many smaller cells. As the division rate of these blastomeres slows, widespread transcription of the embryo's genome begins (Figure 21–43). This event, where the embryo's own genome largely takes over control of development from maternal macromolecules, is called the **maternal-zygotic transition (MZT)**, and it occurs with roughly similar timing in most animal species, with the exception of mammals.

One trigger for the MZT appears to be the nuclear-to-cytoplasmic ratio. During cleavage, the total amount of cytoplasm in the embryo remains constant, but the number of cell nuclei increases exponentially. As a critical threshold is reached in the ratio of cytoplasm to DNA, the cell cycles lengthen and transcription is initiated. Thus, haploid embryos undergo the MZT one cell cycle later than diploid embryos, which contain twice as much DNA per cell. According to one

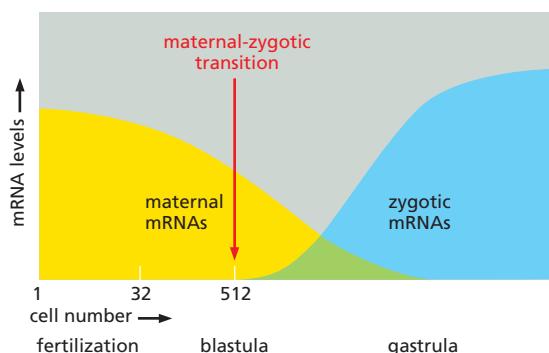


Figure 21–43 The maternal-zygotic transition in a zebrafish embryo. Maternal mRNAs are deposited by the mother into the egg and drive early development. These mRNAs are degraded during different stages of embryogenesis, including blastula and gastrula stages, but a relatively abrupt change occurs at the maternal-zygotic transition (MZT). Before this, the embryonic (zygotic) genome is transcriptionally inactive; afterward, zygotic genes start to be transcribed. In zebrafish embryos, the zygotic genome begins to be activated at the 512-cell stage.

model, the nuclear-to-cytoplasmic ratio might be measured through the titration of a transcription repressor against the increasing amount of nuclear DNA. The total amount of repressor would stay constant during cleavage divisions, but the amount of repressor per genome would decrease, falling by a half with each round of DNA synthesis, until loss of repression allowed the zygotic genome to become transcriptionally active. The newly synthesized transcripts include the miRNAs that recognize many of the transcripts deposited in the egg by the mother, directing their translational repression and rapid degradation.

Hormonal Signals Coordinate the Timing of Developmental Transitions

We have so far emphasized timing mechanisms that operate locally and separately in the different parts of the embryo, or in specific subsystems of the molecular control machinery. Evolution has tuned each of these largely independent processes to run at an appropriate rate, matched to the needs of the organism as a whole. For some purposes, however, this is not enough: a global coordinating signal is required. This is especially true where changes have to occur throughout the body in response to a cue that depends on the environment. For example, when an insect or amphibian undergoes *metamorphosis*—the transition from larva to adult—almost every part of the body is transformed. The timing of metamorphosis depends on external factors such as the supply of food, which determines when the animal reaches an appropriate size. All the bodily changes have to be triggered together at the right time, even though they are occurring in widely separated sites. The coordination in such cases is provided by **hormones**—signal molecules that spread throughout the body.

The metamorphosis of amphibians provides a spectacular example. During this developmental transition, amphibians switch from an aquatic to a terrestrial life. Larva-specific organs such as gills and tail disappear, and adult-specific organs such as legs form. This dramatic transformation is triggered by thyroid hormone, produced in the thyroid gland. If the gland is removed or if thyroid hormone action is blocked, metamorphosis does not occur, although growth continues, producing a giant tadpole. Conversely, a dose of thyroid hormone given to a tadpole by an experimenter can trigger metamorphosis prematurely.

The thyroid hormone is distributed through the vascular system and induces changes throughout the animal by binding to intracellular nuclear hormone receptors, which regulate hundreds of genes. This does not mean, however, that target tissues all respond in the same way to the hormone: organs differ not only in their levels of thyroid hormone receptors and levels of extracellular proteins that locally regulate the amount of active hormone, but also in the sets of genes that respond. Thyroid hormone induces muscle in the limbs to grow and muscle in the tail to die. The timing of the responses also differs: for example, the legs form early in response to a very low concentration of circulating hormone, but it requires a high level of the hormone to induce resorption of the tail.

A surge of thyroid hormone triggers metamorphosis, but how is the timing of the surge controlled? One mechanism depends on coupling hormone synthesis to the size of the thyroid gland, which reflects the size of the tadpole. Only when the gland attains a certain size does it produce enough thyroid hormone to initiate metamorphosis. However, environmental cues other than nutrition also play a part: conditions such as temperature and light are sensed by the nervous system, which regulates the secretion of another tier of hormones (neurohormones) that stimulate the secretion of thyroid hormone. Thus, tadpole-intrinsic factors such as size combine with environmental factors to determine when metamorphosis begins.

Environmental Cues Determine the Time of Flowering

Another striking example of environmentally controlled developmental timing is the flowering of plants. Flowering involves a transformation of the behavior of the

cells at the growing apex of the plant shoot—the *apical meristem*. During ordinary vegetative growth, these cells behave as stem cells, generating a steady succession of new leaves and new segments of stalk. In flowering, the meristem cells switch to making the components of a flower, with its sepals and petals, its stamens carrying pollen, and its ovary containing the female gametes.

To time the switch correctly, the plant has to take account of both past and present conditions. 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, including the phytochromes and cryptochromes that act as light receptors (discussed in Chapter 15). The flowering signal that is carried from the leaves to the stem cells via the vasculature depends on the product of *Flowering locus T* (*Ft*).

But this signal will trigger flowering only if the plant is in a receptive condition from prior long-term cold exposure. Many plants need winter before they will flower—a process called *vernalization*. Cold over a period of weeks or months progressively reduces the level of expression of a remarkable gene called *Flowering locus C* (*Flc*). *Flc* encodes a transcriptional repressor that suppresses expression of the *Ft* flowering promoter.

How does vernalization shut down *Flc* so as to lift the block to flowering? The effect involves a noncoding RNA called *Coolair* that overlaps with the *Flc* gene and is produced when the temperature is low (Figure 21–44). Together with cold-induced chromatin modifiers, including Polycomb-group proteins, *Coolair* coordinates the switching of *Flc* chromatin to a silent state (discussed in Chapters 4 and 7). The degree of silencing depends on the length of cold exposure enabling the plants to distinguish the odd chilly night from the whole of winter.

The effect on the chromatin is long lasting, persisting through many rounds of cell division even as the weather grows warmer. Thus vernalization creates a persistent block in production of *Flc*, enabling the *Ft* signal to be generated when day length is sufficiently long.

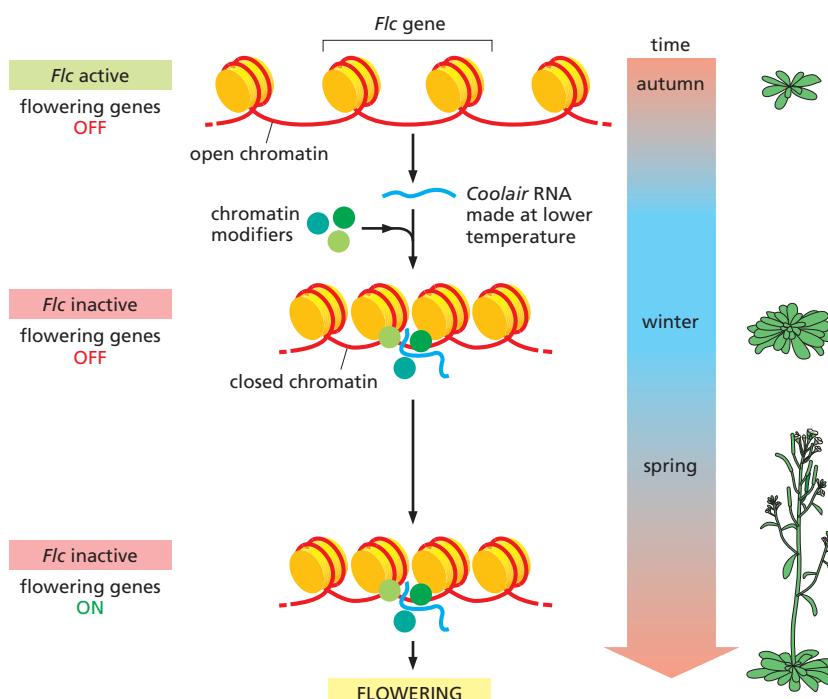


Figure 21–44 Temporal control of flowering in *Arabidopsis*. The *FLC* gene is active and blocks flowering when plants have been grown without exposure to winterlike temperatures. Exposure to a prolonged period of cold leads to the production of the noncoding RNA *Coolair*, which overlaps with the *FLC* gene. *Coolair* induces long-term chromatin changes that turn off *FLC*. These changes persist after the end of the cold period and allow the plant to flower when other environmental conditions are favorable for flowering.

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 control system governing the switch to flowering is thus of vital importance for agriculture, especially in an era of rapid climate change.

The example of vernalization suggests a general point about the role of chromatin modification in developmental timing. The plant uses changes in chromatin to record its experience of prolonged cold. It may be that in other organisms—animals as well as plants—slow, progressive changes in chromatin structure provide long-term timers for those mysterious developmental processes that unfold slowly, over a period of days, weeks, months, or years. Such chromatin timers may be among the most important clocks in the embryo, but as yet we understand very little about them.

Summary

Developmental timing is controlled at many levels. It takes time to switch a gene on or off, and this time delay depends on the lifetimes of the molecules involved, which can vary widely. Cascades of gene regulation involve cascades of delays. Feedback loops can give rise to temporal oscillations in gene expression, and these may serve to generate spatially periodic structures. During vertebrate segmentation, for example, expression of the Hes genes oscillates, and one new pair of somites is formed during each oscillation cycle. Hes genes encode transcription repressor proteins that can act back on expression of the Hes genes themselves. This negative feedback generates oscillations with a period that reflects the delay in the autoregulatory gene switching loop. The period of oscillation of this “segmentation clock” controls the sizes of the somites. Notch signaling between neighboring cells synchronizes their oscillations: when Notch signaling fails, the cells drift out of synchrony because of genetic noise in their individual clocks, and the segmental organization of the vertebral column is disrupted.

Timing does not always depend on cell-cell interactions; many developing animal cells have intrinsic developmental programs that play out even in isolated cells in culture. Neuroblasts in Drosophila embryos, for example, go through set programs of asymmetric divisions, generating different neural cell types at each division with a predictable sequence and timing, through a cascade of gene switching events. Studies in both vertebrates and invertebrates show that such programs are rarely governed by the timing of cell division and can unfold even when cell division is blocked. MicroRNAs produced at critical moments sharpen developmental transitions by blocking the translation and promoting the degradation of specific sets of mRNAs. Global coordination of developmental timing is achieved by hormones: as a tadpole grows, for example, thyroid hormone levels surge and trigger its metamorphosis into a frog. Environmental control of developmental timing is especially striking in plants and reveals the presence of molecular timers that act over the long term. In vernalization, for example, prolonged cold induces changes in chromatin that chart the passage through winter so as to allow flowering only in the spring. Slow, progressive changes in chromatin structure are likely to be important timers in the long-term programming of development in animals too.

MORPHOGENESIS

The specialization of cells into distinct types at specific times is important, but it is only one aspect of animal development. Equally important are the movements and deformations that cells go through to assemble into tissues and organs with specific shapes and sizes. Like developmental timing, this process of **morphogenesis** (“form generation”) is less well understood than the processes of differential gene expression and inductive signaling that lead to cell-type specialization. The cell movements can be readily described, but the underlying molecular mechanisms that coordinate the movements are much harder to decipher.

In Chapter 19, we saw how cells cohere to form epithelial sheets or surround themselves with extracellular matrix to create connective tissues. We also discussed how the basic features of tissues, such as the polarity of epithelia, arise

from the properties of individual cells. In this section, we consider how the rearrangements of cells during animal development give shape to the embryo and to all the individual organs and appendages of the body.

A small number of cell processes are basic to morphogenesis. Individual cells can migrate through the embryo along defined tracks. They can crawl over one another in a coordinated way to elongate, constrict, or thicken a tissue. They can segregate from their neighbors and form physically separate groups. They can change their shape so as to deform an epithelial sheet into a tube or a vesicle. By stretching out while holding on to their companions, specialized sets of cells can form growing tubular networks such as the system of blood or lymph vessels. Mass migrations, as occur in gastrulation, can transform the entire topology of the embryo. Underlying all these processes are changes in cell shape and changes in cell contacts—either with other cells or with extracellular matrix. We begin by considering the migration of individual cells.

Cell Migration Is Guided by Cues in the Cell's Environment

The birthplace of cells is often far from their ultimate location in the body. Our skeletal muscles, for example, derive from muscle cell precursors, or *myoblasts*, in somites, from which they migrate into the limbs and other regions. The routes that the migrant cells follow and the selection of sites that they colonize determine the eventual pattern of muscles in the body. The embryonic connective tissues form the framework through which the myoblasts travel, and these tissues provide the cues that guide myoblast distribution. No matter which somite they come from, the 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. It is the connective tissue that provides the patterning information.

As a migrant cell travels through the embryonic tissues, it repeatedly extends surface projections that probe its immediate surroundings, testing for cues to which it is particularly sensitive by virtue of its specific assortment of cell-surface receptor proteins. Inside the cell, these receptors are connected to the cortical actin and myosin cytoskeleton, which moves the cell along. Some extracellular matrix molecules, such as the protein fibronectin, provide adhesive sites that help the cell advance; others, such as chondroitin sulfate proteoglycan, inhibit locomotion and repel immigration. The nonmigrant cells along the migration pathway may likewise have inviting or repellent macromolecules on their surface; some may even extend filopodia to make their presence known.

Among the many guiding influences, a few stand out as especially important. In particular, many types of migrating cells are guided by chemotaxis that depends on a G-protein-coupled receptor (called CXCR4), which is activated by an extracellular ligand called CXCL12. Cells expressing this receptor can snuffle their way along tracks marked out by CXCL12 (Figure 21–45). Chemotaxis toward sources of CXCL12 plays a major part in guiding the migrations of lymphocytes

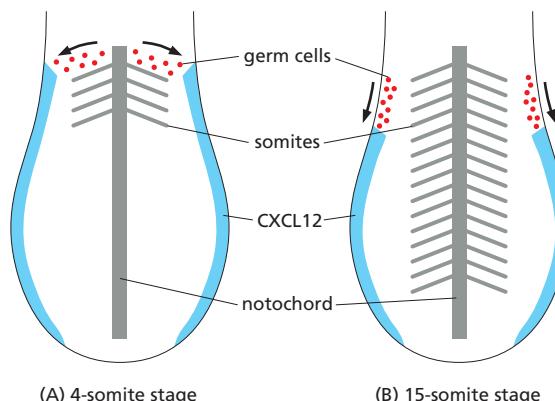


Figure 21–45 CXCL12 guides migrating germ cells. Zebrafish germ cells migrate to domains that express CXCL12. As the sites of CXCL12 expression change, cells follow the CXCL12 track and are guided to the region where the gonad develops at a later developmental stage. (A) At the 4-somite stage, germ cells move from a position that is close to the midline to more lateral regions where CXCL12 is expressed. (B) As the CXCL12 expression retracts, germ cells are guided to more posterior positions.

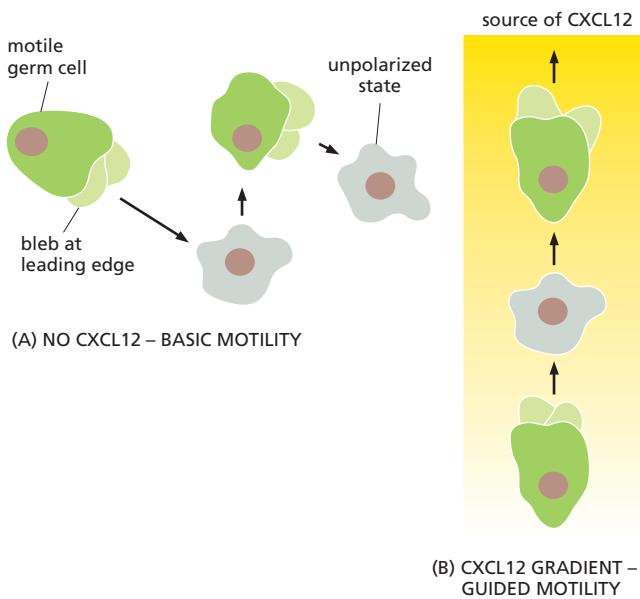


Figure 21–46 Directional migration by local blebbing. Germ cells migrate via protrusions that define the leading edge of the cell. The persistence and site of the protrusions are biased toward higher levels of CXCL12. Thus, germ cells migrate up the CXCL12 gradient.

and various other white blood cells; of neurons in the developing brain; of myoblasts entering limb buds; of primordial germ cells as they travel toward the developing gonads; and of cancer cells when they metastasize.

Detailed studies of primordial-germ-cell migration have shown that CXCL12 signaling does not induce cell migration per se but rather serves to control its direction. In the absence of CXCL12 signaling, germ cells still display the membrane blebbing associated with cell migration, but the position of the cell front where blebs form is randomly chosen (Figure 21–46); if CXCL12 signaling is intact, blebbing is more frequent on the side of the cell that faces the source of CXCL12, resulting in directional migration.

The Distribution of Migrant Cells Depends on Survival Factors

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 for specific types of migrant cells to survive.

Among the most important sets of migrant cells in the vertebrate embryo are those of the **neural crest**. They arise from the border region between the part of the ectoderm that will form epidermis and the part that will form the central nervous system. As the neural ectoderm rolls up to form the neural tube, the neural crest cells break loose from the epithelial sheet along this border region and set out on their long migrations (see Figure 19–8 and [Movie 21.5](#)). They settle ultimately in many sites and give rise to a surprising diversity of cell types. Some lodge in the skin and specialize as pigment cells; still others form skeletal tissue in the face. Still others will differentiate into the neurons and glial cells of the peripheral nervous system—not only in the sensory ganglia that lie close to the spinal cord, but also, following a much longer migration, in the wall of the gut.

The neural crest cells that give rise to the pigment cells of the skin and those that develop into the nerve cells of the gut depend on a secreted peptide called endothelin-3, which is produced by tissues along the migration pathways and acts as a survival factor for the migrating crest cells. In mutants with a defect in the gene for endothelin-3 or its receptor, many of these migrating crest cells die. As a result, the mutant individuals have nonpigmented (albino) patches of skin and a deficit of nerve cells in the intestine, especially its lower end, the large bowel, which becomes abnormally distended for lack of proper neural control—a potentially lethal condition called megacolon.



Figure 21–47 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 gene product as a receptor for a survival factor. (Courtesy of R.A. Fleischman, from R.A. Fleischman et al., *Proc. Natl Acad. Sci. USA* 88:10885–10889, 1991.)

Another important survival signal for many types of migratory cells, including primordial germ cells, blood cell precursors, and neural-crest-derived pigment cells, depends on a receptor tyrosine kinase called *Kit*. This is expressed on the surface of the migrant cells, and a protein ligand, called Steel factor, is 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 have deficits in pigmentation, blood cells, and germ cells (Figure 21–47).

Changing Patterns of Cell Adhesion Molecules Force Cells Into New Arrangements

Patterns of gene expression govern embryonic cell movements in many ways. They regulate cell motility, cell shape, and the production of proteins that guide migration. Importantly, they also determine the sets of adhesion molecules that the cells display on their surface. 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 with different surface chemistry.

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 postgastrulation embryo after the cells have been artificially dissociated and mixed up. When these cells are reaggregated into a random mixture, the cells spontaneously sort themselves out according to their original germ-layer origins (Figure 21–48). As discussed in Chapter 19, cadherin proteins have a central role in the sorting process (see Figure 19–9). Cadherins belong to a large and varied family of Ca^{2+} -dependent cell-cell adhesion proteins, and they and other cell-cell adhesion proteins are differentially expressed in the various tissues of the early embryo. Antibodies against these proteins 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 various developmental processes, including gastrulation, neural tube formation, and somite formation. These cell rearrangements are likely to be regulated and driven in part by

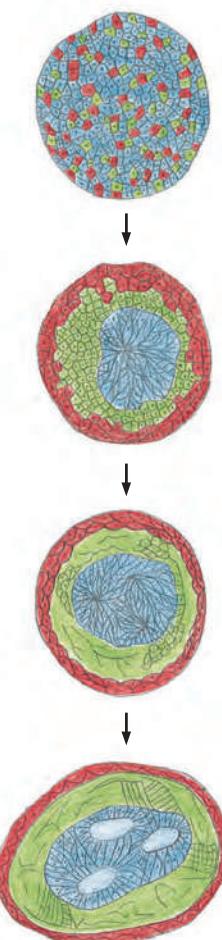


Figure 21–48 Sorting out by adhesion. Cells from different parts of an early amphibian embryo will sort out according to their origins. In the classical experiment shown here, mesoderm cells (green), neural plate cells (blue), and epidermal cells (red) have been disaggregated and then reaggregated in 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.)

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 (see Movie 19.1). 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 cell adhesions.

Repulsive Interactions Help Maintain Tissue Boundaries

The different types of cadherins enable different types of cells to cohere selectively: cells expressing one type of cadherin will maximize their contact with cells expressing the same cadherin and thereby segregate from other cells, creating specific tissue boundaries. Cell mixing can be inhibited and boundaries created and maintained in another way as well: cells of different types can sometimes actively repel one another. The bidirectional activation of Eph receptors and ephrins discussed in Chapter 15 often mediates such repulsion, acting at interfaces between different groups of cells to keep the groups from mixing, and repelling invasion by inappropriate visitors. Ephrin–Eph signaling operates, for example, at the boundaries of the rhombomeres discussed earlier. Neighboring rhombomeres express complementary combinations of ephrins and Eph receptors, and this keeps the cells in adjacent rhombomeres strictly segregated, with a boundary between them that is sharply defined (**Figure 21–49**).

Groups of Similar Cells Can Perform Dramatic Collective Rearrangements

Cadherin-mediated cell sorting and ephrin–Eph-mediated repulsion exemplify how differences in cell-surface properties can drive tissue arrangements, causing cells that express different sets of genes to separate from one another. However, groups of cells that are all similar can also undergo dramatic rearrangements. During frog gastrulation, for example, cells in one region of the surface epithelium invaginate and migrate as a sheet into the interior of the embryo and converge toward the embryonic midline. The movement is driven mainly by an active rearrangement of the migrating cells, called **convergent extension**. Here the cells crawl over one another in a coordinated way, displacing their neighbors as they migrate, causing the cell sheet to narrow along one axis (converge) and elongate along another (extend). Strikingly, small, square fragments of tissue from the

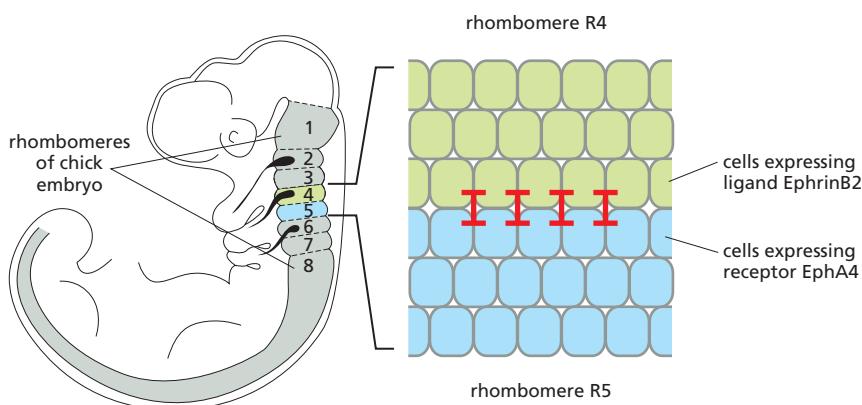


Figure 21–49 Sorting out by repulsion. Ephrin–Eph signaling in hindbrain segmentation in a chick embryo. Each pair of rhombomeres (segments in the hindbrain) is associated with a branchial arch (a modified gill rudiment) to which it sends innervation. Rhombomeres are distinguished from one another by expression of different *Hox* genes (see Figure 21–32). Mutual repulsion (red bars) between cells that express EphrinB2 in rhombomere 4 and EphA4 in rhombomere 5 creates a sharp boundary.

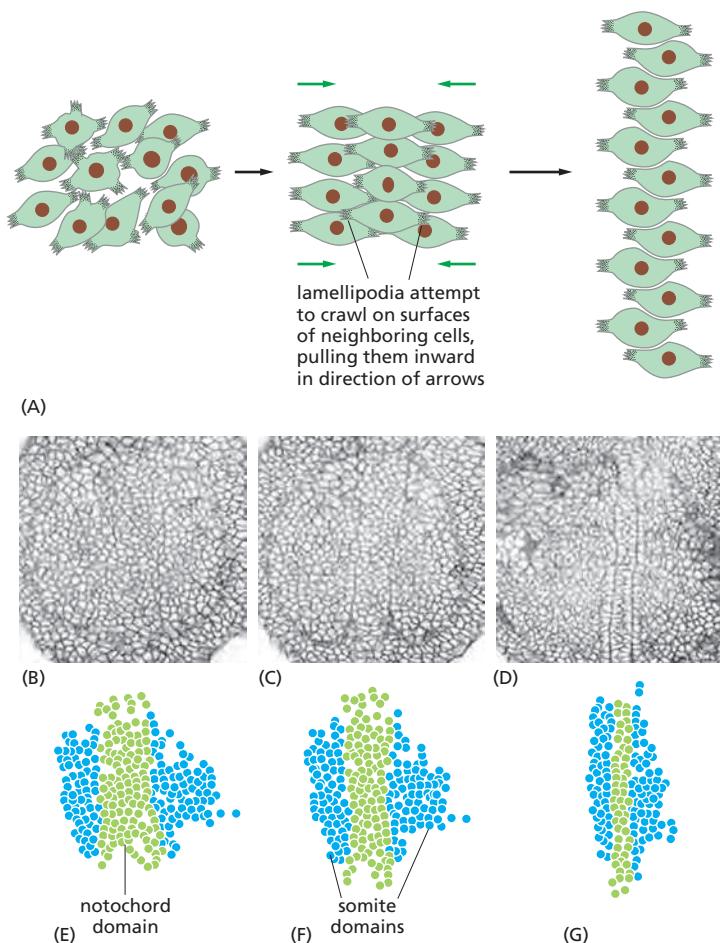


Figure 21–50 Convergent extension and its cellular basis. (A) Schematic diagram of cell behaviors that underlie convergent extension. The cells form lamellipodia, with which they attempt to crawl over one another. Alignment of the lamellipodial movements along a common axis leads to convergent extension. The process depends on the Wnt–Frizzled/planar-cell-polarity signaling pathway and is cooperative, presumably because cells that are already aligned exert forces that tend to align their neighbors in the same way.

(B–G) The pattern of convergent extension of dorsal mesoderm during zebrafish gastrulation at 8.8 (B, E), 9.3 (C, F), and 11.3 (D, G) hours after fertilization. Cells that will give rise to the notochord are labeled in green, and cells that will give rise to somites and muscle are labeled in blue. The notochord and somite domains are spatially separate from the start of the recording (B, E), but their boundaries are at first barely visible and only a little later become obvious. Convergence narrows the notochord domain to a width of about two cells at the last time point (D, G). (A, after J. Shih and R. Keller, *Development* 116:901–914, 1992; B–G, after N.S. Glickman et al., *Development* 130:873–887, 2003. With permission from The Company of Biologists.)

appropriate region of the embryo, isolated in culture, will spontaneously narrow and elongate, just as they would in the embryo (Figure 21–50). The alignment of the cell movements depends on the same signaling pathway that is involved in generating *planar cell polarity* within developing epithelia, as we discuss next.

Planar Cell Polarity Helps Orient Cell Structure and Movement in Developing Epithelia

Cells within an epithelium always have an apical–basal polarity (discussed in Chapter 19), but the cells of many epithelia show an additional polarity at right angles to this axis: the cells are all arranged 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**. 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 a precisely oriented asymmetric bundle of actin-filled, rodlike protrusions called stereocilia sticking up from its apical plasma membrane as a detector of sound and of forces such as gravity. Tilting the bundle in one direction causes ion channels in the membrane to open, electrically activating the cell; tilting in the opposite direction has the opposite effect. For the ear to function properly, the hair cells must be oriented correctly. Planar cell polarity is also important in the respiratory tract, where every ciliated cell must orient the beating of its cilia so as to sweep mucus upward, away from the lungs.

Screens for mutants with misoriented wing hairs in *Drosophila* have identified a set of genes that is critical for planar cell polarity. Some of these genes code for

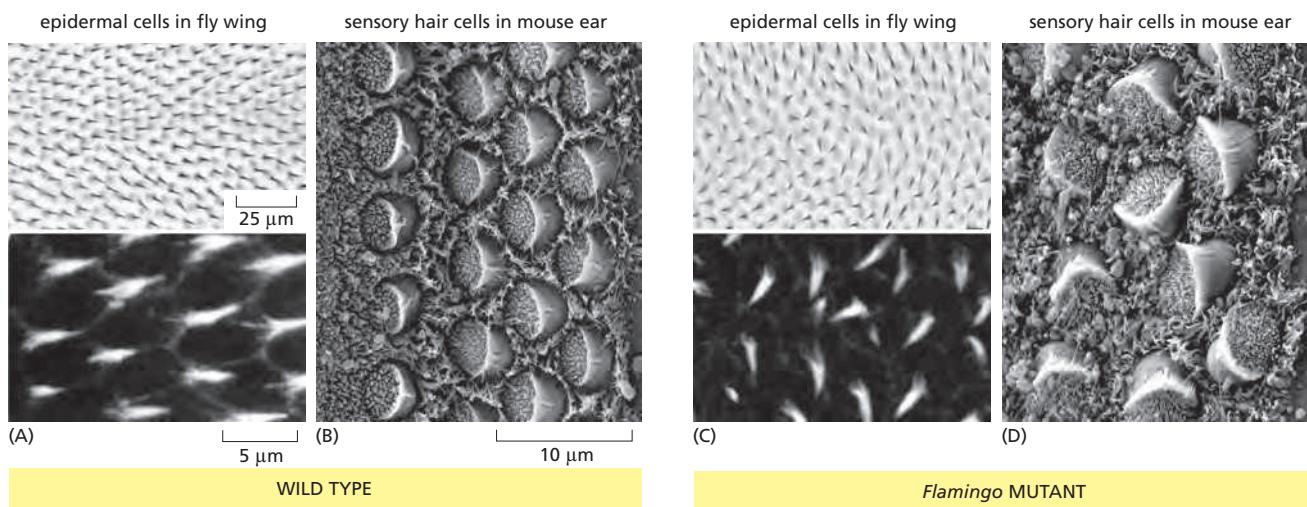


Figure 21–51 Planar cell polarity. (A) Wing hairs on the wing of a fly. Each cell in the wing epithelium forms a small, spiky protrusion or “hair” at its apex, and all the hairs point the same way, toward the tip of the wing. This reflects a planar polarity in the structure of each cell. (B) Sensory 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 nonclassical 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.)

components of the Wnt signaling pathway, others code for specialized members of the cadherin superfamily, while the functions of others are uncertain. These components of planar-cell-polarity signaling are assembled at cell-cell junctions in the epithelium in such a way as to exert a polarizing influence that can propagate from cell to cell. Essentially the same system of proteins controls planar cell polarity in vertebrates; mice deficient in homologs of the *Drosophila* planar polarity genes have a variety of defects, including incorrectly oriented hair cells in the inner ear, making them deaf (Figure 21–51).

Interactions Between an Epithelium and Mesenchyme Generate Branching Tubular Structures

Animals require specialized types of epithelial surfaces for many functions, including excretion, absorption of nutrients, and gas exchange. Where large surfaces are required, they are often organized as branching tubular structures. The lung is an example. It originates from epithelial buds that grow out from the floor of the foregut and invade neighboring mesenchyme to form the bronchial tree, a system of tubes that branch repeatedly as they extend. Endothelial cells that form the lining of blood vessels invade the same mesenchyme, thereby creating a system of closely apposed airways and blood vessels, as required for gas exchange in the lung (Figure 21–52). This whole process of *branching morphogenesis* depends on signals that pass in both directions between the growing epithelial buds and the mesenchyme. Genetic studies in mice indicate that FGF proteins and their receptor tyrosine kinases play a central part in these signaling processes. FGF signaling has various roles in development, but it is especially important in the many interactions that occur between a developing epithelium and mesenchyme.

In the case of lung development, FGF10 is expressed in clusters of mesenchyme cells that lie near the tips of the growing epithelial tubes, and its receptor is expressed in the invading epithelial cells. In FGF10-deficient mutant mice, a primary bud of lung epithelium is formed but fails to grow out into the mesenchyme to create a branching bronchial tree. Conversely, a microscopic bead soaked in

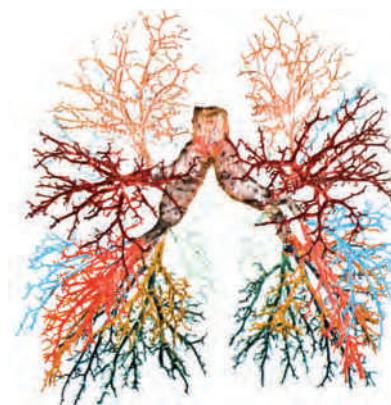
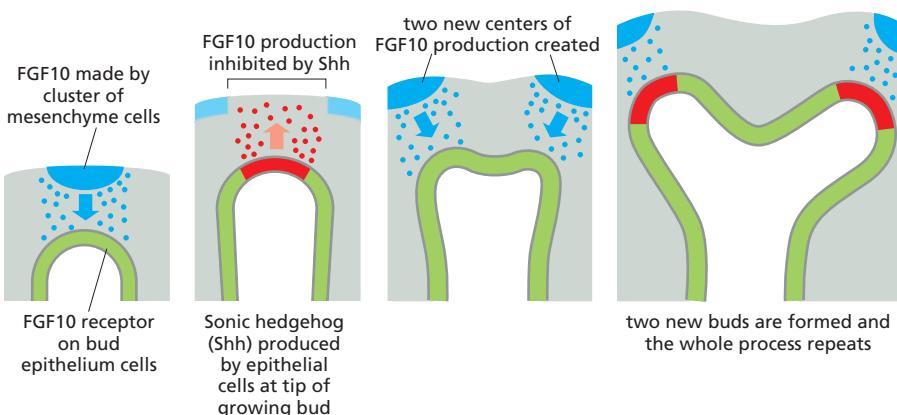


Figure 21–52 The airways of the lung, shown in a cast of the adult human bronchial tree. Resins of different colors have been injected into different branches of the tree of airways. (From R. Warwick and P.L. Williams, Gray's Anatomy, 35th ed. Edinburgh: Longman, 1973.)



FGF10 and placed near embryonic lung epithelium in culture will induce a bud to form and grow out from the epithelium toward the bead. Evidently, the epithelium invades the mesenchyme only by invitation, in response to FGF10.

But what makes the growing epithelial tubes of the lung branch repeatedly as they invade the mesenchyme? This depends 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, as shown in **Figure 21–53**. In mice lacking Sonic hedgehog, the lung epithelium grows and differentiates, but it forms a sac instead of a branching tree of tubules.

FGF signaling acts in a remarkably similar way in the formation of the air-exchange system of insects, which consists of a pattern of fine, air-filled channels called *tracheae* and *tracheoles*. These originate from the epidermis covering the surface of the body and extend inward to invade the underlying tissues, branching and narrowing as they go (**Figure 21–54**). The FGF acts on cells at the tips of the advancing tracheae, causing them to extend filopodia and migrate toward the source of the FGF signal. Because the tip cells remain connected to the remainder of the tracheal epithelium, the pulling force that they generate elongates the tracheal tube.

Initially, the pattern of FGF production in fly embryos is defined by the D-V and A-P patterning systems discussed earlier. In later stages of development, however, FGF expression is induced by transcription regulators called *hypoxia-inducible factors* (*HIFs*) that are activated by hypoxia (low oxygen levels). In this way, hypoxia stimulates the formation of finer and finer and more extensively branched trachea, until the oxygen supply is sufficient to stop the process. Hypoxia and HIFs have similar roles in vertebrates, especially in the development of blood vessels, as we shall see in the next chapter.

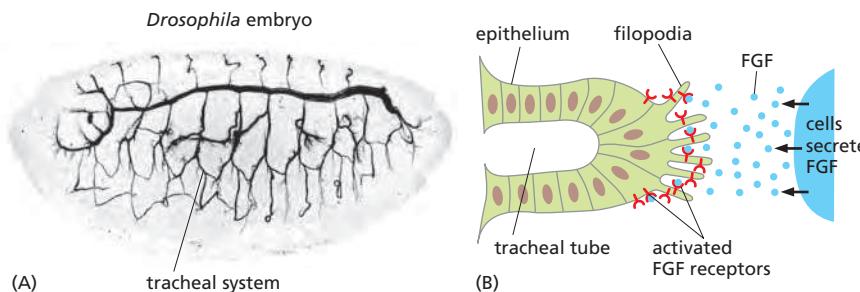
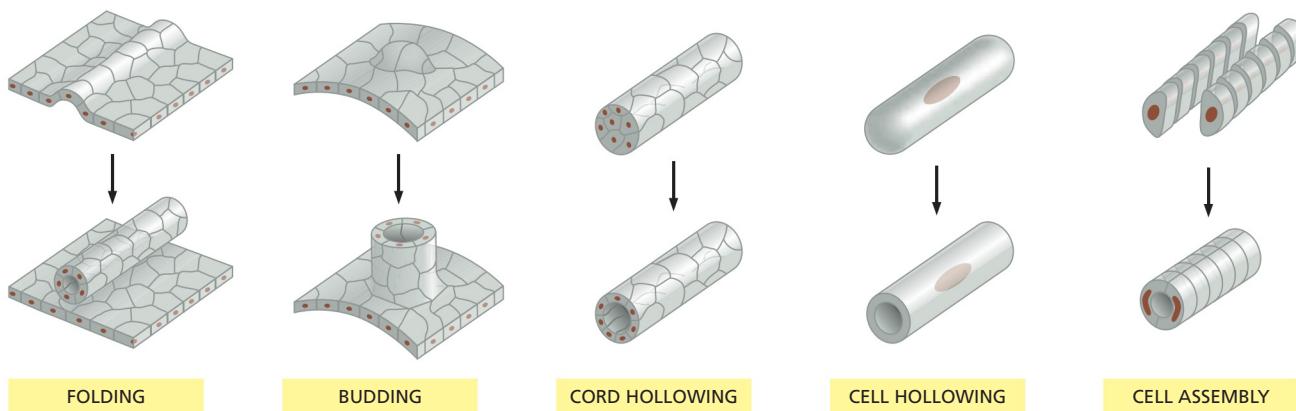


Figure 21–54 Branching morphogenesis of airways in a fly. (A) *Drosophila* embryonic tracheal system. (B) FGF (produced in *Drosophila* by the *Branchless* gene) signals from surrounding cells to the tracheal epithelium and activates its FGF receptors, leading to filopodia formation and tube elongation. [A, from G. Manning and M.A. Krasnow, in *The Development of Drosophila* (A. Martinez-Arias and M. Bate, eds), Vol. 1, pp. 609–685. New York: Cold Spring Harbor Laboratory Press, 1993.]

Figure 21–53 Branching morphogenesis of the lung. 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.

As indicated, FGF10 protein is expressed in clusters of mesenchyme cells near the tips of the growing epithelial tubes, and its receptor is expressed in the epithelial cells themselves. The Sonic hedgehog signal is sent in the opposite direction, from the epithelial cells at the tips of the buds back to the mesenchyme. The patterns of gene expression and their timing suggest 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.



An Epithelium Can Bend During Development to Form a Tube or Vesicle

The creation of systems of tubes such as blood vessels and airways is a complex process, and it can involve various additional forms of cell behavior, as sketched in **Figure 21–55**.

As explained in Chapter 19, the process that converts an epithelial sheet into a tube depends on contraction of specific bundles of actin filaments. With the help of myosin motor proteins, actin filament bundles can shorten, causing the epithelial cells to narrow at their apex. These actin bundles are connected from cell to cell by adherens junctions, and if their contraction is coordinated along a specific axis, the result will be that the sheet bends and rolls up into a tube (**Figure 21–56**). The vertebrate neural tube, which we discuss in the last section of this chapter, originates in this way.

Figure 21–55 The forms of cell behavior involved in tube formation. Folding generates the neural tube, budding underlies the formation of lungs and trachea, cord hollowing occurs during the formation of mammalian salivary glands, cell hollowing is involved in the formation of tracheal terminal cell tubes, and cell assembly generates the heart tube that forms at the earliest stage of heart development.

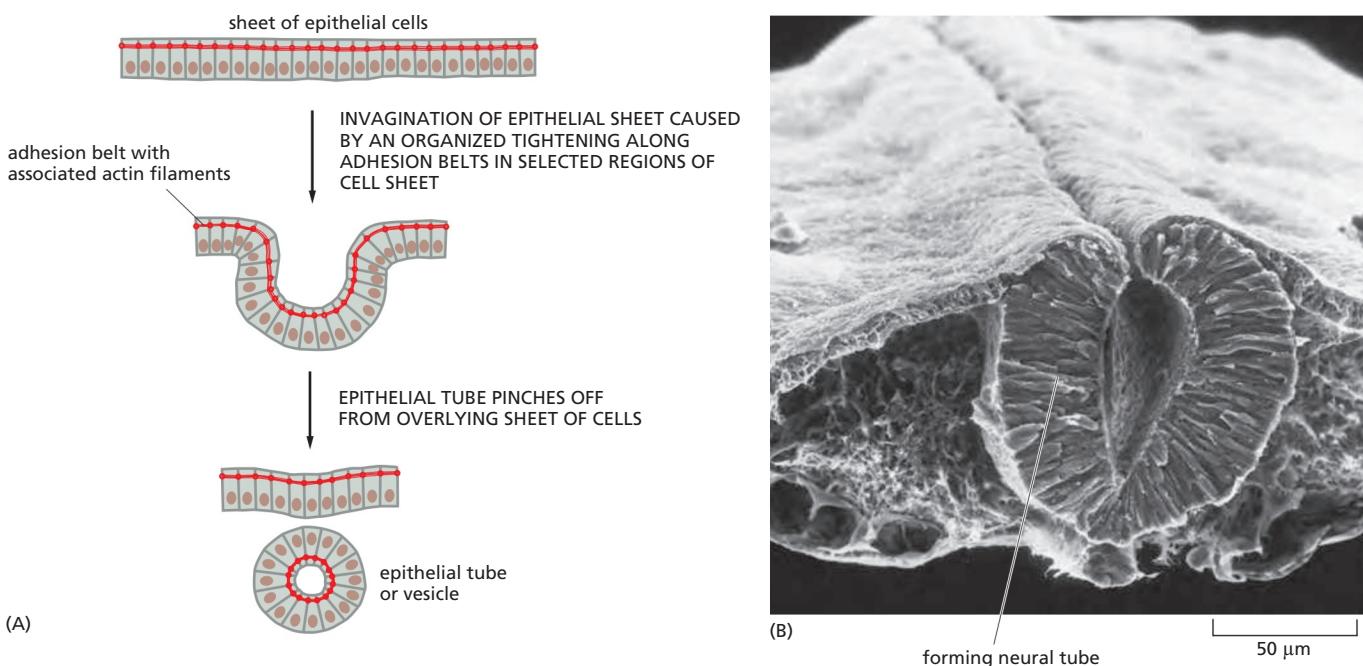


Figure 21–56 Bending of an epithelial sheet to form a tube. Contraction of apical bundles of actin filaments linked from cell to cell via adherens junctions causes the epithelial cells to narrow at their apex. Depending on whether the contraction is oriented along one axis of the sheet or is equal in all directions, the epithelium will either roll up into a tube or invaginate to form a vesicle. (A) Diagram showing how an apical contraction along one axis of an epithelial sheet can cause the sheet to form a tube. (B) Scanning electron micrograph of a cross section through the trunk of a two-day chick embryo, showing the formation of the neural tube by the process diagrammed in (A). (B, courtesy of Jean-Paul Revel.)

Summary

Animal development involves dramatic cell movements, including the guided migration of individual cells, the adhesion and repulsion of groups of cells, and the complex extension, branching, or rolling up of epithelial tissues. Migrant cells, such as those of the neural crest, break loose from their original neighbors and travel through the embryo to colonize new sites. Many migrant cells, including primordial germ cells, are guided by chemotaxis dependent on the receptor CXCR4 and its ligand CXCL12. In general, cells that have similar adhesion molecules on their surfaces cohere and tend to segregate from other cell groups with different surface properties. Selective cell-cell adhesion is often mediated by cadherins; repulsion is often driven by ephrin-Eph signaling. Within an epithelial sheet, cells can rearrange themselves to drive epithelial convergence and extension, as in gastrulation. Many movements are coordinated through a Wnt-dependent planar-polarity signaling pathway that is also responsible for orienting cells correctly in various types of epithelium. Elaborate branched tubular structures, such as the airways of the lung, are generated through bidirectional signaling between an epithelial bud and the mesenchyme that it invades, in a process called branching morphogenesis. Epithelial tubes and vesicles can originate in various ways, most simply by the rolling up and pinching off of a segment of epithelium, as in the formation of the neural tube.

GROWTH

One of the most fundamental aspects of animal development is one we know surprisingly little about—how the size of an animal or an organ is determined. Why, for example, do we grow to be so much larger than a mouse? Even within a species, size can vary greatly; a Great Dane, for instance, can weigh over 40 times more than a Chihuahua (**Figure 21–57**).

Three variables define the size of an organ or organism: the number of cells, the size of the cells, and the quantity of extracellular material per cell. Size differences can arise from changes in any of these factors (**Figure 21–58**). If we compare a mouse with a human, for example, we find that the difference lies chiefly in the number of cells, there being roughly 3000 times more cells in a human, corresponding to a body that is roughly 3000 times more massive. Wild and cultivated species of food plants, on the other hand, often differ in body size chiefly because of differences of cell size.

The challenge, therefore, is to understand how cell numbers, cell size, and extracellular matrix production are regulated. First of all, we need to identify the signals that drive or inhibit growth. Then we need to discover how the signals themselves are regulated. In many cases, the size of an organ or of the body as a whole seems to be controlled homeostatically, so that the correct size is reached and maintained even in the face of drastic disturbances. This suggests that the developing structure somehow senses its own size and uses this information to regulate the signals for its own growth or shrinkage. In most cases, the nature of this feedback control remains a profound mystery.

In other cases, the duration of growth and the final size seem to be dictated by intracellular programs that take no cognizance of the size the structure has attained. These intracellular programs, too, present many mysteries, as we saw in our discussion of developmental timing. Very often, it seems, the sizes and proportions of body parts must depend on combinations of size-measuring feedback controls and intracellular programs, as well as on environmental influences such as nutrition.

The variation in control strategies is nicely illustrated by some classic transplantation experiments. If several fetal thymus glands are transplanted into a developing mouse, each grows to its characteristic adult size. In contrast, if multiple fetal spleens are transplanted, each ends up smaller than normal, but collectively they grow to the size of one adult spleen. Thus, thymus growth is regulated by local mechanisms intrinsic to the individual organ, whereas spleen growth is controlled by a feedback mechanism that senses the quantity of spleen tissue in the body as a whole. In neither case is the mechanism known.



Figure 21–57 Members of the same species can have dramatically different sizes. The Chihuahua weighs 2–5 kilograms, whereas a Great Dane weighs 45–90 kilograms. (Courtesy of Deanne Fitzmaurice.)

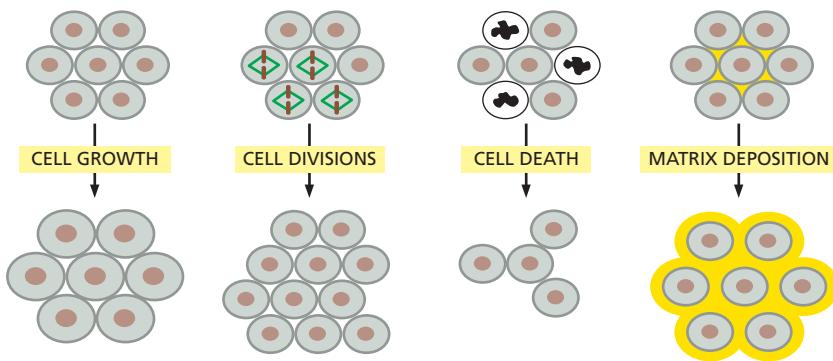


Figure 21–58 Determinants of organ size.

The Proliferation, Death, and Size of Cells Determine Organism Size

The nematode worm *C. elegans* illustrates the different ways in which size differences can arise. This creature follows an astonishingly precise and predictable developmental program. Each individual of a given sex is generated by almost exactly the same sequences of cell divisions and cell deaths, and consequently has precisely the same number of somatic cells—959 in the adult hermaphrodite (the sex of the majority of these animals)—although the number of germ cells is more variable from worm to worm. The stereotyped development makes it possible to trace somatic cell lineages in exhaustive detail. More than 1000 cell divisions generate 1090 somatic cells during hermaphrodite development, but 131 of these cells undergo apoptotic cell death. Thus, precise regulation of both cell division and cell death determines the final numbers of somatic cells in the worm. In fact, genetic screens in *C. elegans* identified the first genes responsible for apoptosis and its regulation—thereby revolutionizing our molecular understanding of this form of programmed cell death (discussed in Chapter 18).

The final number of somatic cells in the adult worm is already present at sexual maturity (around three days after fertilization), after which no more somatic cells are generated. Yet the worm continues to grow, doubling in size between sexual maturity and death 2–3 weeks later. This doubling results from somatic cell growth: although the cells no longer divide, they continue to go through rounds of DNA synthesis; this *endoreplication* of the genome makes the cells *polyploid*. As in all organisms, the size of a cell is proportional to its ploidy—that is, the number of genome copies that it contains: a doubling of ploidy roughly doubles cell volume. By artificial manipulation of somatic cell ploidy, and thereby somatic cell size, the size of the worm as a whole can be increased or decreased. Thus the worm's final size is set by a combination of programmed cell divisions and cell deaths, along with regulation of the sizes of individual cells through changes in ploidy.

In plants, as in animals, cell size increases as ploidy increases (Figure 21–59). This effect has been exploited in the agricultural breeding of plants for large size: most of the major fruits and vegetables that we consume are polyploid.

Animals and Organs Can Assess and Regulate Total Cell Mass

The size of an animal or organ depends on both cell number and cell size—that is, on total cell mass. Remarkably, many animals and organs can somehow assess their total cell mass and regulate it, providing evidence for feedback controls of the sort highlighted earlier in our introductory account of general principles of growth control. In contrast with *C. elegans*, if cell size is artificially increased or decreased in these cases, cell numbers adjust to maintain a normal total cell mass. This has been beautifully illustrated by experiments done long ago in salamanders, where cell size can be manipulated by altering the animal's ploidy. As shown in Figure 21–59E, salamanders of different ploidies end up being the same size with very different numbers of cells. The individual cells in a pentaploid salamander,

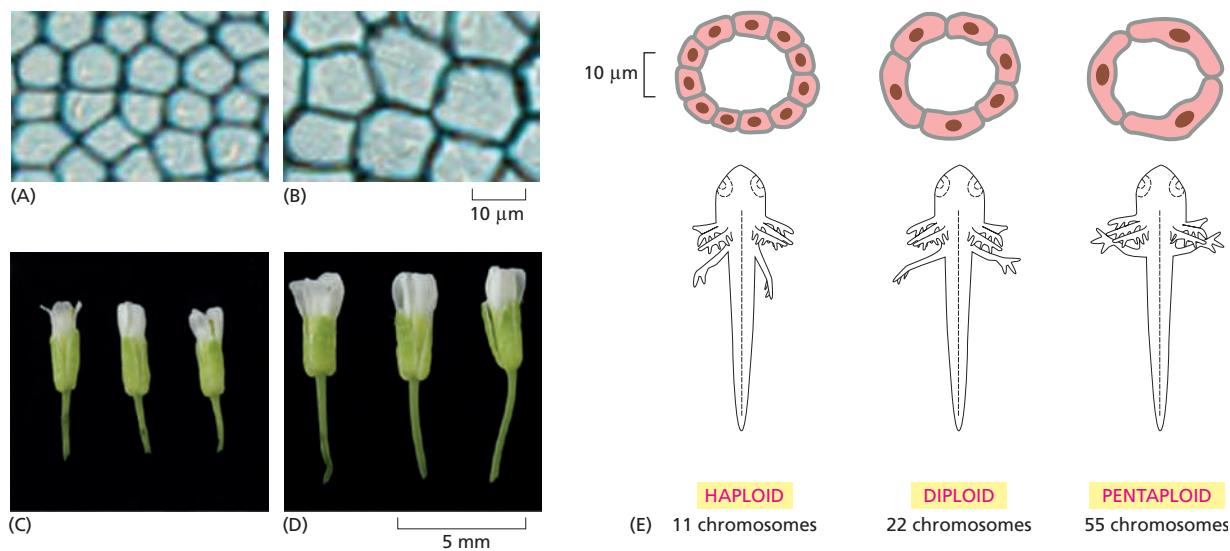


Figure 21–59 Effects of ploidy on cell size and organ size. In all organisms, from bacteria to humans, cell size is proportional to ploidy—the number of copies of the genome per cell. This is illustrated for (A–D) *Arabidopsis* flowers and (E) for salamanders. In each case, the upper panels show cells in a specific tissue [a petal for *Arabidopsis*, a pronephric (kidney) tubule for the salamander]; the lower panels show the gross anatomy—flowers for *Arabidopsis*, the whole body for the salamander. In the case of *Arabidopsis* flowers, increase in cell size increases organ size. By contrast, the salamander and its individual organs attain their normal standard size regardless of ploidy, because large cell size is compensated for by fewer cells. This indicates that the size of an organism or organ in this species is not controlled simply by counting cell divisions or cell numbers; size must somehow be regulated at the level of total cell mass. [A–D, from C. Breuer et al., *Plant Cell* 19:3655–3668, 2007. With permission from the American Society of Plant Biologists; E, adapted from G. Fankhauser, in *Analysis of Development* (B.H. Willier, P.A. Weiss and V. Hamburger, eds), pp. 126–150. Philadelphia: Saunders, 1955.]

for example, are about five times the size of those in a haploid salamander, but there are only one-fifth as many cells. This scaling operates not only in the body as a whole, but in its individual organs.

The **imaginal discs** of *Drosophila* provide another striking example of homeostatic size control. These are epithelial pouches that grow by cell proliferation during the larval period and, during the pupal stage, form the organs and extremities of the adult fly (Figure 21–60). Experiments have been chiefly done on the wing imaginal disc. Mutations in components of the cell-cycle control machinery can be used to speed up or slow down the rate of cell division in the disc. Remarkably, such mutations can result in an excessive number of abnormally small cells

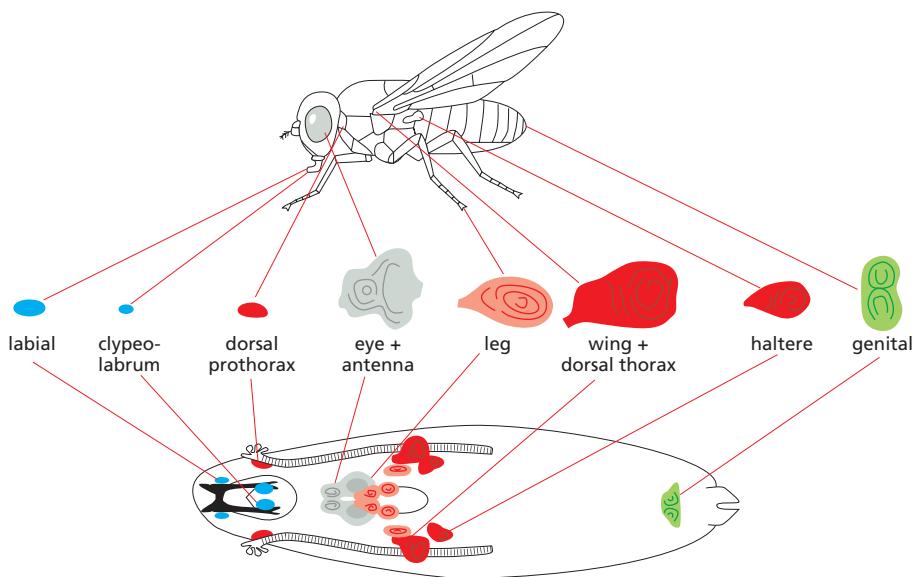


Figure 21–60 The imaginal discs in the *Drosophila* larva (below) and the structures in the adult (above) that they give rise to. [After J.W. Fristrom et al., in *Problems in Biology: RNA in Development* (E.W. Hanley, ed.), p. 382. Salt Lake City: University of Utah Press, 1969.]

Figure 21–61 Pituitary dwarf and pituitary giant. The “giant” on the right is Robert Wadlow (1914–1940), the tallest recorded man at 8 feet 11 inches (2.72 m), together with his father, who was almost 6 feet tall (1.82 m). The dwarf on the left is General Tom Thumb, which was the stage name of Charles Sherwood Stratton (1838–1883). On his 18th birthday, he was measured at 2 feet 8.5 inches (82.6 cm) tall, and at his death, he was 3 feet 4 inches (102 cm). (Images from http://en.wikipedia.org/wiki/File:Robert_Wadlow.jpg. © Bettmann/CORBIS.)

or a reduced number of abnormally large cells, respectively, leaving the size (area) and patterning of the adult wing practically unchanged. Thus, the size of the disc is not regulated so as to contain a set number of cells. Instead, there must be a regulatory mechanism that halts growth when the disc’s total cell mass reaches the appropriate value, so that the size and pattern of the adult wing that develops from the disc are normal. Remarkably, developing discs—or even disc fragments, taken out of their normal context and transplanted into the abdomen of an adult female—will grow until they reach their normal size. Clearly, the mechanisms that regulate disc size are intrinsic to the disc.

We still have very little idea how organisms or organs assess their total cell mass or monitor their own growth. Nevertheless, we are beginning to understand some of the signal molecules that drive or halt growth in response to the mysterious cues that convey information about the size attained.

Extracellular Signals Stimulate or Inhibit Growth

We have already seen how some signals act systemically as hormones to regulate the development of the animal as a whole. Some of these serve to regulate growth. In mammals, for instance, **growth hormone (GH)** is secreted by the pituitary gland into the bloodstream and stimulates growth throughout the body: excessive production of growth hormone leads to gigantism, and too little leads to dwarfism (**Figure 21–61**). Pituitary dwarfs have bodies and organs that are proportionately small, unlike achondroplastic dwarfs, for example, whose limbs are disproportionately short, usually because of a mutation in a gene encoding an FGF receptor that disrupts normal cartilage development (**Figure 21–62**).

Growth hormone stimulates growth largely by inducing the liver and other organs to produce insulin-like growth factor 1 (IGF1), which acts mainly as a local signal within many tissues to increase cell survival, cell growth, cell proliferation, or some combination of these, depending on the cell type. Large breeds of dogs such as Great Danes owe their great size to high levels of IGF1, while miniature breeds such as Chihuahuas have low levels (see Figure 21–57).

Not all growth-regulating extracellular signals stimulate growth; some inhibit it, by promoting cell death or inhibiting cell growth, cell division, or both. *Myostatin* is a TGF β family member that specifically inhibits the growth and proliferation of myoblasts—the precursor cells that fuse to form the huge, multinucleated cells of skeletal muscle. When the *Myostatin* gene is deleted in mice, muscles grow to be several times larger than normal. Remarkably, two breeds of cattle that were bred for large muscles have both turned out to have mutations in the *Myostatin* gene; whippet dogs mutant for *Myostatin* develop similarly (**Figure 21–63**).

Figure 21–62 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.)

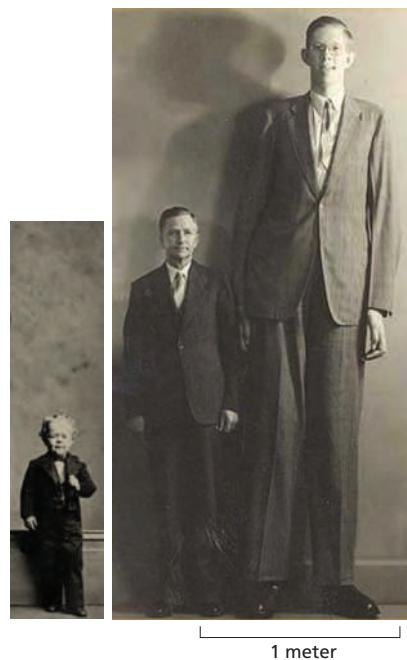




Figure 21–63 Myostatin limits muscle growth. A wild-type whippet dog and a bully whippet that lacks myostatin. (A, from <http://www.merlinanimalrescue.co.uk/dogs/?m=201211>; B, from <http://animalslook.com/schwarzenegger-dog/>.)

Like TGF β itself, myostatin acts through the Smad intracellular signaling pathway (see Figure 15–57) to inhibit muscle growth specifically. Another intracellular signaling pathway, called the *Hippo pathway*, inhibits organ and organism growth more generally. It was discovered in *Drosophila*, but it operates in vertebrates as well. It inhibits growth both by promoting cell death (by blocking an apoptosis inhibitor) and by inhibiting cell-cycle progression (by inhibiting the expression of the cell-cycle gene *Cyclin E*). Some components of the pathway in *Drosophila* are shown in Figure 21–64. The organs of animals that are abnormally resistant to Hippo repression can grow to a monstrous size (Figure 21–65).

It is important to note that in all species nutritional conditions also play a fundamental part in regulating the pace and extent of growth, and in animals they do so through hormonal signal networks that are highly conserved between vertebrates and invertebrates. Although we do not have space for details here, genetic experiments, especially in *Drosophila*, have begun to unravel the logic of these controls, and to indicate how they may operate alongside other machinery, such as the Hippo pathway, to determine final size.

Summary

The sizes of animals and their organs vary widely and largely depend on total cell mass. This in turn depends on the size and number of cells, which are increased through cell growth and cell division, respectively. Cell numbers are reduced by programmed cell death. Each of these processes depends on both intracellular and

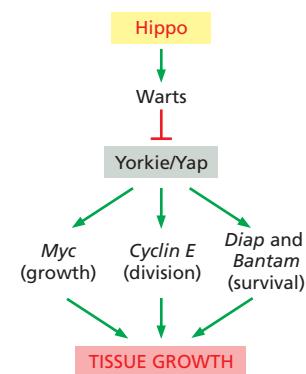


Figure 21–64 Hippo pathway. Hippo, a protein kinase, limits growth by phosphorylation and activation of the kinase Warts, which in turn phosphorylates and inactivates the transcriptional coactivator Yorkie/Yap (called Yap in vertebrates). When unphosphorylated, Yorkie/Yap drives tissue growth: it activates the transcription of the growth-promoting gene *Myc*, the cell-cycle progression gene *Cyclin E*, the anti-apoptotic gene *Diap*, and the microRNA *Bantam*. Hippo-induced phosphorylation of Yorkie/Yap blocks this effect.

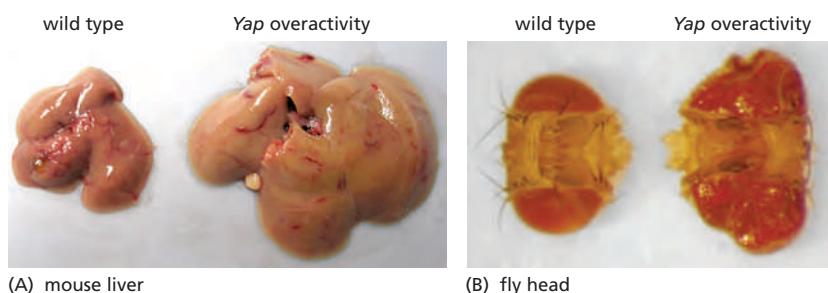


Figure 21–65 Overcoming Hippo repression increases organ size. (A) Livers from control and Yap-overexpressing mice. In these mice, Hippo signaling is insufficient to block Yap. (B) Adult heads from control and Yap-overexpressing flies. In the mutant flies, Hippo signaling is unable to block Yap. (From J. Dong et al., *Cell* 130:1120–1133, 2007. With permission from Elsevier.)

extracellular signals. The mystery is how these processes are regulated and coordinated to produce and maintain the characteristic final size of the adult organ or animal.

Some signals such as survival factors, growth factors, and mitogens stimulate growth by promoting cell survival, cell growth, and cell division, respectively, while other signal molecules do the opposite. Although most of these signals operate locally to help sculpt the size and shape of the animal, its organs, and appendages, others act as hormones to regulate the growth of the animal as a whole. Nutrients can regulate growth through hormonal signals in the entire body.

Many animals and organs can, by unknown mechanisms, assess their total cell mass and regulate it. If, for example, cell size is artificially increased or decreased in these cases, cell numbers adjust to maintain a normal total cell mass. Conversely, if cell numbers are artificially increased or decreased, cell size adjusts to compensate.

NEURAL DEVELOPMENT

The development of the nervous system poses problems that have little parallel in other tissues. A typical nerve cell, or neuron, has a structure unlike that of any other class of cells, with a long axon and branching dendrites, both of which make many synaptic connections to other cells (Figure 21–66). 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 neural network—an electrical signaling system—that functions correctly to guide behavior (Figure 21–67). 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, because the brain performs its computations in a different way and is more tolerant of vagaries in individual components. But the human 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. Thus, in the first phase of neural development, the different parts of the nervous system 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 transcription regulators, by mechanisms of the types we have already discussed. In the next phase, newborn neurons extend axons and dendrites along specific routes toward their target cells, guided by extracellular signals that attract or repel them. In the third phase, neurons form synapses with other neurons or muscle cells, setting up a provisional but orderly network of connections. In the final phase, which continues into adult life, the synaptic connections are adjusted and refined through mechanisms that usually depend on synaptic signaling between the cells involved

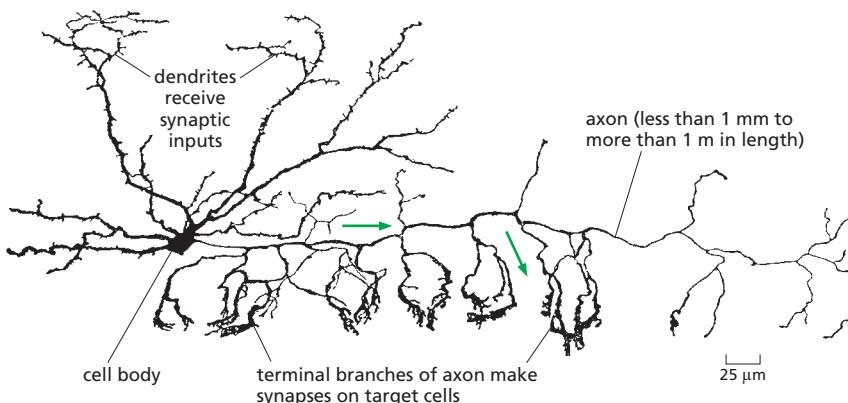


Figure 21–66 A typical neuron of a vertebrate. The arrows indicate the direction in which signals are conveyed. The neuron shown is a basket cell, a type of neuron in the cerebellum. (Adapted from S. Ramón y Cajal, *Histologie du Système Nerveux de l'Homme et des Vertébrés*, 1909–1911. Paris: Maloine; reprinted, Madrid: C.S.I.C., 1972.)

Figure 21–67 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.)

(**Figure 21–68**). At all stages, neurons are in intimate contact with various types of non-neuronal supporting cells—the **glial cells**.

Neurons Are Assigned Different Characters According to the Time and Place of Their Birth

We start our account here with the first phase of neural development: the generation of neural progenitors and their differentiation into hundreds of different neuronal subtypes, along with a much smaller number of glial types. Although the nervous system is exceptional in the extent of cell diversity, the process depends on the same principles that generate different cell types in other organs. We have already discussed some of the underlying machinery in the developing *Drosophila* nervous system. We turn now to vertebrates.

The vertebrate spinal cord, the brain, and the retina of the eye together constitute the central nervous system (CNS). They all originate as parts of the **neural tube**, whose formation was described earlier (see Figure 21–56). The brain and eyes develop from the anterior neural tube and the spinal cord from the posterior.

The developmental anatomy is seen at its simplest in the **spinal cord**. As it develops, the epithelium forming the walls of the posterior neural tube becomes enormously thickened as the cells proliferate and differentiate, creating a highly organized structure of neurons and glial cells, surrounding a small central channel. Bands of neurons with different future functions—and expressing different genes—are laid out along the dorsoventral axis of the tube. Motor neurons (those that control the muscles) are located ventrally, whereas neurons that process sensory information are found dorsally. This pattern is established by opposing gradients of morphogens. These are secreted by specialized groups of cells that run the length of the ventral and dorsal midlines of the neural tube (**Figure 21–69**). The two morphogen gradients—consisting of Sonic hedgehog protein from the ventral source and BMP and Wnt from the dorsal source—help induce different groups of proliferating neural progenitor cells and differentiating neurons to express different combinations of transcription regulators. These regulators in turn drive the production of different combinations of neurotransmitters, receptors, cell-cell adhesion proteins, and other molecules, creating terminally differentiated neurons that will form synaptic connections selectively with the right partners and exchange appropriate signals with them.

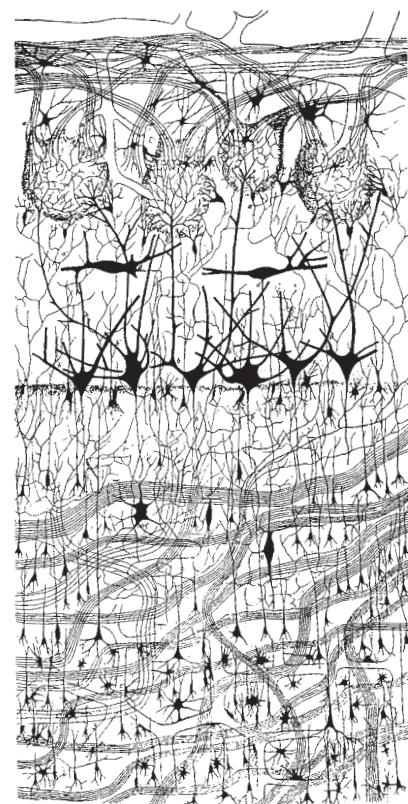
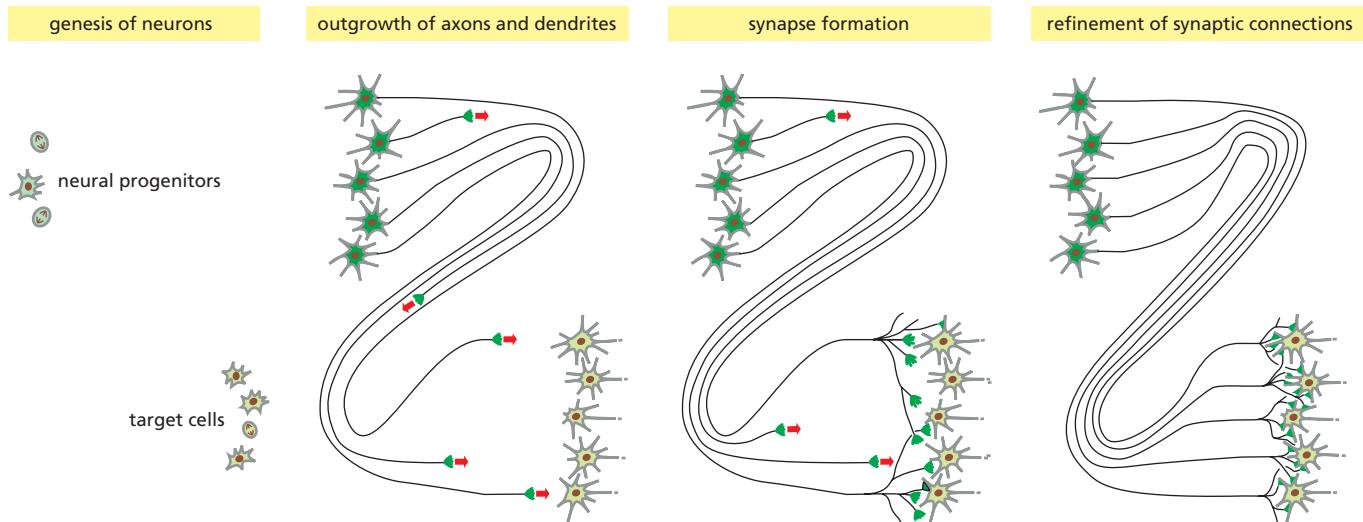


Figure 21–68 The four phases of neural development.



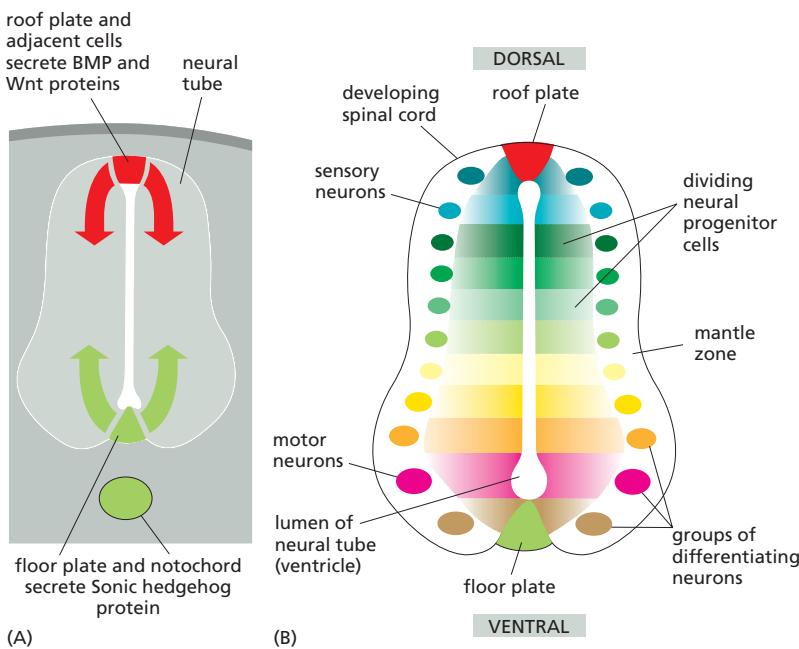


Figure 21–69 A schematic cross section of the spinal cord of a chick embryo, showing how cells at different levels along the dorsoventral axis acquire different characters. (A) Signals that direct the dorsoventral pattern. Sonic hedgehog protein from the notochord and the floor plate (the ventral midline of the neural tube) and BMP and Wnt proteins from the roof plate (the dorsal midline) act as morphogens to control gene expression. (B) The resulting patterns of cell fates in 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 transcription regulators. Neurons expressing different transcription regulators will form connections with different partners and may make different combinations of neurotransmitters and receptors. Colors represent different cell types and combinations of regulatory proteins.

Extracellular morphogen gradients, however, are not the only way to generate cell diversity. As we saw earlier in our discussion of *Drosophila* neuroblasts (see Figure 21–36), different cell types can also be generated by temporal patterning, in which an intracellular program changes the character of a progenitor cell over time, giving rise to different cell types as development progresses. This mechanism also seems to operate in vertebrate neurogenesis. The most striking illustration comes from study of another part of the CNS—the mammalian cerebral cortex.

Although the **cerebral cortex** is the most complex structure in the human body, it has a simple beginning—from the anterior neural tube. As in the spinal cord, the cells that form the walls of the tube proliferate, and the neuroepithelium thickens and expands as they divide. On a predictable schedule, the divisions of the neuroepithelial cells begin to produce a succession of cells committed to terminal differentiation as neurons. These future neurons are born close to the lumen (the central cavity) of the tube. From here, they migrate outward, losing attachment to the luminal surface and crawling outward along neighboring cells that continue to span the full thickness of the neuroepithelium. These latter neuroepithelial cells do double duty, functioning as progenitors of neurons and glia, and as supporters of the epithelial architecture. They become stretched out as **radial glial cells**, forming a scaffold that continues to span the neuroepithelium even as this grows to an enormous thickness (Figure 21–70). At the same time, the radial glial cells continue to divide as neural precursors, giving rise to both neurons and glial cells—new radial glial cells as well as glial cells of other types. The newborn neurons, migrating along the radial glial cells, find their appropriate resting places in the developing cortex, where they mature, and from these sites they send out their axons and dendrites. The first-born neurons settle closest to their birthplace near the lumen, while neurons born later crawl past them to settle farther out (Figure 21–71). The successive generations of neurons thus build up as a series of cortical layers, ordered by birthdate and endowed with different intrinsic characters.

Strikingly, single cortical progenitor cells isolated in culture generate distinct types of cortical neurons and glial cells, with the timing and characteristics appropriate to specific cortical layers. These observations suggest that the neural progenitors in the developing mammalian cortex, much like the *Drosophila* neuroblasts, step through an intracellular developmental program that generates the ordered succession of different nerve cell types.

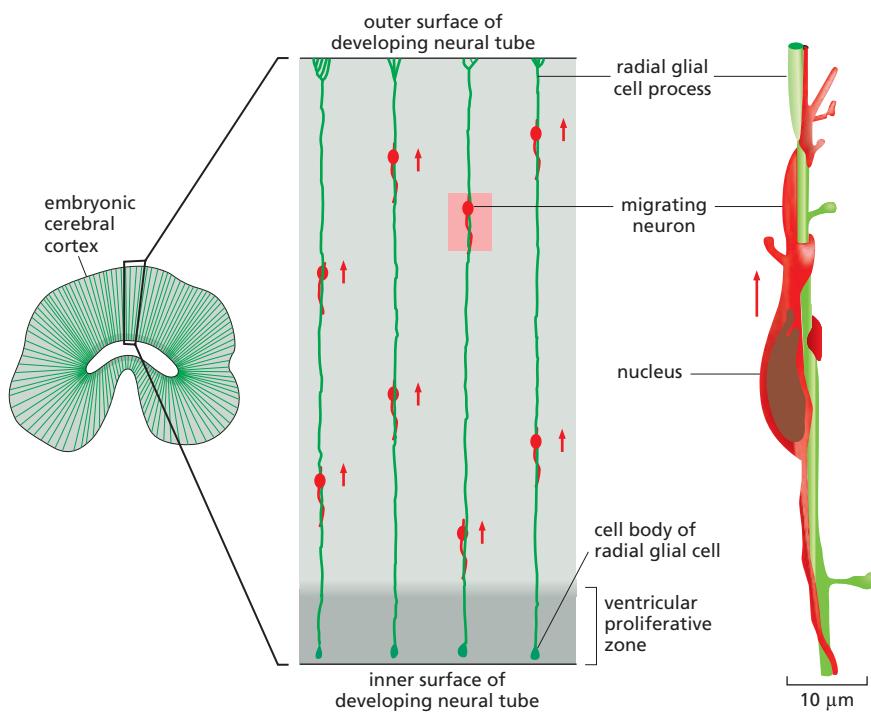


Figure 21–70 Migration of immature neurons. Before sending out axons and dendrites, newborn neurons often migrate from their birthplace and settle in another location. The diagrams are based on reconstructions from sections of the cerebral cortex (part of the neural tube) of a monkey and rely on a staining technique that picks out at random a small subset of the whole dense mass of neuroepithelial cells. The neurons go through their final cell division close to the inner, luminal face of the neural tube (in the ventricular proliferative zone) and then migrate outward by crawling along radial glial cells that form a scaffold. Each of these latter cells extends from the inner to the outer surface of the tube, a distance that may be as long 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. They also serve as neural stem cells: depending on stage and region, the newborn neurons can be generated from radial glial cells that undergo mitosis while their nuclei are close to the inner surface of the tube, or they can be generated from a nearby class of specialized progenitors in the ventricular proliferative zone. (After P. Rakic, *J. Comp. Neurol.* 145:61–84, 1972. With permission from John Wiley & Sons, Inc.)

The Growth Cone Pilots Axons Along Specific Routes Toward Their Targets

According to the character assigned to it during its early development, a neuron will proceed to make connections with specific partners. This phase of neural development involves a type of morphogenesis unique to the nervous system, in which axons and dendrites extend along specific routes toward their target cells. A typical neuron sends out one long axon and many dendrites, which are usually shorter. The axon projects to distant target cells to which the neuron will eventually send signals. The dendrites will receive incoming signals from axon terminals of other neurons. Axons and dendrites extend by growth at their tip, where one sees an irregular, spiky enlargement called a **growth cone** (Figure 21–72 and Movie 21.6). The growth cone is both the engine that produces the crawling movement and the steering apparatus that directs the tip along the proper path. Cytoskeletal machinery in the growth cone creates active protrusions, in the form of filopodia and lamellipodia (see Chapter 16 for details): when such a protrusion

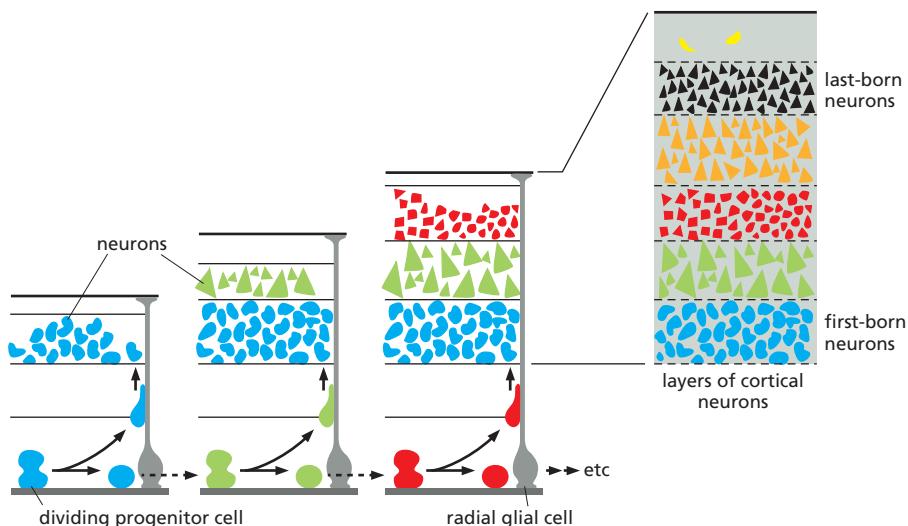
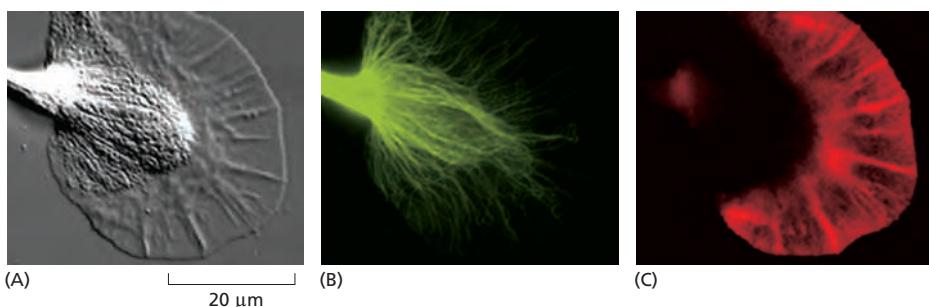


Figure 21–71 Programmed production of different types of neurons at different times from dividing progenitors in the cerebral cortex of the brain of a mammal. Close to one face of the cortical neuroepithelium, progenitor cells divide, in stem-cell fashion, to produce successive generations of neurons (colored here blue, green, red, orange, and black). The neurons migrate out toward the opposite face of the epithelium by crawling along the surfaces of radial glial cells, as shown in Figure 21–70. 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.



contacts an unfavorable surface, it withdraws; when it contacts a more favorable surface, it persists longer, steering the growth cone in that direction. In this way, the growth cone is guided by subtle variations in the properties of the surfaces over which it moves. At the same time, it is sensitive to specific signaling molecules, which—as we discuss next—can either encourage or hinder its advance.

A Variety of Extracellular Cues Guide Axons to their Targets

Growth cones generally travel toward their targets along predictable routes, according to programs stored in the memory of the particular neuron to which they belong ([Movie 21.7](#)). In the simplest case, a growth cone can take a route that has 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 partly mediated by homophilic cell–cell adhesion molecules—membrane glycoproteins that help a cell displaying them to stick to any other cell that displays the same molecules. As discussed in Chapter 19, many homophilic adhesion molecules fall into one of two main classes: they are members of either the immunoglobulin superfamily, such as *N-CAM*, or 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. Growth cones also migrate over components of the extracellular matrix. When tested with neurons growing in a culture dish, some of the matrix molecules, such as laminin, favor axon outgrowth, while others, such as chondroitin sulfate proteoglycans, discourage it. But exactly how the matrix functions to guide axons in intact animals remains to be discovered.

Growth cones are generally guided by a succession of different cues at different stages of their journey, as summarized in [Figure 21-73](#). Many of these cues involve specific signaling molecules. Some of these are encountered in the extracellular matrix, while others are attached to the plasma membrane of cells that the growth cones touch. Another important part is played by chemotactic factors; these are proteins secreted from cells that act as beacons at strategic points along the path—some attracting, others repelling. The trajectory of *commissural axons*—axons that cross from one side of the body to the other—provides a well-studied example.

Commissural axons are a general feature of bilaterally symmetrical animals, such as us, because they are required to coordinate behavior of the two sides of the body. 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 (the same structure that we encountered earlier as a source of the morphogen Sonic hedgehog—see [Figure 21-69](#)). 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 21-74](#)). The first stage of the journey depends on a concentration gradient of the signal protein **Netrin**, secreted by the cells of the floor plate: the commissural growth cones sniff their way toward its source.

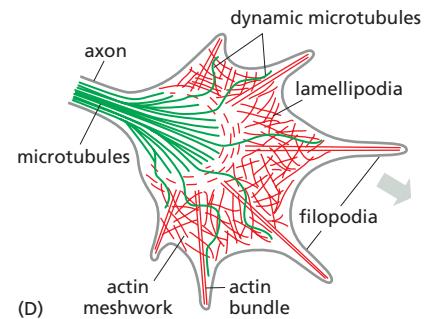


Figure 21-72 Internal architecture of a neuronal growth cone, as seen in culture on a flat substratum. The growth cone forms as an expansion of the tip of the growing axon. (A) Image by interference-contrast microscopy. (B) Immunostaining to show microtubules (green). (C) Immunostaining to show actin filaments (red). (D) Diagram of the cytoskeletal machinery. Filopodia form and push forward by assembly of actin filaments at the leading edge of the growth cone. Microtubules stabilize the directional decisions made by the actin-rich protrusions. Filopodia adhering to the flat substratum contract and pull the growth cone forward. (Images by Chi-Hung Lin, Paul Forscher Laboratory, Yale University, New Haven, CT.)

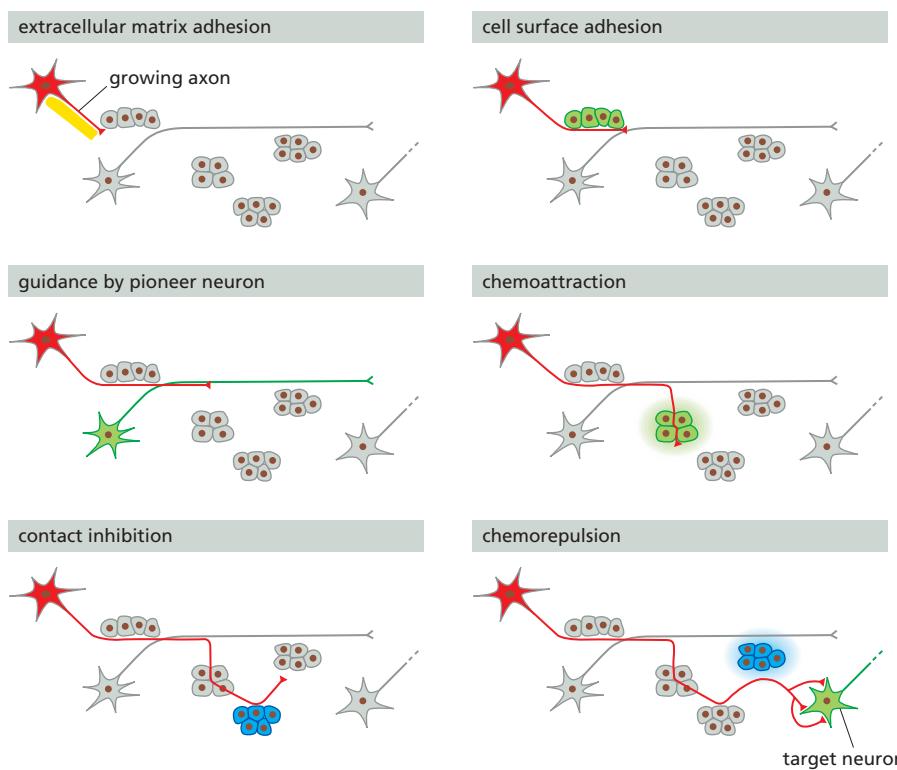


Figure 21–73 Mechanisms of growth-cone guidance. Growth cones use a variety of extracellular cues to navigate to distant targets. They can adhere to the extracellular matrix or to the surfaces of other cells, or they can be repelled by them; they can crawl, for example, by homophilic adhesion along the axons of pioneer neurons; and they can be attracted or repelled by soluble guidance signals. (After E. Kandel et al., Principles of Neural Science, 5th ed., New York: McGraw Hill Medical, 2012.)

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 cross back again? The answers lie in a change in the responsiveness of the growth cones during their journey. As the growth cones cross the midline, they lose sensitivity to Netrin and become sensitive instead to a signal protein called Slit (see Figure 21–74). Slit is also produced by the floor plate, but it has the opposite effect to that of Netrin: it repels the growth

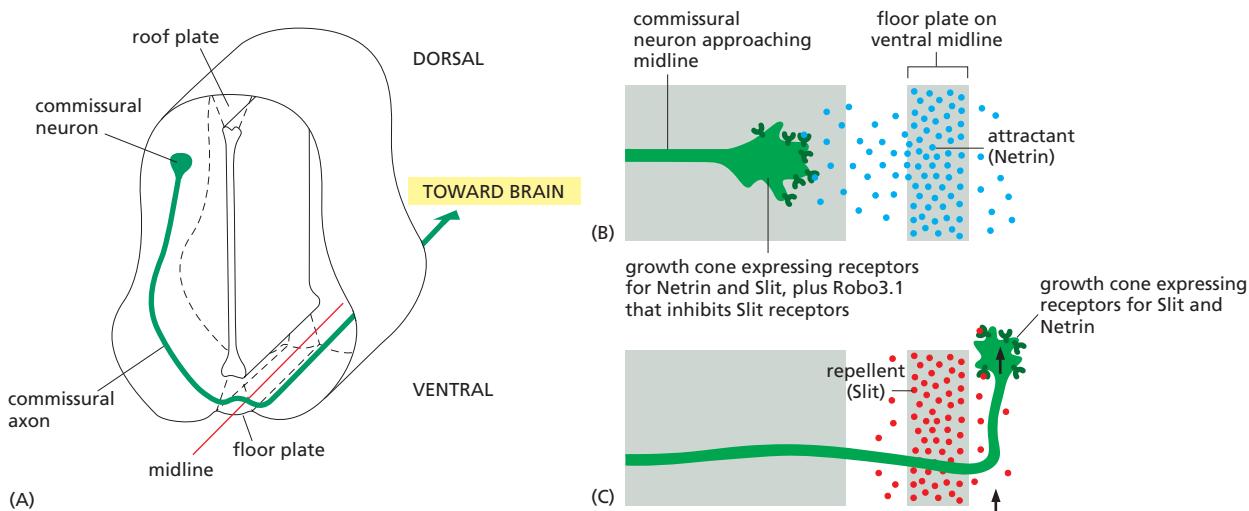


Figure 21–74 The guidance of commissural axons. (A) The pathway taken by commissural axons in the embryonic spinal cord of a vertebrate. (B) Attraction to the midline. The growth cone is first attracted to the floor plate by Netrin, which is secreted by the floor-plate cells and acts on the receptor DCC in the axonal membrane. (C) Repulsion from the midline after crossing it. As the growth cone crosses the floor plate, Slit comes into play: it binds to its receptors Robo1 and Robo2 and acts as a repellent to keep the growth cone from re-entering the floor plate. In addition, it blocks responsiveness to the attractant Netrin. Before crossing the midline, the commissural neurons express Robo3.1, an alternative splice form of Robo3 that is related to Robo proteins but blocks Slit signaling. As neurites cross the midline, Robo3.1 is lost and growth cones become responsive to Slit and are repelled from the midline.

cones, preventing them from re-entering the midline territory. The responses of the growth cone depend on the receptors that it expresses: as commissural neurons approach the floor plate, the Slit receptors are kept inactive by an inhibitory protein (Robo3.1) in the same membrane, allowing the commissural axons to grow to the midline without being repelled. Robo3.1 is lost as the growth cones cross the midline; now the growth cones become sensitive to repulsion by Slit and are thereby prevented from crossing back to the other side. At the same time, signals from the Slit receptors interfere with those from the Netrin receptors, making the growth cones deaf to the signal that attracted them to the floor plate initially. A similar mechanism, using similar proteins, seems to govern midline crossing of commissural axons in other animals, including flies and worms.

The guidance of commissural axons illustrates how axons rarely navigate directly to their targets. Instead, they use intermediate targets, or guideposts, and switch their sensitivities as they move from one local guidepost to the next, steering their way through a complex environment to a far-away destination.

The Formation of Orderly Neural Maps Depends on Neuronal Specificity

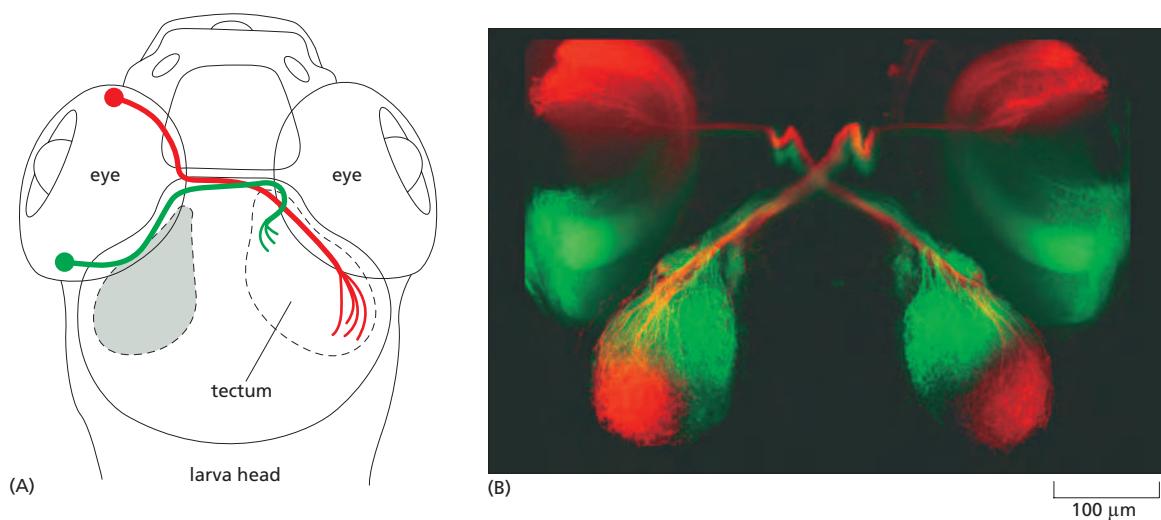
In many cases, neurons of a similar type are laid out in a broad array of different positions, but send out axons that come together for their journey and arrive at the target region in a tight bundle. There the axons disperse again, to terminate at different sites in the target territory. This they do in an orderly way, creating a regular mapping from one territory to another—a neural map.

The axon projection from the eye to the brain provides an important example. The neurons in the retina that convey visual information back to the brain are called *retinal ganglion cells* (RGCs). There are more than a million of them in humans, 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 developing optic nerve toward 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 tectal neurons, the RGC axons distribute themselves in a predictable pattern according to the arrangement of their cell bodies in the retina: RGCs that are neighbors in the retina connect with target cells that are neighbors in the tectum. The orderly projection creates a *retinotopic map* of visual space on the tectum (**Figure 21–75**).

Orderly maps of this sort are found in many brain regions. In the auditory system, for example, the neurons that project from the ear to the brain form a tonotopic 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

Figure 21–75 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.)



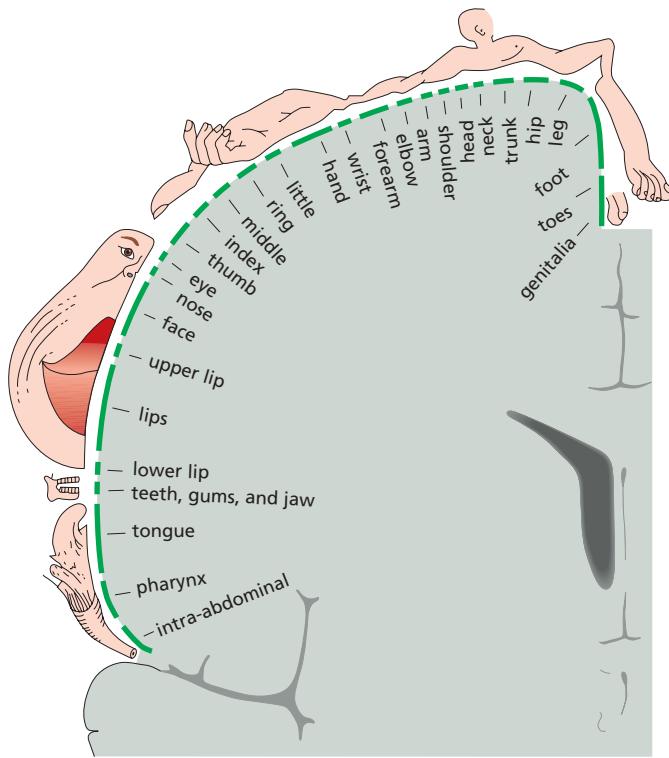


Figure 21–76 A map of the body surface in the human brain. The surface of the body is mapped onto the somatosensory region of the cerebral cortex by using an orderly system of nerve cell connections to pair body sites with the brain sites that receive their sensory information. This means that the map 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.)

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 21–76**).

The retinotopic map of visual space in the optic tectum is the best characterized of all these maps. How does it arise? A famous experiment in the 1940s on frogs provided an important clue. 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, however, the eye is in addition 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 (**Figure 21–77**). If food is dangled in front of it, for example, it will lunge perversely backward. This is because the misplaced retinal cells make the connections appropriate to their original, not their actual, positions. It seems that the retinal ganglion cells (RGCs) have positional values—position-specific biochemical properties representing records of their original location in the retina, assigned perhaps by earlier morphogen gradients, and making RGCs on opposite sides of the retina 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

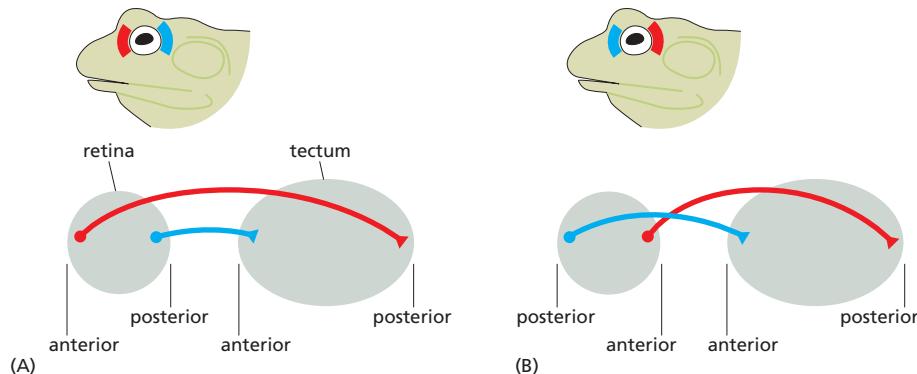


Figure 21–77 Neurons in different regions of the retina project axons to different regions in the tectum. (A) Neurons (RGCs) in the anterior retina project axons to the posterior tectum (as shown in Figure 21–75 for zebrafish). (B) Regeneration experiments show that retinal neurons have an intrinsic preference for the part of the tectum they normally connect to. If the eye is surgically rotated when the optic nerve is cut, the regenerating retinal axons connect to their original targets, creating an inverted map. (After E. Kandel et al., *Principles of Neural Science*, 5th ed., New York: McGraw Hill Medical, 2012.)

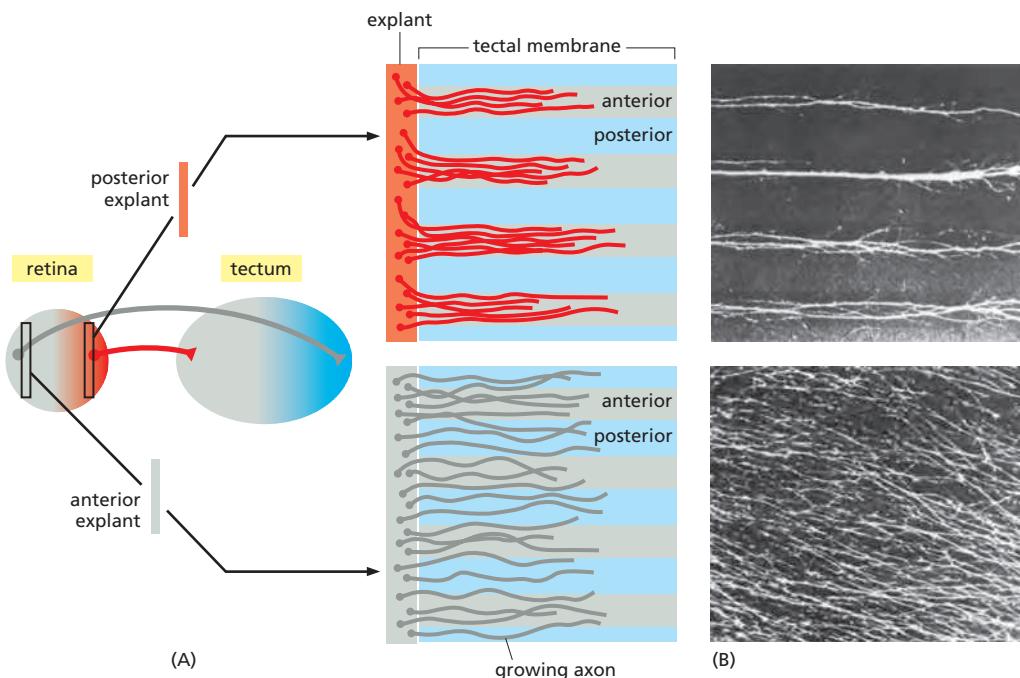


Figure 21–78 Selectivity of retinal axons growing over tectal membranes. (A) Diagram of an experiment performed with cells from a chick embryo. The culture substratum is coated with alternating stripes of membrane prepared either from posterior tectum or from anterior tectum. Axons from posterior retina grow on anterior tectal membrane but are repelled by posterior tectal membrane. Axons from anterior retina show different (less selective) behavior. (B) Photograph of results. The retinal axons, growing out from the left, are made visible by staining them with a fluorescent marker. The selective pattern of outgrowth shows that anterior tectum differs from posterior tectum, and anterior retina correspondingly differs from posterior retina. In the intact organism, this serves to orient a retinotopic map; the map is refined by subsequent competitive interactions among the anterior and posterior retinal axons, which push the anterior retinal cells off anterior tectal territory. (From J. Walter et al., *Development* 101:685–696, 1987. With permission from the Company of Biologists.)

retinal axons because the tectal cells also carry positional labels. Thus, the neural map depends on a correspondence between two systems of positional markers, one in the retina and the other in the tectum.

How are these markers used to make the map? When posterior axons are allowed to grow out over a carpet of anterior or posterior tectal membranes in a culture dish, they show selectivity. Posterior axons strongly prefer the anterior tectal membranes, as *in vivo*, whereas anterior axons show no preference or prefer posterior tectal membranes (Figure 21–78). The key difference between anterior and posterior tectum is not an attractive factor on the anterior tectum but a repulsive factor on the posterior tectum, to which posterior retinal axons are sensitive but anterior retinal axons are not. If a posterior retinal growth cone touches posterior tectal membrane, it collapses its filopodia and withdraws.

In this system, as in others that we have mentioned, the repulsive interactions are mediated by ephrin-Eph signaling—specifically, EphrinA-EphA signaling for the anteroposterior axis (Figure 21–79). An analogous mechanism based on EphB-EphrinB signaling orients the dorsoventral axis of the retinotopic map.

These mechanisms serve to orient the map along both axes, but they are not enough by themselves to ensure accurate point-to-point detail. This is brought about through a long process of adjustment that fills in and refines the map through interactions among the RGC axon terminals as they compete for territory on the tectum. This refinement of the pattern of connections involves electrical signaling in the system of developing synapses—a topic that we return to shortly.

Both Dendrites and Axonal Branches From the Same Neuron Avoid One Another

Axons and dendrites from different neurons can repel one another, or they can cohere; they can collaborate to form synapses, or they can compete. Remarkably,

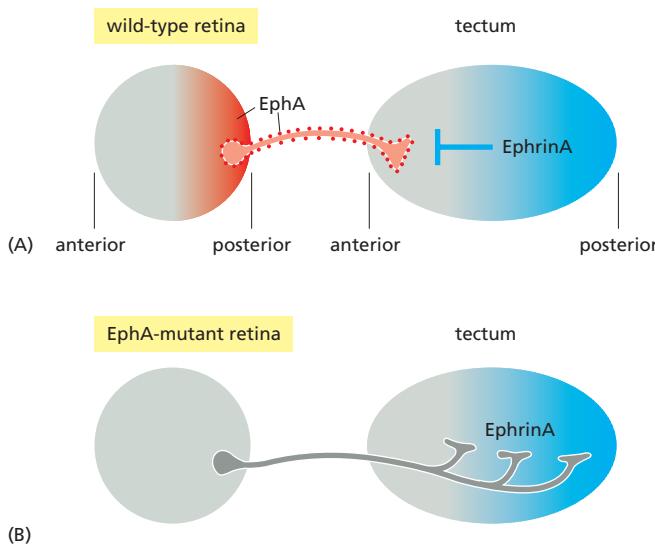


Figure 21–79 Ephrin signaling orients the retinotopic map. (A) Neurons in the posterior retina express EphA. As their axons reach the tectum, they are repelled by high levels of EphrinA protein in the posterior tectum and project preferentially to the anterior tectum. (B) In EphA-mutant mice, posterior retinal axons feel no such repulsion and project more widely within the tectum. (After E. Kandel et al., Principles of Neural Science, 5th ed., New York: McGraw Hill Medical, 2012.)

axons or dendrites can also repel each other when they arise from a single neuron. Such self-avoidance prevents the neuron from making purposeless synapses with itself; it also helps the cell spread out its processes widely so as to innervate a broad territory.

Self-avoidance poses a problem. If the same self-recognition molecule were used in every neuron, all neurons in the brain would repel each other. Some classes of neurons do show this sort of mutual repulsion, creating solitary territories—a phenomenon called *tiling*; but in most cases, axons and dendrites from different neurons can overlap with one another. How then can the processes put out by a single neuron distinguish between self and non-self? This conundrum has been partially resolved by the discovery of a remarkable set of proteins that endow each neuron with a label unlike that of its neighbors. These are the *DSCAM* proteins in *Drosophila* and the *protocadherins* in vertebrates. As described in Chapter 7, DSCAM proteins are extraordinary for the number of isoforms that can be generated by alternative RNA splicing—more than 30,000 variants for DSCAM1 (see Figure 7–57). Diversity arises from alternative exons that code for three highly variable extracellular immunoglobulin domains. Each DSCAM1 isoform engages in homophilic binding (see Figure 19–5), but remarkably, all the variable domains need to be identical for this to occur. Thus, one cell surface will bind to another via DSCAM only when the two cell surfaces express identical isoforms. The result of binding is repulsion, although the detailed mechanisms are poorly understood.

If alternative splicing occurs in a random fashion in each cell, neighboring processes from different neurons are unlikely to express the same DSCAM1 variant, so only the processes of the same cell will repel one another. Neurons that lack all DSCAM1 variants have severe defects in neuronal self-avoidance. Engineering *Drosophila* so that all of its neurons produce a single isoform restores self-avoidance; but now the processes of neighboring neurons express the same isoform and repel each other, resulting in the phenomenon of *tiling* (Figure 21–80).

Vertebrate neurons use a similar self-avoidance strategy to pattern their axons and dendrites, but instead of DSCAMs, they use protocadherins for self/non-self discrimination. The *Protocadherin* locus encodes 58 related cadherin-like transmembrane proteins that are expressed in different combinations in single neurons. Homophilic recognition results in self-avoidance of dendrites emanating from the same neuron; neighboring dendrites of different neurons express different protocadherins and thus evade repulsion. Thus, although insect DSCAM and vertebrate protocadherin proteins share no sequence homology, they mediate similar self-avoidance strategies.

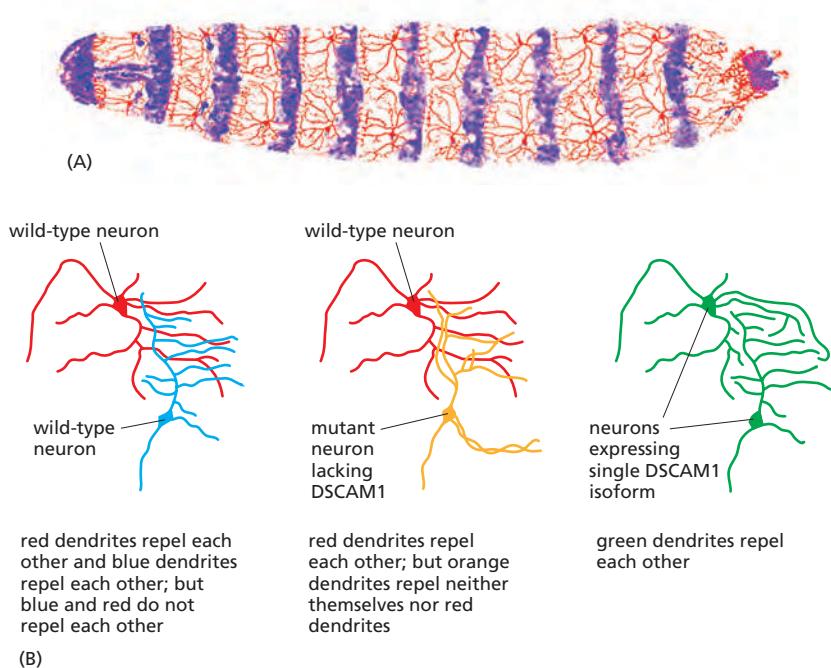


Figure 21–80 DSCAM mediates self-avoidance of dendrites. (A) Sensory neurons in the *Drosophila* peripheral nervous system extend dendrites along the larval body wall. The image shows the dendrites of a regular array of photosensing neurons (red), which allow the larva to detect and avoid harmful light. The posterior epidermal cells of each segment are labeled in blue. There are many neurons, and those shown here spread out their dendrites into overlapping fields. (B) Mutations at the *Dscam* locus upset the way the various dendrites interact, changing the rules of self-avoidance and the distribution of innervation. (A, courtesy of Chun Han; B, after D. Hattori et al., *Annu. Rev. Cell Dev. Biol.* 24:597–620, 2008. With permission from Annual Reviews.)

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. These synapses, as a rule, are destined to transmit neural signals in one direction, from axon to target cell. The development of synapses, however, depends on signaling in both directions: signals from the target tissue not only help control which growth cones synapse where (as we discuss shortly), but can also regulate how many of the innervating neurons survive.

Many types of vertebrate neurons are produced in excess; up to 50% or more of some of them 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 *normal neuronal death* often seems to reflect the outcome of a competition, in which the target tissue releases a limited amount of a specific **neurotrophic factor** that the neurons innervating the tissue require to survive; 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 same neurons all die (Figure 21–81). 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 body surface. The strategy of overproduction followed by death of surplus cells may seem wasteful, but it provides a simple and effective means to adjust the number of innervating neurons according to the amount of tissue requiring innervation.

The first neurotrophic factor to be identified, and still the best characterized, is called *nerve growth factor* (*NGF*)—the founding member of the **neurotrophin** family of signal proteins. It promotes the survival and growth of specific classes of sensory neurons and of sympathetic neurons (a subclass of peripheral neurons that control contractions of smooth muscle and secretion from exocrine glands).

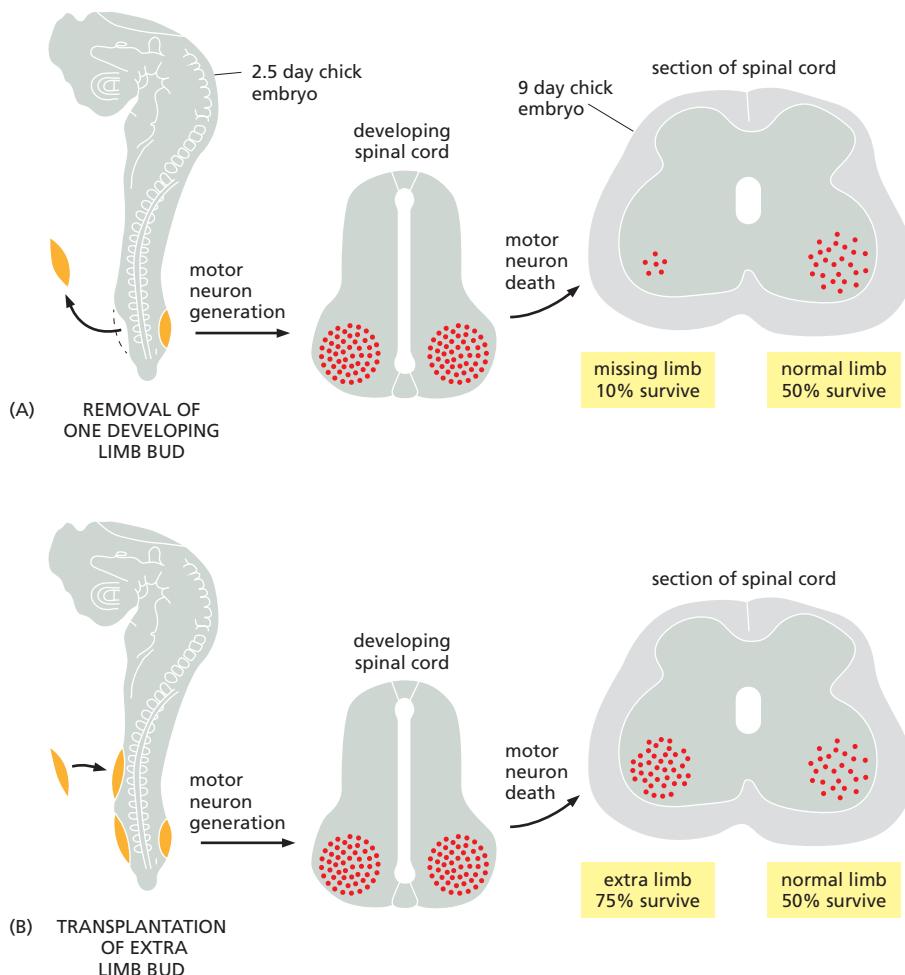
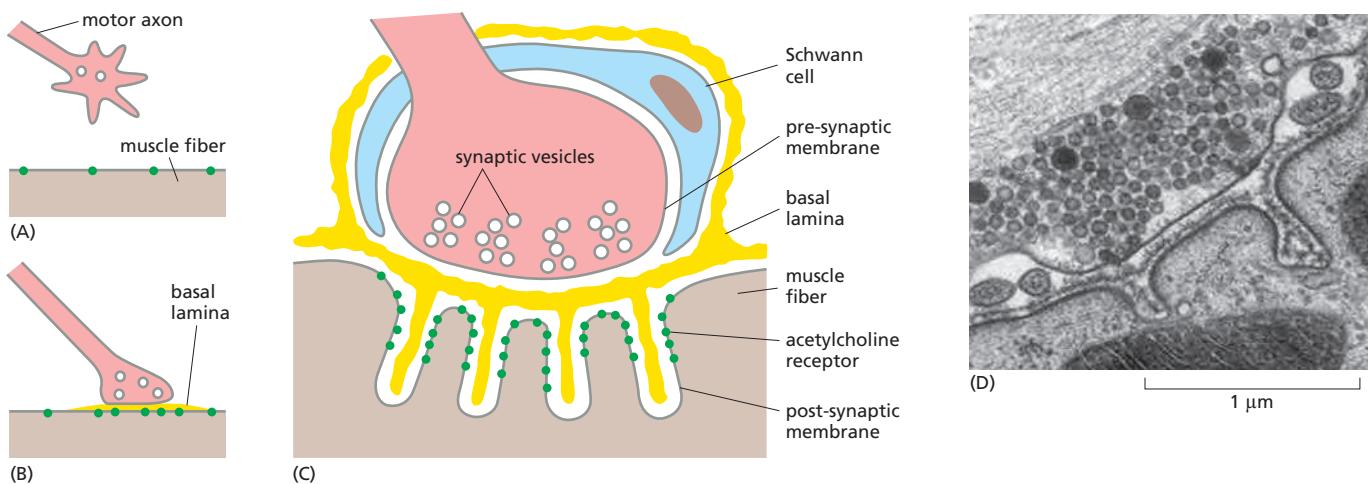


Figure 21–81 The survival of motor neurons depends on signals provided by the target muscles. (A) Removal of the limb bud shortly after arrival of motor axons results in the death of motor neurons in the spinal cord on the amputated side. (B) Transplantation of an extra limb bud increases the survival of motor neurons. (After E. Kandel et al., Principles of Neural Science, 5th ed., New York: McGraw Hill Medical, 2012.)

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 inactivates the gene for NGF or for its receptor (a receptor 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 the survival and growth of different classes of neurons.

Formation of Synapses Depends on Two-Way Communication Between Neurons and Their Target Cells

At journey's end, the task of a growth cone is to halt its travels and make synapses with specific target cells. Synapses were introduced in Chapter 11, where we discussed channels and the electrical properties of membranes. Two main classes of synapses are found in vertebrates; those made with muscle cells and those made with other neurons. Synapse formation is best understood in the case of the highly specialized connections between motor neurons and skeletal muscle cells—so-called **neuromuscular junctions** (see Figure 11–38). During synapse formation, the axonal growth cone differentiates into a *nerve terminal* that contains synaptic vesicles filled with the neurotransmitter acetylcholine, while acetylcholine receptors become clustered in the muscle cell plasma membrane at the site of synapse formation. A synaptic cleft separates the pre- and postsynaptic plasma membranes, and a thin sheet of basal lamina lies in this space between them (Figure 21–82).



Formation of the synapse involves two-way communication between the muscle cell and axonal growth cone: each of them, under the influence of the other, must reorganize the molecules on its side of the junction. The growth cone releases the signal protein **Agrin**, while the muscle expresses the Agrin receptor LRP4. Agrin binding to LRP4 stimulates association of LRP4 with MuSK, a receptor tyrosine kinase. LRP4 also serves as a signal in the reverse direction, from the muscle to the axon (Figure 21-83). During synapse formation, MuSK and LRP4 cluster in the muscle cell plasma membrane in the general neighborhood of the future synapse. As the growth cone approaches, it recognizes LRP4, which stimulates the differentiation of presynaptic structures in the nerve cell. At the same time, Agrin released from the growth cone binds to LRP4 in the muscle cell; this activates MuSK, and promotes a more focused clustering of acetylcholine receptors in the muscle cell membrane. Through these mechanisms, the reciprocal signaling of LRP4 from muscle to growth cone—and of Agrin from growth cone to muscle—induces the coordinated, localized differentiation of pre- and postsynaptic structures.

Synapse formation between neurons in the CNS is far more challenging, both for the neurons and for the scientists trying to understand the molecular basis of its specificity, and it remains poorly understood.

Figure 21–82 Formation of the neuromuscular junction. (A) The growth cone of a motor axon approaches the muscle fiber. (B) Initial synapse formation is characterized by the accumulation of synaptic vesicles at the axon terminal and the formation of a specialized basal lamina in the synaptic cleft. (C) As the neuromuscular junction matures, the synaptic cleft accumulates basal lamina and extracellular matrix proteins, synaptic vesicles cluster at presynaptic release sites, and neurotransmitter receptors cluster at postsynaptic sites. Schwann (glial) cells accompany the motor axon and wrap around its terminus outside the region of synaptic contact. [D, courtesy of John Heuser, from *J. Electron Microsc.* 60 (Suppl 1), 2011. With permission from Oxford University Press.]

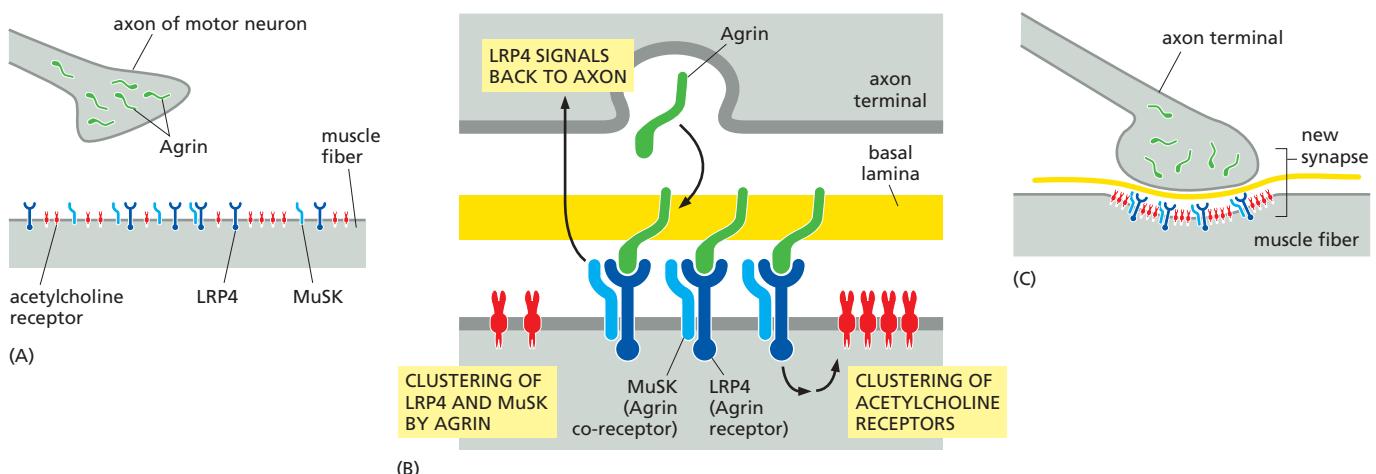


Figure 21–83 Reciprocal signaling during neuromuscular synapse differentiation. (A) The Agrin receptor LRP4 and its co-receptor MuSK cluster in the muscle cell membrane in the general neighborhood of the future synapse. (B) As the growth cone approaches, it recognizes LRP4, which stimulates differentiation of presynaptic structures. Reciprocally, Agrin is released from the nerve terminal, binds to a complex of LRP4 and MuSK in the muscle, and (C) promotes the further and more focused clustering of the LRP4 and acetylcholine receptors in the muscle cell. Although the Agrin/MuSK/LRP4 machinery organizes the synapse, the process also depends on electrical signaling via the acetylcholine receptors. It is not yet known how LRP4 signals to the motor axon.

Synaptic Pruning Depends on Electrical Activity and Synaptic Signaling

The two-way exchange of signals between axon growth cones and muscle cells controls the initial formation of neuromuscular junctions, but it is only the first step in the establishment of the final pattern of the synaptic connections. Each muscle cell at first receives synapses from several motor neurons, but in the end it is left innervated by only one. This process of **synapse elimination** depends on active synaptic communication and electrical activity. If synaptic transmission is blocked by a toxin that binds to the acetylcholine receptors in the muscle cell membrane, or if axonal electrical activity is blocked by a toxin that binds to sodium channels in the axon plasma membrane, the muscle cell retains its multiple innervation beyond the normal time of elimination.

The phenomenon of *activity-dependent synapse elimination* is encountered in almost every part of the developing vertebrate nervous system (Figure 21–84). It has a key role, for example, in the refinement of the retinotopic map discussed earlier. 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. Synapse remodeling during neural development, however, involves more than just synapse elimination; it also involves synapse reinforcement, as we discuss next.

Neurons That Fire Together Wire Together

Throughout the nervous system, and throughout life, activity-dependent elimination and reinforcement of synapses plays a fundamental part in adjusting the detailed anatomy of the neural network according to functional requirements. The importance of these processes, and their underlying rules, emerged half a century ago from a groundbreaking series of experiments on the developing visual system of young mammals.

In the brain of most mammals, axons relaying visual inputs from the two eyes are brought together in a specific neuronal 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 may be 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 form synapses together on shared target neurons in the visual cortex. A period of early electrical signaling activity,

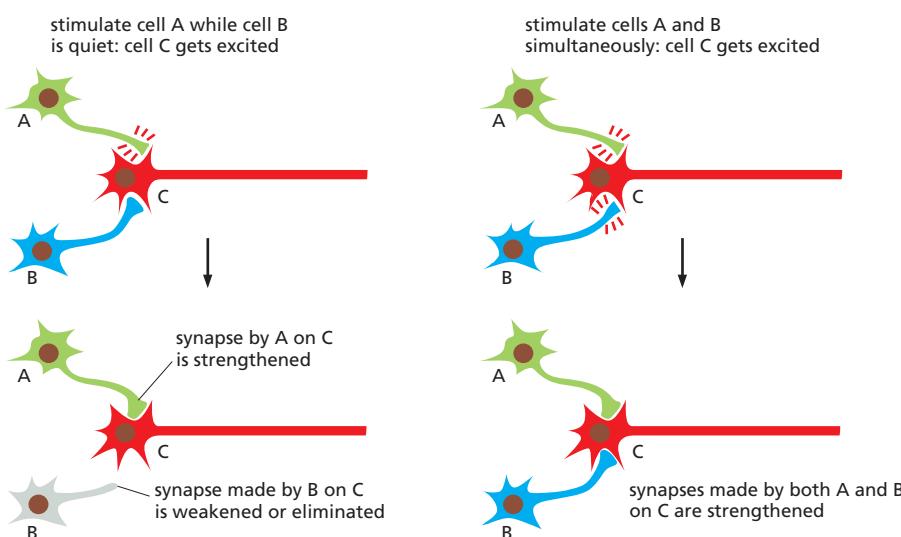


Figure 21–84 Synapse modification and its dependence on electrical activity. Experiments in several systems indicate that synapses are strengthened or weakened by electrical activity according to the rule shown in the diagram. The underlying principle appears to be that each excitation of a target cell tends to weaken any synapse where the presynaptic axon terminal has been quiet, but to strengthen any synapse where the presynaptic axon terminal has just been active. As a result, any synapse that is repeatedly weakened and rarely strengthened is eventually eliminated altogether.

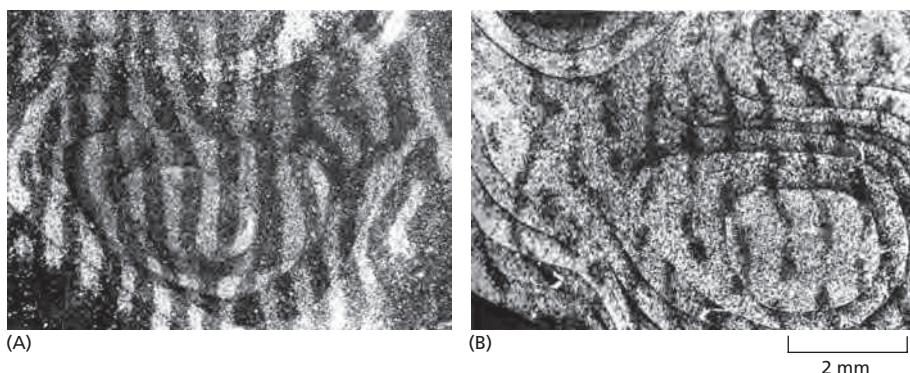


Figure 21–85 Ocular dominance columns in the visual cortex of a monkey's brain, and their sensitivity to visual experience. (A) Normally, stripes of cortical cells driven by the right eye alternate with stripes, of equal width, driven by the left eye. The stripes, set up before birth, are revealed here 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 after birth, during the sensitive 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. LeVay, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 278:377–409, 1977. With permission from The Royal Society.)

however, occurring spontaneously and independently in each retina before birth, leads to a remarkable pattern of *ocular dominance columns* in the visual cortex: stripes of cells driven by inputs from the right eye alternating with stripes driven by inputs from the left eye (Figure 21–85).

The basis for these phenomena became clear from ingenious experiments interfering artificially with visual experience and altering the coordination of electrical signaling in the two eyes. These studies, and many others subsequently, have highlighted a simple but profoundly important principle that seems to govern synapse reinforcement and elimination throughout the nervous system. When two (or more) neurons synapsing on the same target cell fire at the same time, they reinforce their connections to that cell; when they fire at different times, they compete, so that all but one of them tend to be eliminated. This **firing rule** is expressed in the catchphrase “neurons that fire together wire together.”

The firing rule provides a simple interpretation of the developmental phenomenon we have just described in the mammalian visual system. A pair of axons bringing information from neighboring sites in the left eye will frequently fire together, and therefore wire 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 toxins that block axonal electrical activity or synaptic signaling, as described above, the inputs fail to segregate correctly.

The segregation of inputs from the two eyes is only the first of a series of activity-dependent adjustments of visual connections, whose maintenance is extraordinarily sensitive to experience early in life. If, during a certain *sensitive period* (ending at about 5 years of age 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.

Activity-dependent synaptic changes are not confined to early life. They also occur in the adult brain, where many synapses show both functional and morphological alterations with use. This *synaptic plasticity* is thought to have a

fundamental role in learning and memory. Clearly, for the nervous system as for other parts of the body, developmental processes do not end at birth, as we discuss in the next chapter.

Summary

The development of the nervous system proceeds in four phases. First, neurons and glial cells are generated from dividing neural progenitor cells. Then, the newborn neurons send out axons and dendrites toward their targets. Next, they make synaptic connections with appropriate target cells so that communication can begin. Finally, excessive neurons are eliminated by normal neuronal cell death, after which the system of synaptic connections is refined and remodeled according to the pattern of electrical and synaptic activity in the neural network.

Neurons born at different times and places are specialized to express different sets of genes, and they have a cell memory that plays a major role in determining the connections they will form. Their specialization depends not only on spatial patterning by morphogens but also on intrinsic developmental programs that unfold as the neural progenitors proliferate. Axons and dendrites grow out from the neurons by means of growth cones, which follow specific pathways delineated by attractive and repellent signals along the way, including cell-surface and extracellular matrix molecules and soluble signal proteins to which growth cones from different classes of neurons respond differently. In many parts of the nervous system, neural maps are set up—orderly projections of one array of neurons onto another. In the retinotopic 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. Other cell-surface molecules such as DSCAM proteins in Drosophila and protocadherins in vertebrates mediate self-avoidance between the branches arising from a single neuron, helping the cell spread out its processes.

The formation of synapses involves back-and-forth signaling between target cells and the growth cone. After the growth cones have reached their targets and initial connections have formed, individual synapses are eliminated in some places and reinforced in others by mechanisms that depend on synaptic and electrical activity. These mechanisms adjust the architecture of the neural network according to the way in which it is used.

WHAT WE DON'T KNOW

- What regulates the pace of development? Why does a mouse embryo develop faster than a human embryo, for example?
- What are the mechanisms that allow cell memory to be stored during development, explaining how each cell's history determines its future behavior?
- How do signals move through tissues? What are the roles of the extracellular matrix and of elongated cell projections?
- How does a cell know exactly where it is in a multicellular organism? How does it know that its neighbors are the correct ones and that, if not, it should move or kill itself?
- How do cells respond to tiny gradients of molecules in their environment, as required for knowing their positions? How are morphogen gradients reliably interpreted?
- What are the genetic changes that allow the repurposing of existing body parts during evolution? For example, how did bat wings evolve from arms?
- How do cells use genetic instructions to form the shape of something as complex as the human nose?

PROBLEMS

Which statements are true? Explain why or why not.

21–1 In the early cleavage stages, when the embryo cannot yet feed, the developmental program is driven and controlled entirely by the material deposited in the egg by the mother.

21–2 Because of the many later developmental transformations that produce the elaborately structured organs, the body plan set up during gastrulation bears little resemblance to the body plan in the adult.

21–3 As development progresses, individual cells become more and more restricted in the range of cell types they can give rise to.

21–4 At different stages of embryonic development, the same signals are used over and over again by different cells, but with different biological outcomes.

21–5 Changes in the coding regions of genes involved in development are primarily responsible for the differences between species.

21–6 The cell cycle is the ticking clock that sets the tempo of developmental processes, with maturational changes in gene expression being dependent on cell-cycle progression.

Discuss the following problems.

21–7 Name the four processes that are fundamental to animal development, and describe each of them in a single sentence.

21–8 What are the three germ layers formed during gastrulation, and what are the principal structures each gives rise to in the adult?

21–9 In the early *Drosophila* embryo, there seems to be no requirement for the usual forms of cell-cell signaling; instead, transcriptional regulators and mRNA molecules move freely between nuclei. How can that be?

21–10 Morphogens play a key role in development, creating concentration gradients that inform cells of where they are and how to behave. Examine the simple patterns represented by the flags in **Figure Q21–1**. Which do you suppose could be created by a gradient of a single morphogen? Which would require gradients of two morphogens? Assuming that such patterns were present in a sheet of cells, explain how they could be created by morphogens.



Figure Q21–1 National flags from three countries (Problem 21–10).

21–11 Two adjacent cells in the nematode worm normally differentiate into an anchor cell (AC) and a ventral uterine precursor (VU) cell, but which of the two becomes the AC and which becomes the VU cell is completely random: the cells have an equal chance of adopting either fate, but they always adopt different fates. Mutations of *Lin12* alter these fates. In hyperactive *Lin12* mutants, both cells become VU cells, while in inactive *Lin12* mutants, both cells become ACs. Thus, *Lin12* is central to the decision-making process. In genetic mosaics in which one precursor cell has the hyperactive *Lin12* and the other precursor has the inactive *Lin12*, the cell with the hyperactive *Lin12* always becomes the VU cell and the cell with inactive *Lin12* always becomes the AC. Assuming that one cell sends a signal and the other cell receives it, explain how these results suggest that *Lin12* encodes a protein required to receive the signal. Offer a suggestion for how the fates of these two precursor cells are normally decided in wild-type worms.

21–12 It was clear from the early days of studying development that certain “morphogenetic” substances were present in the egg and segregated asymmetrically into cells of the developing embryo. One such investigation in ascidian (sea squirt) embryos examined endodermal alkaline phosphatase, which could be visualized by a histochemical stain. Treatment of embryos with cytochalasin B stopped cell division, but did not block expression of alkaline phosphatase at the appropriate time. Treatment with actinomycin D, which blocks transcription, did not interfere with expression of alkaline phosphatase. Treatment with puromycin, which blocks translation, eliminated expression of alkaline phosphatase. What is the likely nature of the morphogenetic substance that gives rise to alkaline phosphatase?

21–13 The mouse *HoxA3* and *HoxD3* genes are paralogs that occupy equivalent positions in their respective *Hox* gene clusters and share roughly 50% identity in their protein-coding sequences. Mice with defects in *HoxA3* have deficiencies in pharyngeal tissues, whereas mice with defects in *HoxD3* have deficiencies in the axial skeleton, suggesting quite different functions for the paralogs. Thus, it came as a surprise when it was found that replacing a defective *HoxD3* gene with the normal *HoxA3* gene corrected the deficiency, as did the reciprocal experiment of replacing a mutant *HoxA3* gene with a normal *HoxD3* gene. Neither transplanted gene, however, could supply its normal function; that is, a normal *HoxA3* gene at the *HoxD3* locus could not correct the deficiency caused by a mutant *HoxA3* gene at the *HoxA3* locus. The same was true for the *HoxD3* gene. If the *HoxA3* and *HoxD3* genes are equivalent, how do you suppose they can play such distinct roles in development? Why do you suppose they cannot perform their normal function in a new location?

21–14 The segmentation of somites in vertebrate embryos is thought to depend on oscillations in the expression of the *Hes7* gene. Mathematical modeling explains these oscillations in terms of the delays in production of the unstable *Hes7* protein, which acts as a transcription regulator to shut off its own expression. Once *Hes7* decays, with a half-life of about 20 minutes, its transcription resumes. To test this model, you decide to reduce the total delay by removing one, two, or all three of the introns from the *Hes7* gene in mice. Why do you expect that intron removal would reduce the delay? What would you predict would happen to the oscillation time, and somite formation, if the model were correct?

21–15 The oscillatory clock that drives somite formation in vertebrates involves three essential components Her7 (an unstable repressor of its own synthesis), Delta (a transmembrane signaling molecule), and Notch (a transmembrane receptor for Delta). Notch is bound by Delta on neighboring cells, activating the Notch signaling pathway, which then activates *Her7* transcription. Normally, this system works flawlessly to create sharply defined somites (**Figure Q21–2A**). In the absence of Delta, however, only the first five somites form normally, and the rest are poorly defined (**Figure Q21–2B**). If a pulse of Delta is supplied later, somite formation returns to normal in the regions where Delta was present (**Figure Q21–2C**). A diagram of the connections between the components of the clock and how they interact in adjacent cells is shown in **Figure Q21–2D**. In the absence of Delta, why do the cells become unsynchronized? What is it about the presence of Delta that keeps adjacent cells oscillating in synchrony?

21–16 The extracellular protein factor Decapentaplegic (Dpp) is critical for proper wing development in *Drosophila* (**Figure Q21–3A**). It is normally expressed in a narrow stripe in the middle of the wing, along the anterior-posterior boundary. Flies that are defective for Dpp form stunted “wings” (**Figure Q21–3B**). If an additional copy of the gene is placed under control of a promoter that is active in the anterior part of the wing, or in the posterior

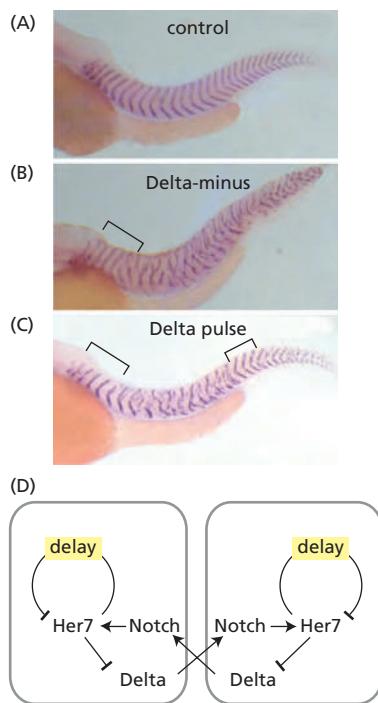


Figure Q21-2 Somite formation in zebrafish embryos (Problem 21–15). (A) Wild-type embryos with normal somites. (B) Somite formation in embryos lacking Delta. The bracket indicates normal-looking somites where they initially form. (C) Somite formation in embryos lacking Delta, but receiving a pulse of Delta expression at the time indicated by the right-hand bracket. (D) Interactions among components of the oscillatory clock in adjacent cells. (Adapted from C. Soza-Ried et al., *Development* 141:1780–1788, 2014. With permission from The Company of Biologists.)

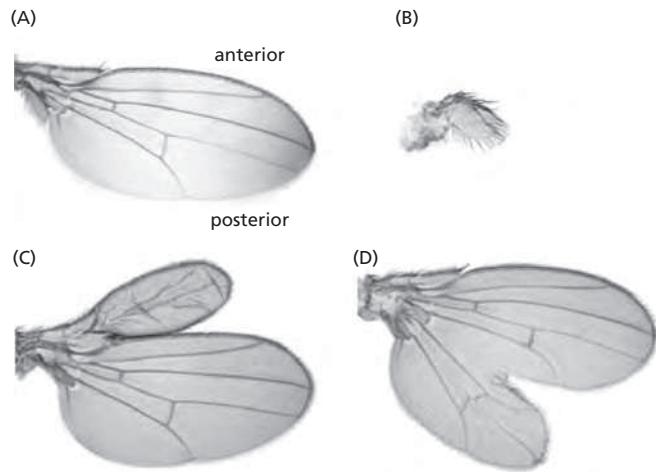


Figure Q21-3 Effects of Dpp expression on wing development in *Drosophila* (Problem 21–16). (A) Normal Dpp expression. (B) Absence of Dpp expression. (C) Additional anterior Dpp expression. (D) Additional posterior Dpp expression. (From M. Zecca, K. Basler and G. Struhl, *Development* 121:2265–2278, 1995. With permission from The Company of Biologists.)

part of the wing, a large mass of wing tissue composed of normal-looking cells is produced at the site of Dpp expression (Figure Q21-3C and D). Does Dpp stimulate cell division, cell growth, or both? How can you tell?

21–17 The highly branched structures of neurons would seem to make it almost inevitable that they should make unproductive synapses with themselves, yet they manage to avoid this outcome very effectively. How is this accomplished in vertebrates?

REFERENCES

General

- Carroll SB (2006) Endless Forms Most Beautiful: The New Science of Evo Devo. New York: W.W. Norton & Co., Inc.
Gilbert SF (2013) Developmental Biology, 10th ed. Sunderland, MA: Sinauer Associates, Inc.
Wolpert L & Tickle C (2010) Principles of Development, 3rd ed. Oxford, UK: Oxford University Press.

Overview of Development

- Gurdon JB (2013) The egg and the nucleus: a battle for supremacy (Nobel Lecture). *Angew. Chem. Int. Ed. Engl.* 52, 13890–13899.
Istrail S & Davidson EH (2005) Logic functions of the genomic cis-regulatory code. *Proc. Natl. Acad. Sci. USA* 102, 4954–4959.
Levine M (2010) Transcriptional enhancers in animal development and evolution. *Curr. Biol.* 20, R754–R763.
Lewis J (2008) From signals to patterns: space, time, and mathematics in developmental biology. *Science* 322, 399–403.
Meinhardt H & Gierer A (2000) Pattern formation by local self-activation and lateral inhibition. *Bioessays* 22, 753–760.
Rogers KW & Schier AF (2011) Morphogen gradients: from generation to interpretation. *Annu. Rev. Cell Dev. Biol.* 27, 377–407.
Shubin N, Tabin C & Carroll S (2009) Deep homology and the origins of evolutionary novelty. *Nature* 457, 818–823.

Mechanisms of Pattern Formation

- Andrey G & Duboule D (2014) SnapShot: Hox gene regulation. *Cell* 156, 856–856.e1.
Baker NE (2011) Proximodistal patterning in the *Drosophila* leg: models and mutations. *Genetics* 187, 1003–1010.
Chan YF, Marks ME, Jones FC et al. (2010) Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* 327, 302–305.
Davis RL, Weintraub H & Lassar AB (1987) Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* 51, 987–1000.
De Robertis EM (2006) Spemann's organizer and self-regulation in amphibian embryos. *Nat. Rev. Mol. Cell Biol.* 4, 296–302.
DiNardo S, Heemskerk J, Dougan S & O'Farrell PH (1994) The making of a maggot: patterning the *Drosophila* embryonic epidermis. *Curr. Opin. Genet. Dev.* 4, 529–534.
Driever W & Nüsslein-Volhard C (1988) A gradient of bicoid protein in *Drosophila* embryos. *Cell* 54, 83–93.
Fowlkes CC, Luengo CL, Keränen VE et al. (2008) A quantitative spatiotemporal atlas of gene expression in the *Drosophila* blastoderm. *Cell* 133, 364–74.
Furman DP & Bukharina TA (2008) How *Drosophila Melanogaster* forms its mechanoreceptors. *Curr. Genomics* 9, 312–323.

- Gaudet J & Mango SE (2002) Regulation of organogenesis by the *Caenorhabditis elegans* FoxA protein PHA-4. *Science* 295, 821–825.
- Halder G, Callaerts P & Gehring WJ (1995) Induction of ectopic eyes by targeted expression of the eyeless gene in *Drosophila*. *Science* 267, 1788–1792.
- Kornberg TB & Roy S (2014) Cytonemes as specialized signaling filopodia. *Development* 141, 729–36.
- Knoblich JA (2010) Asymmetric cell division: recent developments and their implications for tumour biology. *Nat. Rev. Mol. Cell Biol.* 11, 849–860.
- Lander AD (2013) How cells know where they are. *Science* 339, 923–27.
- Lewis EB (1978) A gene complex controlling segmentation in *Drosophila*. *Nature* 276, 565–570.
- Müller P, Rogers KW, Jordan BM et al. (2012) Differential diffusivity of Nodal and Lefty underlies a reaction-diffusion patterning system. *Science* 336, 721–724.
- Nüsslein-Volhard C & Wieschaus E (1980) Mutations affecting segment number and polarity in *Drosophila*. *Nature* 287, 795–801.
- Ringrose L & Paro R (2007) Polycomb/Trithorax response elements and epigenetic memory of cell identity. *Development* 134, 223–232.
- Shulman JM & St Johnston D (1999) Pattern formation in single cells. *Trends Cell Biol.* 9, M60–64.
- von Dassow G, Meir E, Munro EM & Odell GM (2000) The segment polarity network is a robust developmental module. *Nature* 406, 188–192.

Developmental Timing

- Brown DD & Cai L (2007) Amphibian metamorphosis. *Dev. Biol.* 306, 20–33.
- Giraldez AJ, Mishima Y, Rihel J et al. (2006) Zebrafish MIR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* 312, 75–79.
- Isshiki T, Pearson B, Holbrook S & Doe CQ (2001) *Drosophila* neuroblasts sequentially express transcription factors which specify the temporal identity of their neuronal progeny. *Cell* 106, 511–521.
- Lee RC, Feinbaum RL & Ambros V (1993) The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75, 843–854.
- Lewis J (2003) Autoinhibition with transcriptional delay: a simple mechanism for the zebrafish somitogenesis oscillator. *Curr. Biol.* 13, 1398–1408.
- Pourquié O (2011) Vertebrate segmentation: from cyclic gene networks to scoliosis. *Cell* 145, 650–663.
- Song J, Irwin J & Dean C (2013) Remembering the prolonged cold of winter. *Curr. Biol.* 23, R807–R811.
- Wightman B, Hall I & Ruvkun G (1993) Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. *Cell* 75, 855–862.

Morphogenesis

- Green AA, Kennaway JR, Hanna AI et al. (2010) Genetic control of organ shape and tissue polarity. *PLoS Biol.* 8, e1000537.
- Le Douarin NM & Kalcheim C (1999) The Neural Crest, 2nd ed. Cambridge, UK: Cambridge University Press.
- Matis M & Axelrod JD (2013) Regulation of PCP by the fat signaling pathway. *Genes Dev.* 27, 2207–20.
- Ochoa-Espinosa A & Affolter M (2012) Branching morphogenesis: from cells to organs and back. *Cold Spring Harb. Perspect. Biol.* 4, pii: a008243.

- Raz E & Reichman-Fried M (2006) Attraction rules: germ cell migration in zebrafish. *Curr. Opin. Genet. Dev.* 16, 355–359.
- Revenu C & Gilmour DE (2009) MT2.0: shaping epithelia through collective migration. *Curr. Opin. Genet. Dev.* 19, 338–342.
- Simons M & Mlodzik M (2008) Planar cell polarity signaling: from fly development to human disease. *Annu. Rev. Genet.* 42, 517–540.
- Solnica-Krezel L & Sepich DS (2012) Gastrulation: making and shaping germ layers. *Annu. Rev. Cell Dev. Biol.* 28, 687–717.
- Takeichi M (2011) Self-organization of animal tissues: cadherin-mediated processes. *Dev. Cell* 21, 24–26.
- Walck-Shannon E & Hardin J (2014) Cell intercalation from top to bottom. *Nature* 15, 34–48.

Growth

- Andersen DS, Colombani J & Léopold P (2013) Coordination of organ growth: principles and outstanding questions from the world of insects. *Trends Cell Biol.* 23, 336–344.
- Enderle L & McNeill H (2013) Hippo gains weight: added insights and complexity to pathway control. *Sci. Signal.* 6, re7.
- Hariharan IK & Bilder D (2006) Regulation of imaginal disc growth by tumor-suppressor genes in *Drosophila*. *Annu. Rev. Genetics* 40, 335–61.
- Johnston LA (2009) Competitive interactions between cells: death, growth, and geography. *Science* 324, 1679–1682.
- Lawrence PA & Casal J (2013) The mechanisms of planar cell polarity, growth and the hippo pathway: some known unknowns. *Dev. Biol.* 377, 1–8.
- Pan D (2010) The hippo signaling pathway in development and cancer. *Dev. Cell* 19, 491–505.
- Restrepo S, Zartman JJ & Basler K (2014) Coordination of patterning and growth by the morphogen DPP. *Curr. Biol.* 24, R245–R255.

Neural Development

- Burden SJ, Yumoto N & Zhang W (2013) The role of MuSK in synapse formation and neuromuscular disease. *Cold Spring Harb. Perspect. Biol.* 5, a009167.
- Dessaud E, McMahon AP & Briscoe J (2008) Pattern formation in the vertebrate neural tube: a sonic hedgehog morphogen-regulated transcriptional network. *Development* 135, 2489–2503.
- Hubel DH & Wiesel TN (1965) Binocular interaction in striate cortex of kittens reared with artificial squint. *J. Neurophysiol.* 28, 1041–1059.
- Kolodkin AL & Tessier-Lavigne M (2011) Mechanisms and molecules of neuronal wiring: a primer. *Cold Spring Harb. Perspect. Biol.* 3, pii: a001727.
- Luo L & Flanagan JG (2007) Development of continuous and discrete neural maps. *Neuron* 56, 284–300.
- Rakic P (1988) Specification of cerebral cortical areas. *Science* 241, 170–176.
- Reichardt LF (2006) Neurotrophin-regulated signalling pathways. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361, 1545–1564.
- Sanes DH, Reh TA & Harris WA (2011) Development of the Nervous System, 3rd ed. San Diego, CA: Academic Press.
- Sperry RW (1963) Chemoaffinity in the orderly growth of nerve fiber patterns and connections. *Proc. Natl. Acad. Sci. USA* 50, 703–710.
- Zipursky SL & Sanes JR (2010) Chemoaffinity revisited: dscams, protocadherins, and neural circuit assembly. *Cell* 143, 343–353.

Stem Cells and Tissue Renewal

CHAPTER 22

Cells evolved originally as free-living individuals, and such cells still dominate the Earth and its oceans. But the cells that matter most to us, as human beings, are specialized members of a multicellular community. These cells 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 in structure, chemistry, and behavior. There are more than 200 different named cell types in the human body that 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 cells 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 technology used to build multicellular tissues—the devices that bind cells together and the extracellular materials that give them support. But the adult body is not static: it is a structure in dynamic equilibrium, where new cells are continually being born, differentiating, and dying. Homeostatic mechanisms maintain a proper balance, so that the tissue architecture is preserved despite the constant replacement of old cells by new. In this chapter, we focus on these developmental processes that continue throughout life. In doing so, we shall illustrate some of the diversity of specialized cell types and see how they work together to perform their tasks.

We shall examine in particular the role played in many tissues by *stem cells*—cells that are specialized to provide a fresh supply of differentiated cells where these need to be continually replaced, or when they are required in great number for purposes of repair and regeneration. We shall see that while many tissues renew and repair themselves, some others do not; there, lost cells are lost forever, causing deafness, blindness, dementia, and other ills.

In the final section of the chapter, we discuss how stem cells can be generated and manipulated artificially, and we confront the practical question that underlies the current storm of interest in stem-cell technology: 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?

STEM CELLS AND RENEWAL IN EPITHELIAL TISSUES

Among all the self-renewing tissues in a mammal, the champion—for speed at least—is the lining of the small intestine: the long, convoluted portion of the gut tube that is chiefly responsible for absorption of nutrients from the gut lumen. To introduce stem cells, we take the small intestine as our starting point—not only because it renews itself at a greater rate than any other tissue in the body, but also because the molecular mechanisms that control its organization are particularly well understood. It thereby provides a beautiful illustration of the principles of stem-cell systems that have broad applicability.

IN THIS CHAPTER

[STEM CELLS AND RENEWAL IN EPITHELIAL TISSUES](#)

[FIBROBLASTS AND THEIR TRANSFORMATIONS: THE CONNECTIVE-TISSUE CELL FAMILY](#)

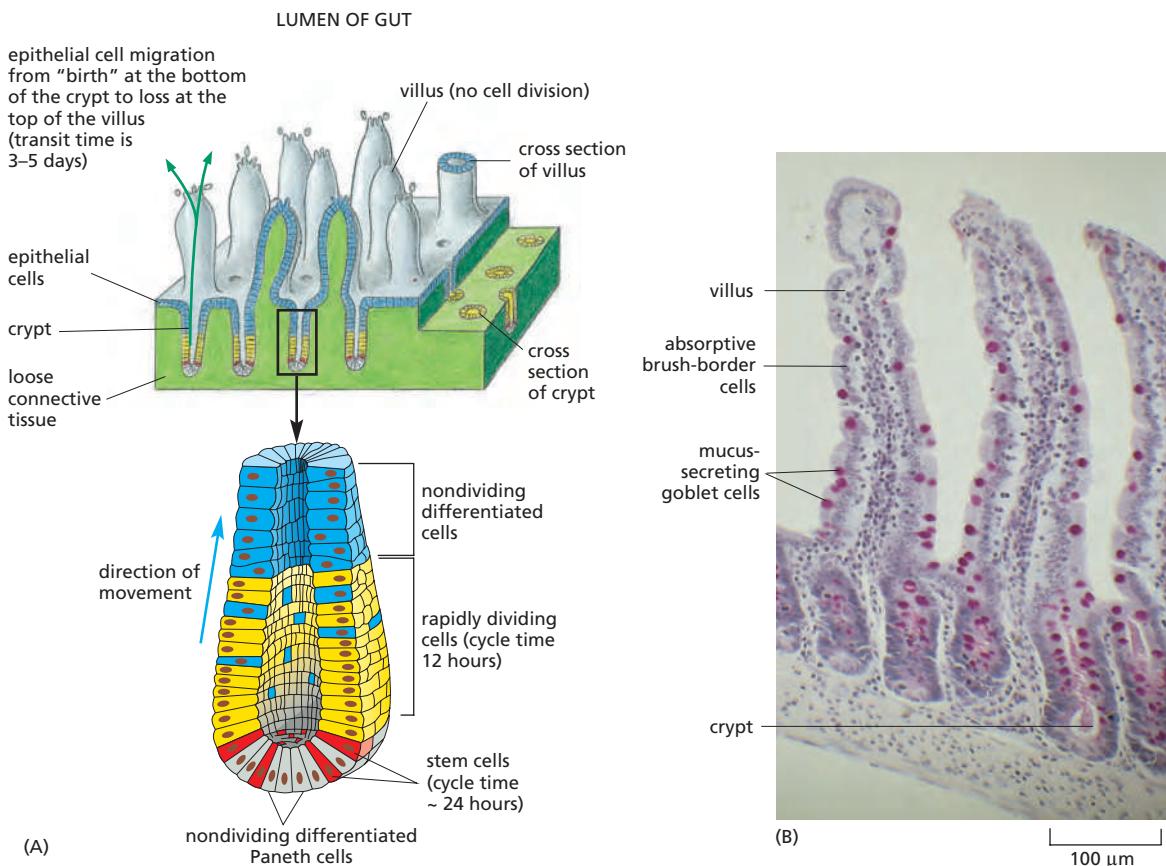
[GENESIS AND REGENERATION OF SKELETAL MUSCLE](#)

[BLOOD VESSELS, LYMPHATICS, AND ENDOTHELIAL CELLS](#)

[A HIERARCHICAL STEM-CELL SYSTEM: BLOOD CELL FORMATION](#)

[REGENERATION AND REPAIR](#)

[CELL REPROGRAMMING AND PLURIPOTENT STEM CELLS](#)



The Lining of the Small Intestine Is Continually Renewed Through Cell Proliferation in the Crypts

The lining of the small intestine (and of most other regions of the gut) is a single-layered epithelium, only one cell thick. This epithelium covers the surfaces of the *villi* that project into the lumen, and it lines the *crypts* that descend into the underlying connective tissue (Figure 22–1). Dividing cells are restricted to the crypts, and differentiated cells, no longer dividing, pour out of the crypts in a steady stream onto the villi. There are four main types of nondividing differentiated cells—one absorptive and three secretory (Figure 22–2):

1. *Absorptive cells* (also called *brush-border cells* or *enterocytes*) have densely packed microvilli on their exposed surfaces. Their job is to take up nutrients from the gut lumen. To this end, they also produce hydrolytic enzymes that perform some of the final steps of extracellular digestion. They are the majority cell type in the epithelium.
2. *Goblet cells* secrete mucus into the gut lumen that covers the epithelium with a protective coat.
3. *Paneth cells* form part of the innate immune defense system (discussed in Chapter 24) and secrete proteins that kill bacteria.
4. *Enteroendocrine cells*, of more than 15 different subtypes, secrete serotonin and peptide hormones 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.

As if on a conveyor belt, the absorptive, goblet, and enteroendocrine cells travel mainly upward from their site of birth in the crypt, by a sliding movement in the plane of the epithelial sheet, to cover the surfaces of the villi. Within 3–5 days (in the mouse) after emerging from the crypts, the cells reach the tips of the villi, where they undergo apoptosis and are finally discarded into the gut lumen (see

Figure 22–1 Renewal of the gut lining. (A) The pattern of cell turnover and proliferation in the epithelium that forms the lining of the small intestine. Stem cells (red) lie at the crypt base, interspersed among nondividing differentiated cells (Paneth cells). Progeny of the stem cells move mainly upward from the crypts onto the villi; after a few quick divisions, they cease dividing and differentiate—some of them while still in the crypt, most of them as they emerge from the crypt. The Paneth cells, like the other nondividing differentiated cells, are continually replaced by progeny of the stem cells, but they migrate downward to the crypt base and survive there for many weeks. (B) Photograph of a section of part of the lining of the small intestine, showing the crypts and villi. Note the mixture of differentiated cell types, all generated from the stem cells; these are primarily absorptive cells, with mucus-secreting goblet cells (stained red) interspersed among them. Enteroendocrine cells (not labeled) are less numerous and less easy to identify without special stains.

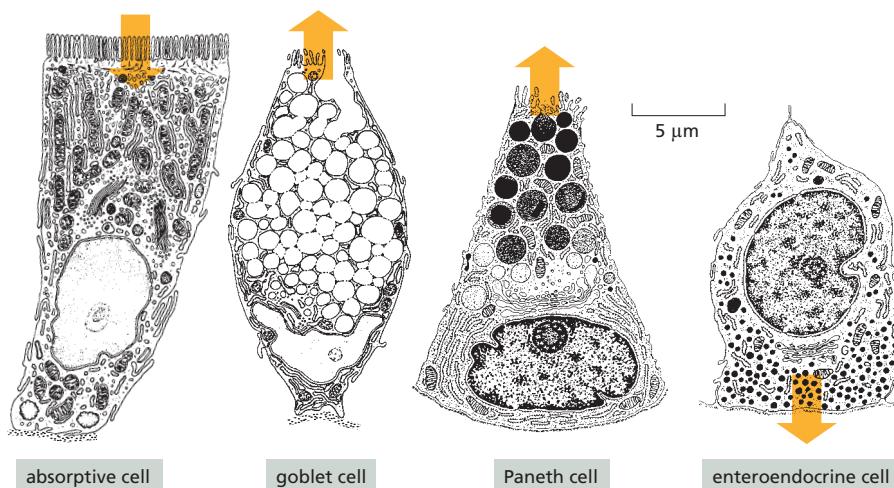


Figure 22–2 The four main differentiated cell types found in the epithelial lining of the small intestine. All cells are oriented with the gut lumen at top. Broad orange arrows indicate direction of secretion or uptake of materials for each type of cell. All of these cells are generated from undifferentiated multipotent stem cells living near the bottoms of the crypts (see Figure 22–1). Absorptive (brush-border) cells outnumber the other cell types in the epithelium by about 10:1 or more. The microvilli on their apical surface 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. Goblet cells secrete mucus; these are the commonest of the secretory cell types. Paneth cells secrete (along with some growth factors) *cryptidins*—proteins of the defensin family that kill bacteria. Different subtypes of enteroendocrine cells secrete serotonin and peptide hormones into the gut wall (and thence the blood). Cholecystokinin is a hormone released from enteroendocrine cells in response to the presence of nutrients in the gut. It binds to receptors on nearby sensory nerve endings, which relay a signal to the brain to stop the feeling of hunger once one has eaten enough. (After T.L. Lentz, Cell Fine Structure. Philadelphia: Saunders, 1971; R. Krstic, Illustrated Encyclopedia of Human Histology. Berlin: Springer-Verlag, 1984.)

Movie 20.6). The Paneth cells in the crypts 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 several weeks (in the mouse) before undergoing apoptosis and being phagocytosed by their neighbors.

The central problem is to understand the processes in the crypt that generate a continual supply of all these nondividing, terminally differentiated cell types.

Stem Cells of the Small Intestine Lie at or Near the Base of Each Crypt

The general pattern of cell proliferation and migration in the gut lining is revealed by a simple labeling method that uses injected pulses of tritiated (radioactive) thymidine or of a thymidine analog that can be detected in tissue sections. Cells that are in S phase of the division cycle incorporate the marker molecule into their DNA, and their fate can then be followed over subsequent hours and days. If a cell divides after incorporation of the label, the label becomes diluted, halving with each cell cycle. This can be quantified. Experiments based on this labeling method confirm, first of all, that dividing cells are confined to the crypts and that the differentiated cell types listed above do not divide. Second, the most rapidly dividing cells, with a cycle time of about 12 hours in the mouse, are shown to lie in the middle and upper parts of the crypt, and these cells are all fated to differentiate and stop dividing (see Figure 22–1A). Just above the base of the crypt, interspersed among the Paneth cells, lie cells that divide more slowly. These are the **stem cells**, which feed some of their progeny into the higher levels of the crypt destined for differentiation, while other progeny remain at the crypt base to continue the whole process. The rapidly dividing cells above these stem cells are derived from them, but already committed to differentiation. These cells are called **committed precursors or transit amplifying cells**, since their divisions serve to amplify the number of differentiated cells that ultimately result from each stem-cell division.

The Two Daughters of a Stem Cell Face a Choice

Stem cells have a critical role in a variety of tissues, and it is useful to list their defining properties:

1. A stem cell 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 22–3**).

Stem cells are required wherever there is a recurring need to replace differentiated cells that cannot themselves divide. Although a stem cell must be able to divide, it does not necessarily have to divide rapidly; in fact, many stem cells divide at a relatively slow rate.

Stem cells are of many types, specialized for the genesis of different classes of terminally differentiated cells—intestinal stem cells for intestinal epithelium, epidermal stem cells for epidermis, hematopoietic 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 conditions serve to keep the stem cell in its proper place and to maintain its stem-cell character? What decides whether a given daughter cell commits to differentiation or remains a stem cell? In a tissue where several distinct types of differentiated cells must be produced, are they all derived from a single type of stem cell, or is there a distinct type of stem cell for each one?

Wnt Signaling Maintains the Gut Stem-Cell Compartment

For the gut, the beginnings of an answer to these questions came from studies of cancer of the colon and rectum (the lower end of the gut, also known as the large intestine). Some people have a hereditary predisposition to colorectal cancer and, in advance of the invasive disease, develop large numbers of small precancerous tumors (adenomas) in the lining of this part of the gut (Figure 22–4). 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. As discussed in Chapter 20, the cause has been traced to mutations in the *Apc* (*adenomatous polyposis coli*) gene: the tumors arise from cells that have lost both gene copies. Because *Apc* codes for a protein that prevents inappropriate activation of the Wnt signaling pathway (see Figure 15–60), this 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 that a cessation of exposure to Wnt signaling normally makes them stop dividing as they leave the crypt.

Stem Cells at the Crypt Base Are Multipotent, Giving Rise to the Full Range of Differentiated Intestinal Cell Types

It has long been suspected that all the differentiated cell types in the lining of the intestine derive from a single type of stem cell. But firm proof was lacking, and the precise nature and location of the stem cells were disputed.

To solve the problem, and indeed to understand the organization of any stem-cell system, we need to discover how its cells are related to one another—who

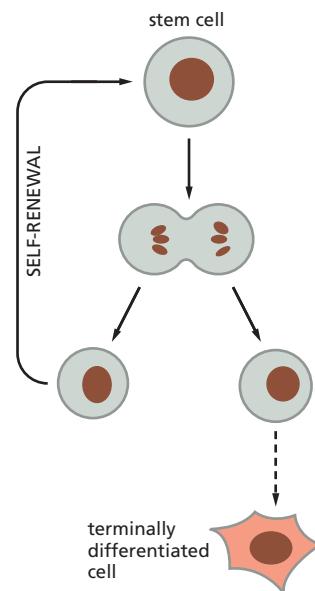


Figure 22–3 The definition of a stem cell. Each daughter produced when a stem cell divides can either remain a stem cell or go on to become terminally differentiated. In many cases, the daughter that opts for terminal differentiation undergoes additional cell divisions before terminal differentiation is completed; such cells are called transit amplifying cells.

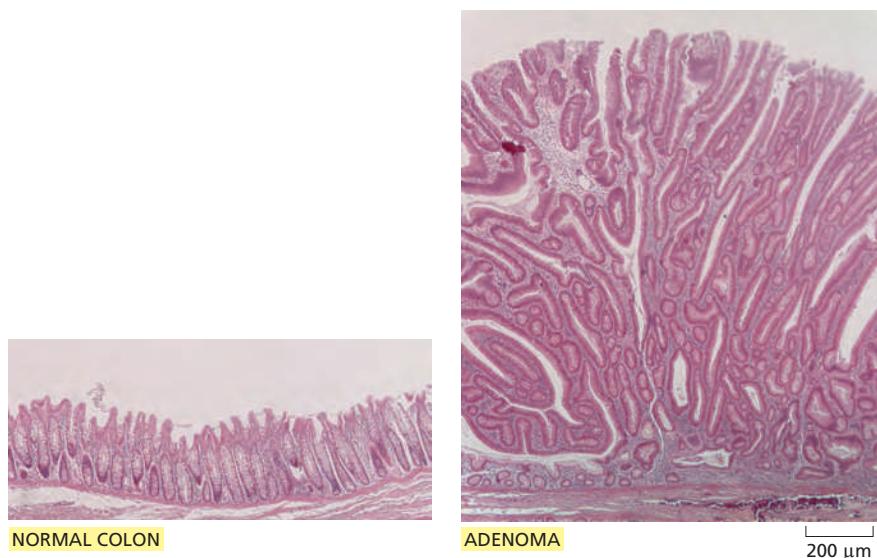


Figure 22–4 An adenoma in the human colon, compared with normal tissue from an adjacent region of the same person's colon. The specimen is from a patient with an inherited mutation in one of his two copies of the *Apc* gene. A mutation in the other *Apc* gene copy, occurring in a colon epithelial cell 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.

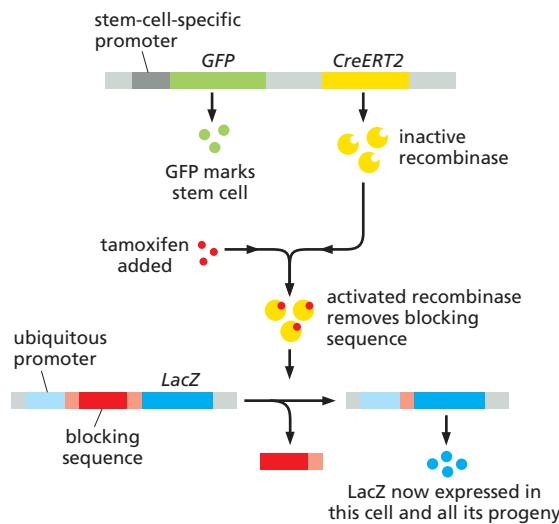


Figure 22–5 Clonal analysis using a genetic marker. A modern method for tracking cell lineage uses transgenic animals containing two transgenes, which together drive expression of a readily detected and heritable marker protein in a small subset of stem cells. The first transgene (top) carries two adjacent protein-coding sequences, *GFP* and *CreERT2*, both expressed under the control of the *Lgr* promoter that is active only in stem cells and not in their differentiated progeny. *GFP* encodes green fluorescent protein (see Chapter 9), which is used here simply to confirm expression in the entire stem-cell population. The *CreERT2* gene encodes a chimeric form of the Cre recombinase called CreERT, which consists of Cre recombinase linked to the estrogen receptor protein; this enzyme becomes active as a recombinase only when it binds the artificial estrogen analog tamoxifen.

The second transgene (bottom) carries a marker gene, *LacZ*, under the control of a promoter that is active in all cells. The *LacZ* gene encodes β-galactosidase, an enzyme that can be detected histochemically in tissues (see Figure 7–28). However, *LacZ* expression in the transgene shown here is prevented by a blocking sequence (red) that is flanked by *LoxP* sites (pink; see Figure 5–66). When tamoxifen is provided, CreERT becomes active—leading to a recombination event that removes the blocking DNA sequence (and leaves one *LoxP* site behind). As a result, the *LacZ* marker is expressed. Because this change is heritable, the marker continues to be expressed in all cells descended from those in which a recombination event has occurred. With a low dose of the inducer molecule tamoxifen, it is possible to activate the marker at random in just a few widely spaced cells, which, in the course of time, give rise to widely separated and easily distinguished clones of progeny (see Figure 22–6).

is descended from whom, or, equivalently, what progeny will be produced from any given cell. This can best be done using a heritable marker that can be activated in an individual cell, thus allowing the identification of the clone of progeny descended from that cell. A modern method uses transgenic animals to create a visible genetic mark in just a few widely spaced cells, which, in the course of time, give rise to widely separated and easily distinguished clones of progeny, as explained in Figure 22–5.

A search among genes that are strongly upregulated in response to Wnt signaling revealed one, called *Lgr5*, that is expressed in gut stem cells specifically. The technique described in Figure 22–5 can be used to create a genetic mark in a random subset of *Lgr5*-expressing cells—a mark that is inherited by the progeny of each cell. These *Lgr5* cells divide with a cycle time of about 24 hours, and within a few days marked clones are seen extending from the crypt bases up along the sides of the villi. After as long as 60 days or more, many of these clones still persist, retaining one or more members at the crypt base and extending all the way up to the tips of the villi (Figure 22–6). Moreover, each single clone typically contains all the major differentiated gut cell types—absorptive, goblet, Paneth, and enteroendocrine—in their normal proportions. The *Lgr5*-expressing cells, therefore, are true stem cells that are *multipotent*—that is, able to generate a diverse set of differentiated cell types.

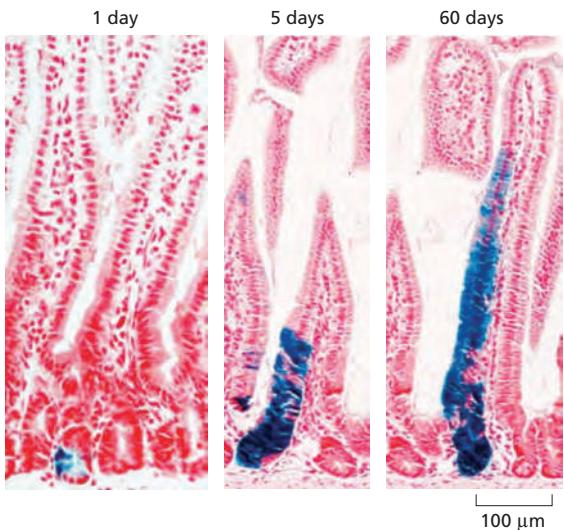


Figure 22–6 *Lgr5*-expressing stem cells and their progeny in the small intestine. The method shown in Figure 22–5 was used here to mark single intestinal stem cells and trace the fates of their progeny. The *Lgr5* gene encodes a member of the family of G-protein-linked transmembrane receptors, and it is expressed specifically in stem cells near the crypt base. Because the *Lgr5* promoter was used to drive expression of *CreERT2*, treatment with a low dose of tamoxifen resulted in occasional stem cells expressing *LacZ*. These cells and all of their progeny could subsequently be detected with a blue histochemical stain. All of the blue cells in these images derive from a single *Lgr5*-expressing stem cell. After 60 days, the blue progeny of this cell are seen to extend all the way up a villus. These progeny can be shown to include all types of differentiated cells, as well as persistent *Lgr5*-expressing cells at the crypt base. This proves that *Lgr5*-expressing cells are multipotent stem cells. (From N. Barker et al., *Nature* 449:1003–1007, 2007. With permission from Macmillan Publishers Ltd.)

The Two Daughters of a Stem Cell Do Not Always Have to Become Different

If the number of stem cells in a crypt is to remain stable, each stem-cell division must on average generate one daughter that remains a stem cell and one that becomes committed to differentiation. In principle, this could be achieved in at least two ways (**Figure 22–7**).

One mechanism—the simplest at first sight—would be through asymmetric division: processes internal to the dividing stem cell could distribute regulatory factors asymmetrically to its two daughters, as occurs in *Drosophila* neuroblast divisions (see Figure 21–36). The factors inherited by one daughter would cause it to remain a stem cell, while those inherited by the other would drive it toward differentiation. This strategy would guarantee that the original stem cell would give rise to precisely one stem cell in every subsequent cell generation.

An alternative strategy would be based on a choice that each daughter makes independently of its sister: in normal circumstances, each would have a 50% probability of remaining as a stem cell and a 50% probability of commitment to differentiation. Sometimes the two daughters of a stem cell would thus have opposite fates, sometimes the same. The choice that each cell makes might either be stochastic, like the flip of a coin, or governed by the environment in which the cell finds itself. A strategy of independent choices is more flexible than that of strict asymmetric division. In particular, environmental factors can control the balance of probabilities, adjusting them in favor of the stem-cell option where more stem cells are needed, as they often are, either for growth or for damage repair.

Clonal analysis gives a way to distinguish between the two strategies, since they give quite different predictions as to the expected number of clones of different sizes produced from individual stem cells (see Figure 22–7). For the gut, the findings seem clear: the independent-choice theory fits the observations, and the asymmetric-division theory does not.

Paneth Cells Create the Stem-Cell Niche

There are about 15 *Lgr5*-expressing stem cells in each crypt. They are slim and columnar, and they sit at the crypt base interspersed among the Paneth cells (see Figure 22–6). This is the intestinal **stem-cell niche**: the Paneth cells generate signals, including a strong Wnt signal, that act over a short range to maintain the stem-cell state. Signal proteins from the connective tissue surrounding the crypt base help to reinforce the localizing signal from the Paneth cells; *Lgr5* itself is a receptor for one of these proteins, called R-spondin.

In the intestine, it seems that the niche created by the Paneth cells has space for only a limited number of stem cells, and when these divide, it is a random

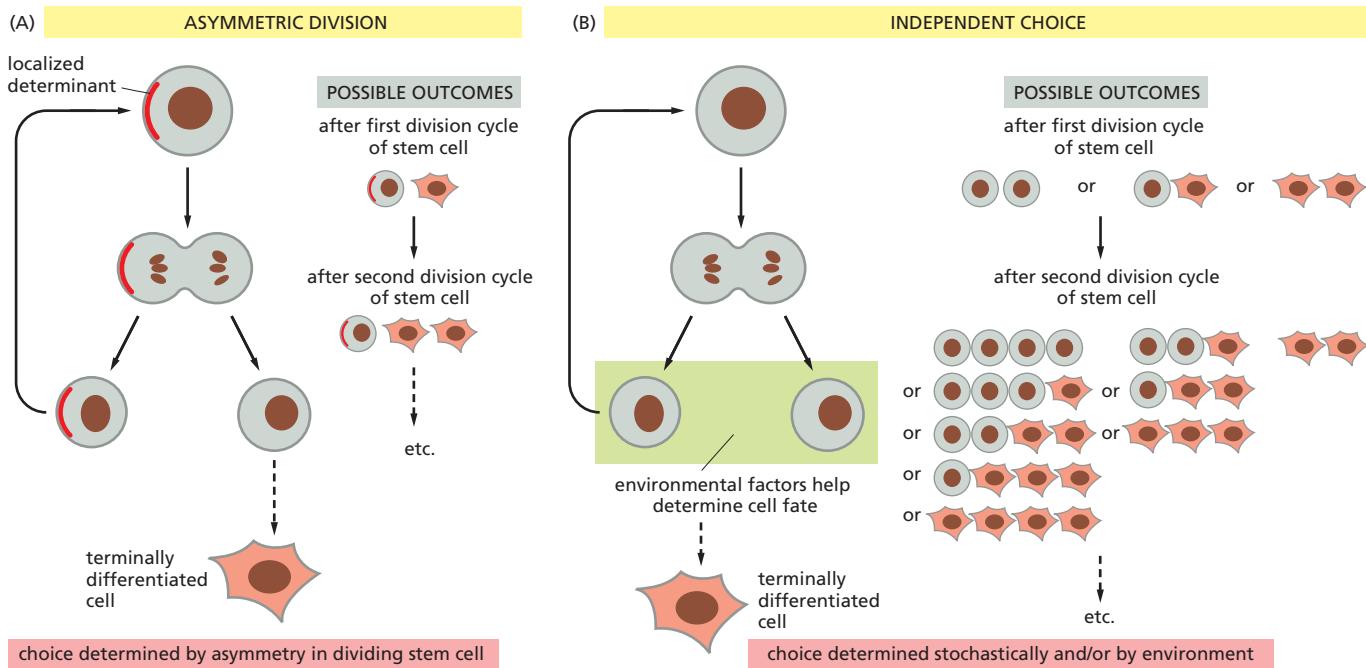


Figure 22–7 Two ways for a stem cell to produce daughters with different fates: asymmetric division and independent choice. (A) The asymmetric-division strategy gives a clone consisting of precisely one stem cell plus a steadily increasing number of differentiating cells, in proportion to the number of cell divisions. (B) The independent-choice strategy is more variable in its outcome. With a choice made at random by each daughter and with a 50% probability for each one to remain a stem cell or differentiate, there is, for example, a 25% chance at the first division that both daughters will differentiate, so that the clone eventually goes extinct. Or, at this division or later, a preponderance of daughters may chance to retain stem-cell character, creating a clone that persists and increases in size. With the help of some mathematics, the probability distribution of clone sizes generated from a single stem cell at any given time can be predicted on this stochastic assumption. The observations in the gut and elsewhere fit the stochastic independent-choice strategy, but not the asymmetric-division strategy.

matter which of them are pushed out of the nest and condemned to differentiation and which stay in place as stem cells for the future. In most other stem-cell systems where the question has been examined, it appears that the fates of the daughters of a stem cell are assigned in a similar way, independently and subject to influence from the cells' environment.

A Single *Lgr5*-expressing Cell in Culture Can Generate an Entire Organized Crypt-Villus System

The Paneth cells themselves are progeny of the stem cells, suggesting that the intestinal stem-cell system is in some way self-maintaining and self-organizing. This is demonstrated in a striking way by taking single dissociated *Lgr5*-expressing cells and allowing them to proliferate in culture, embedded in a cell-free matrix rich in the basal-lamina component laminin (mimicking basal lamina). The cells proliferate, forming at first small, round epithelial vesicles. Within a few days, however, one or another of the cells in the vesicle, at random, begins to differentiate as a Paneth cell. This induces its neighbors to behave as stem cells and initiates transformation of the simple vesicle into an organized structure, or *organoid* (Figure 22–8A,B). Protrusions resembling crypts grow out into the surrounding matrix and contain Paneth cells, *Lgr5*-expressing stem cells, and the transit amplifying cells derived from them; these cell types are confined to the cryptlike structures. Terminally differentiated, nondividing absorptive cells line the other parts of the organoid epithelium, with their microvilli facing the lumen. Goblet and enteroendocrine cells are also present, scattered through the epithelium, and the whole “minigut” structure, with all its cell types, grows and renews itself in much the same way as the lining of the normal intestine.

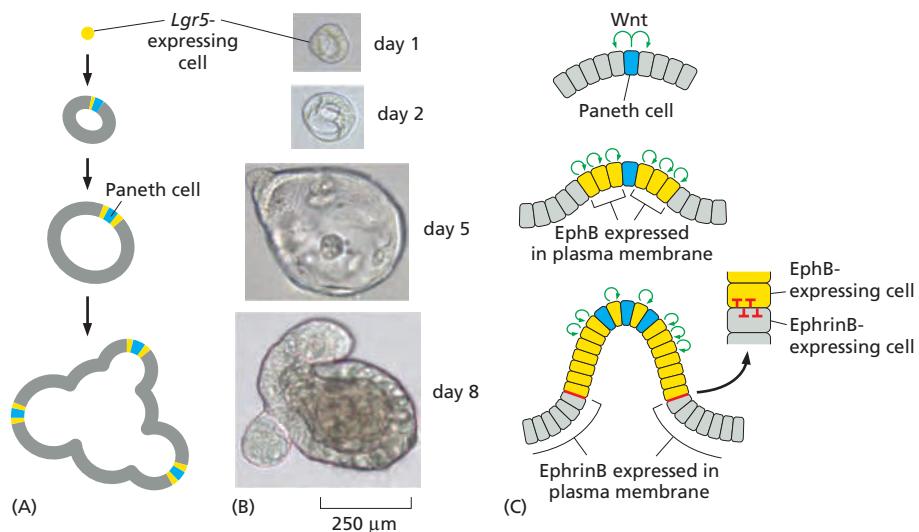


Figure 22–8 Genesis of a minigut from a single *Lgr5*-expressing cell cultured in a cell-free matrix. (A,B) The founder cell first divides to form a small vesicle. At random, one or more of the cells in this vesicle differentiates as a Paneth cell (blue). This cell maintains *Lgr5* expression (yellow) in its immediate neighbors, which persist as stem cells that generate the full range of intestinal cell types. (C) Schematic diagram of the key organizing signals. The Paneth cells organize crypts by producing a Wnt signal that acts on neighboring cells and keeps them proliferating in the stem-cell state. A repulsive interaction based on ephrin–Eph binding causes the crypt cell types (which express EphB, induced by Wnt) to segregate from the nondividing differentiated villus cell types (which express EprinB). Both ephrin and Eph are cell-surface proteins attached to the plasma membrane; in many tissues, two cells that contain a different member of this pair repel each other when they touch (see Figure 21–49). (Adapted from T. Sato and H. Clevers, *Science* 340:1190–1194, 2013. With permission from AAAS.)

Ephrin–Eph Signaling Drives Segregation of the Different Gut Cell Types

The remarkable self-organizing behavior of the cultured organoids suggests that some interaction among the different epithelial cells drives them to segregate from one another. The ephrin–Eph signaling pathway (discussed in Chapter 15) appears to be responsible. The cells that live in the crypts express EphB receptor proteins, while absorptive, goblet, and enteroendocrine cells, as they begin to differentiate, switch off expression of this receptor and instead switch on expression of its ligands, cell-surface proteins of the EprinB family (Figure 22–8C). In various other tissues, cells expressing Eph proteins are repelled by contacts with cells expressing ephrins on their surface (see Figures 21–49 and 21–79). It seems that the same is true in the gut lining, and that this mechanism serves to keep the cells segregated and in their proper places. In EphB knockout mutants, the populations become mixed, so that, for example, Paneth cells wander out onto the villi.

Notch Signaling Controls Gut Cell Diversification and Helps Maintain the Stem-Cell State

If a single type of stem cell generates all the differentiated cell types in the gut lining, what causes the progeny of this stem cell to diversify? Notch signaling has this role in many other systems, where it mediates lateral inhibition—a competitive interaction that drives neighboring cells toward different fates (see Figure 15–58 and Figure 21–35). All the essential components of the Notch pathway are expressed in the crypts; it seems that Wnt signaling maintains them there. If Notch signaling is abruptly blocked, within a few days all the cells in the crypts differentiate as goblet cells, and absorptive cells cease to be produced; conversely, if Notch signaling is artificially activated in all the cells, absorptive cells continue

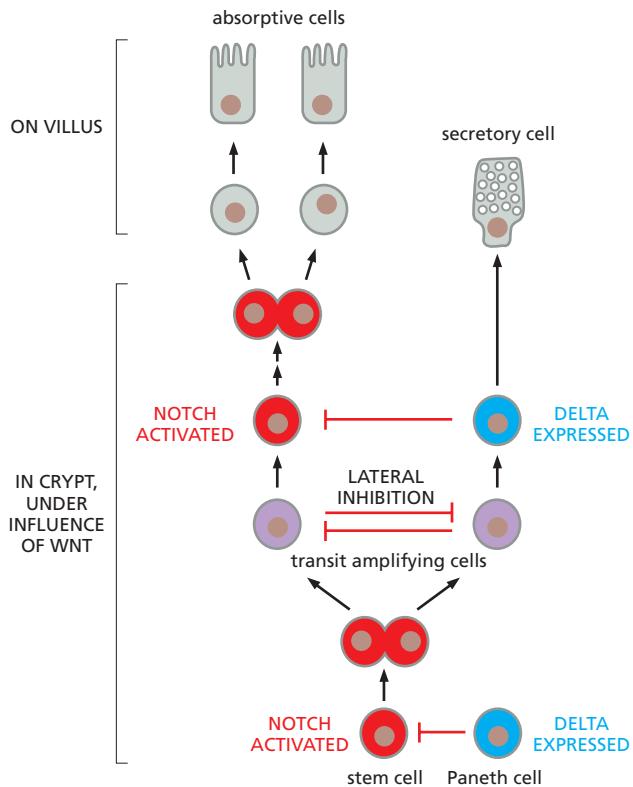


Figure 22–9 How Notch signaling, in combination with Wnt, maintains stem cells and drives cell diversification in the intestine. Wnt signaling leads to expression of Notch and Delta in the cells of the crypt, and Delta-Notch signaling in the crypt mediates lateral inhibition between adjacent cells. Cells expressing higher levels of Delta eventually activate Notch in their neighbors, adopt a secretory fate, and stop dividing; their neighbors, with activated Notch, are prevented from differentiating and keep on dividing. Essentially the same process operates at the crypt base, where the Paneth cells express higher levels of Delta to prevent stem cells from differentiating, and in the transit amplifying population, where nascent secretory cells express higher levels of Delta. Division continues in the Notch-activated cells as they move up the crypt, until they escape from the influence of Wnt and emerge onto the villi to become absorptive cells.

to be generated but no goblet cells are produced. This reflects the lateral inhibition mechanism operating in normal animals: the nascent goblet (and other secretory) cells express the Notch ligand Delta and thereby activate Notch in their neighbors, inhibiting them from differentiating as secretory (Figure 22–9).

Delta-Notch signaling is crucial not only in the transit amplifying population, but also at the crypt base: the Paneth cells express Delta and this activates Notch in the stem cells, inhibiting differentiation. Without this influence, the stem cells lose their special character and differentiate as secretory cells. Thus maintenance of the intestinal stem-cell state requires a combination of signals, with both Wnt and Notch acting as central players.

The Epidermal Stem-Cell System Maintains a Self-Renewing Waterproof Barrier

Stem-cell systems are organized in many different ways, but they share some underlying principles. Consider the **epidermis**, for example—the outer, epithelial covering of the body. The epidermis undergoes continual renewal, but, unlike the lining of the gut, it is multilayered or *stratified*. Stem cells are located in the basal layer, and their progeny move outward toward the exposed surface, differentiating as they go. They end up as lifeless scales or *squames*, which are eventually shed from the surface of the skin (Figure 22–10). Even though the architecture of this tissue is very different from that of the intestine, many of the same basic principles apply. The stem cells depend for their existence on signals from a specific niche, in this case the basal lamina and underlying connective tissue. The daughters of stem cells that are committed to differentiation undergo several divisions as transit amplifying cells (while still in the basal layer) before differentiating. Finally, a stochastic independent-choice mechanism dictates the fates of the daughters of a stem-cell division, allowing for increase in the number of stem cells when needed for growth or wound healing. Most of the same signaling pathways that organize the intestinal stem-cell system are also involved in regulating the epidermal stem-cell system, although with different individual roles.

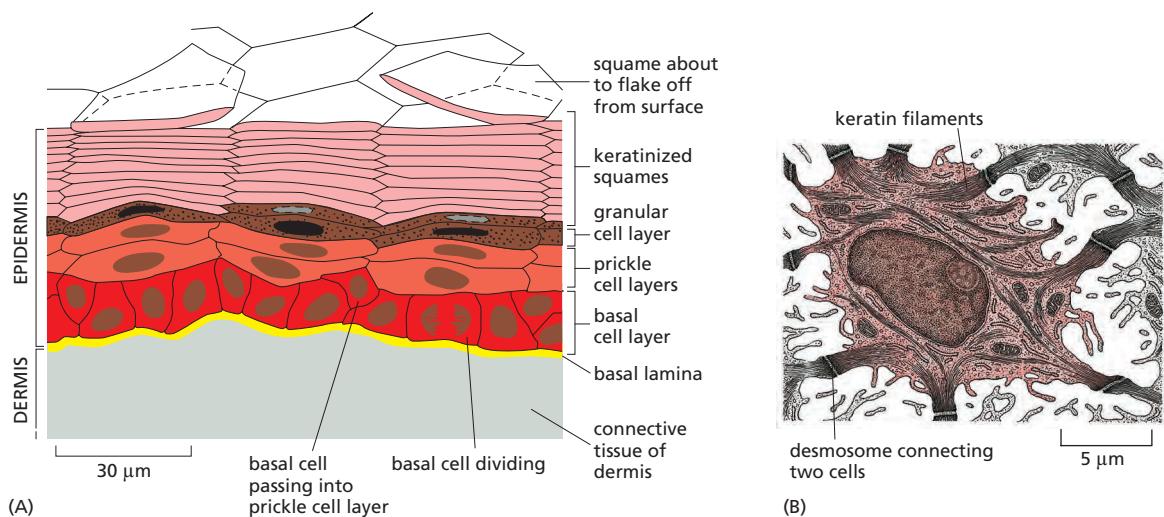


Figure 22–10 The multilayered structure of the epidermis, as seen in thin skin of a mouse. (A) The epidermis forms the outer covering of the skin, creating a waterproof barrier that is self-repairing and continually renewed. Beneath this lies a relatively thick layer of connective tissue, which includes the tough, collagen-rich dermis (from which leather is made) and the underlying fatty subcutaneous layer or hypodermis. The cells of the epidermis are called keratinocytes, because their characteristic differentiated activity is the synthesis of keratin intermediate filament proteins, which give the epidermis its toughness. These cells change their appearance and properties from one layer to the next, progressing through a regular program of differentiation. Those in the innermost layer, attached to an underlying basal lamina, are termed basal cells, and it is usually only these that divide: the basal cell population includes relatively small numbers of stem cells along with larger numbers of transit amplifying cells derived from them. Above the basal cells are several layers of larger prickle cells, shown in top view in (B), 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. Beyond the prickle cells lies the thin, darkly staining granular cell layer, where the cells are sealed together to form a waterproof barrier; this 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, which are eventually shed from the surface of the skin. The time from exit of a cell from the basal layer to its loss by shedding at the surface is a week or two, depending on body region and species.

In addition to the cells destined for keratinization, the deep layers of the epidermis include small numbers of cells (not shown) that invade this tissue and have quite different origins and functions. These immigrants include dendritic cells, called Langerhans cells, derived from bone marrow and belonging to the immune system; melanocytes (pigment cells) derived from the neural crest; and Merkel cells, which are associated with nerve endings in the epidermis. (B, from R.V. Krstic, Ultrastructure of the Mammalian Cell: an Atlas. Berlin: Springer-Verlag, 1979.)

Tissue Renewal That Does Not Depend on Stem Cells: Insulin-Secreting Cells in the Pancreas and Hepatocytes in the Liver

Some types of cells can divide even though fully differentiated, allowing for renewal and regeneration without the use of stem cells. The *insulin-secreting cells* (β cells) of the pancreas are one example. Their mode of renewal has a special importance, because it is the loss of these cells (through autoimmune attack) that is responsible for type 1 (juvenile-onset) diabetes; they are also a significant factor in the type 2 (adult-onset) form of the disease. The β cells are normally sequestered in cell clusters called *islets of Langerhans*. These islets contain no obvious subset of cells specialized to act as stem cells, yet fresh β cells are continually generated within them. Lineage tracing studies, similar to those described above for the gut, show that the renewal of this population normally occurs by simple duplication of the existing insulin-expressing cells, and not by means of stem cells.

Another tissue that can renew by simple duplication of fully differentiated cells is the liver. The main cell type in the liver is the *hepatocyte*, a large cell that performs the liver's metabolic functions. Hepatocytes normally live for a year or more and renew themselves through cell division at a very slow rate. Powerful homeostatic mechanisms 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 or restore it to that size

in case of damage. A dramatic effect is seen if large numbers of hepatocytes are removed surgically or are 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 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 two weeks.

Both the pancreas and the liver contain small populations of stem cells that can be called into play as a backup mechanism for production of the differentiated cell types in more extreme circumstances. This imparts resilience to the mechanisms of renewal and repair.

Some Tissues Lack Stem Cells and Are Not Renewable

The variety among tissues in the capacity for self-renewal is illustrated in a striking way by comparing the olfactory epithelium in the nose, the auditory epithelium of the inner ear, and the photoreceptive epithelium of the retina. These three sensory structures, which like the epidermis develop from the ectodermal layer of the early embryo, differ radically in their self-renewal capabilities. The olfactory epithelium contains a population of stem cells that give rise to differentiated cells that have a limited life-span and are continually replaced. But unlike the epidermis, these differentiated cells (the olfactory receptor cells) are neurons, with cell bodies lying in the olfactory epithelium and axons that extend back to the olfactory lobes in the brain. The continual renewal of this epithelium therefore involves continual production of fresh axons, which have to navigate back to the appropriate sites in the brain.

In contrast, in mammals at least, the auditory epithelium and the retinal epithelium lack stem cells, and their sensory receptor cells—the sensory hair cells in the ear, the photoreceptors in the retina—are irreplaceable. If they are destroyed—whether by too much exposure to loud noise, by looking into the beam of a laser, or through degenerative processes in old age—the loss is permanent.

Summary

Many tissues in the adult mammalian body are continually renewed by stem cells. 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 lining of the gut renews itself more rapidly than any other tissue in the mammalian body and provides a paradigm for the workings of stem-cell systems. In the small intestine, there is a continual upward flow from crypts, where new cells are generated by cell division, onto villi that are composed of nondividing differentiated cells. Wnt signaling maintains cell proliferation in the crypts, and overactivation of the Wnt pathway gives rise to tumors. Stem cells lie at each crypt base and are distinguished by expression of Lgr5 and certain other genes. The Lgr5⁺ stem cells are multipotent, each capable of generating several different types of differentiated cells as well as new stem cells. The balance of fate choices is adjusted according to need, allowing increase in the number of stem cells where more are needed for growth or repair. In a suitable cell-free culture medium, a single Lgr5⁺ stem cell can generate a self-organizing "minigut," containing all the standard intestinal epithelial cell types.

Other self-renewing epithelia, such as the epidermis with its multilayered (stratified) architecture, have stem cells and their differentiating progeny arranged in different ways but are governed by similar basic principles. However, tissue renewal and repair does not always have to depend on stem cells. Thus, the population of insulin-producing cells in the pancreas is enlarged and renewed by simple duplication of existing insulin-producing cells. Similarly, in the liver, differentiated hepatocytes remain able to divide throughout life and can dramatically increase their division rate when the need arises. At an opposite extreme, some tissues, such as the sensory epithelia of the ear and the eye, do not undergo any turnover and are not renewable: their cells, once lost, are lost forever.

FIBROBLASTS AND THEIR TRANSFORMATIONS: THE CONNECTIVE-TISSUE CELL FAMILY

From epithelia, with their varied patterns of renewal and their enormous variety of protective, absorptive, secretory, sensory, and biosynthetic functions, we turn now to **connective tissues**. Connective tissues typically consist of cells dispersed in extracellular matrix that they themselves secrete, as discussed in Chapter 19. They originate from the mesodermal (middle) layer of the early embryo, sandwiched between ectoderm and endoderm (see Chapter 21, Figure 21–3).

In the adult body, virtually all epithelia are supported by a connective-tissue bed, or *stroma*; and specialized types of connective tissue, such as bone, cartilage, and tendon, form the supporting framework of the body as a whole. No less important than its mechanical role, connective tissue also contains the blood vessels that bring the oxygen and nourishment on which all cells depend. Cells of the immune system roam through connective tissue, passing in and out of blood vessels and lymphatics, and providing defence against infection; and through the meshes of connective tissue run peripheral nerves. Also embedded in connective tissue are the muscles that enable us to move. In these many ways, the cells that form connective tissue and synthesize its various types of extracellular matrix contribute to the support and repair of almost every tissue and organ.

Connective-tissue cells belong to a family of cell types that are related by origin, and they are often remarkably 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* (*adipocytes*) and *smooth muscle cells*. **Figure 22–11** illustrates these cell types and the interconversions that are thought to occur between them. The adaptability of the differentiated character of connective-tissue cells is an important feature of responses to many types of damage.

Fibroblasts Change Their Character in Response to Chemical and Physical 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 (**Movie 22.1**), and produce large amounts of collagenous matrix that 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.

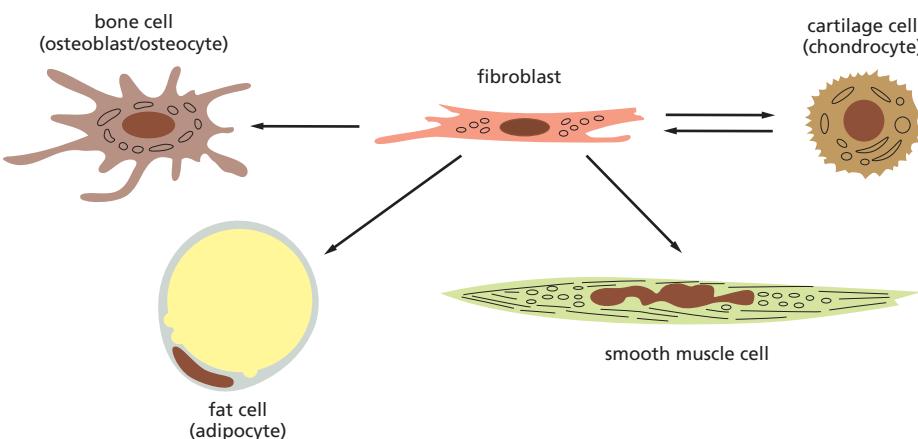
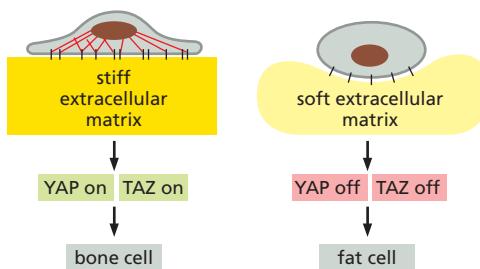


Figure 22–11 The family of connective-tissue cells. Arrows show the interconversions that are thought to occur within the family. For simplicity, the fibroblast is shown as a single cell type, but it is uncertain how many types of fibroblasts exist and whether the differentiation potential of different types is restricted in different ways.



A class of connective-tissue cells in the bone marrow, called *bone marrow stromal* cells, provides an example of radical 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. Depending on the culture conditions, the members of such a clone either can continue proliferating to produce more cells of the same type, or can differentiate as fat cells, cartilage cells, or bone cells. The fate of the cells depends on physical as well as chemical signals: embedded in a stiff, unyielding matrix, they tend to turn into bone cells, whereas in a softer, more elastic matrix, they tend to turn into fat cells. This effect is mediated by an intracellular pathway that responds to tension in actin-myosin bundles and relays a signal to specific transcription regulators in the nucleus (Figure 22–12). Because of their self-renewing, multipotent character, the bone marrow stromal cells, and other cells with similar properties, are referred to as *mesenchymal stem cells*.

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 (discussed in Chapter 19). The cartilage matrix is deformable, and the tissue grows by expanding as the chondrocytes divide and secrete more matrix (Figure 22–13). **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 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 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 additional matrix in small quantities around itself. The osteocyte, like the chondrocyte, occupies a small cavity, or *lacuna*, in the matrix, but unlike the chondrocyte

Figure 22–12 Control of fibroblast differentiation by the physical properties of the extracellular matrix. On a stiff matrix, the cells form strong adhesions, spread out, and tend to turn into bone cells. On a soft matrix, where the cells are unable to form strong anchorages, they fail to spread and tend to differentiate as fat cells. These effects depend on transcription regulators (YAP and TAZ proteins) that move into the cell nucleus in response to tension developed in actin–myosin bundles in the cytoplasm. (Based on S. Dupont et al., *Nature* 474:179–183, 2011.)

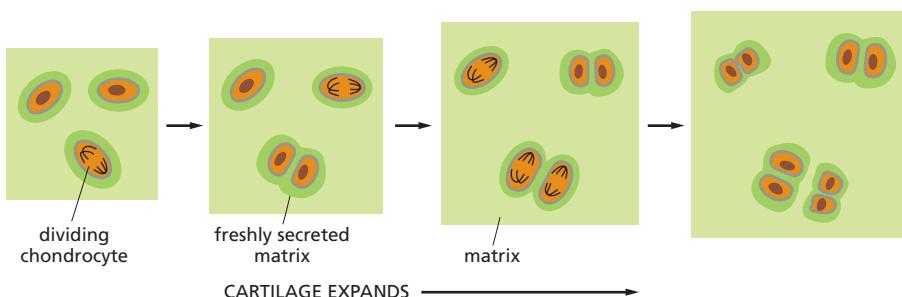


Figure 22–13 The growth of cartilage. The tissue expands as the chondrocytes divide and make more matrix. The freshly synthesized matrix with which each cell surrounds itself is shaded dark green. Cartilage may also grow by recruiting fibroblasts from the surrounding tissue and converting them into chondrocytes.

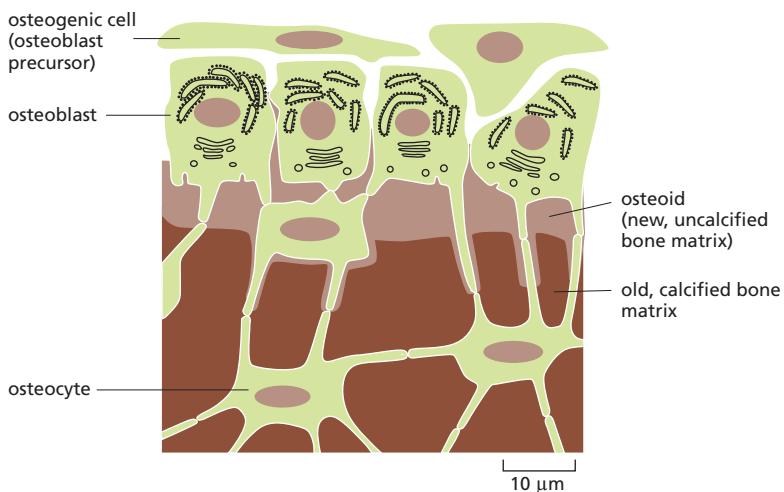


Figure 22-14 Deposition of bone matrix by osteoblasts. Osteoblasts lining the surface of bone secrete the organic matrix of bone (osteoid) and are converted into osteocytes as they become embedded in this matrix. The matrix calcifies soon after it has been deposited. The osteoblasts themselves are thought to derive from osteogenic stem cells that are closely related to fibroblasts.

it is not isolated from its fellows. Tiny channels, or *canalliculi*, radiate from each lacuna and contain cell processes from the resident osteocyte, enabling it to form gap junctions with adjacent osteocytes (Figure 22-14). 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 22-15). 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 explain below.

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 22-16) are large, multinucleated cells that originate, like macrophages, from hematopoietic stem cells in the bone marrow (discussed later

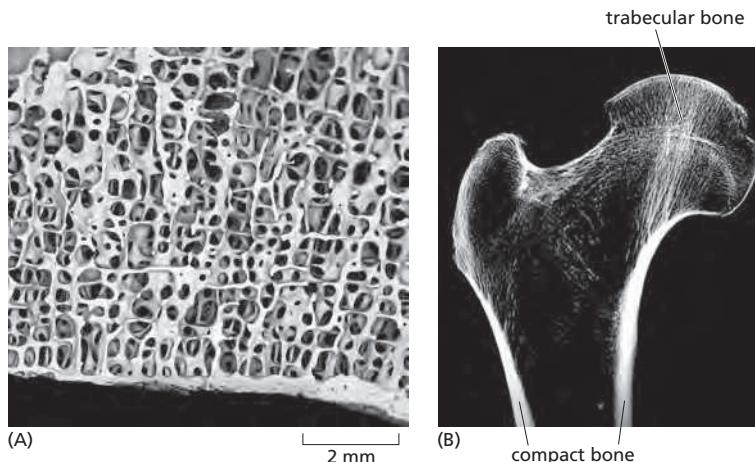


Figure 22-15 Trabecular and compact bone. (A) Low-magnification scanning electron micrograph of trabecular bone in a vertebra of an adult man. The soft marrow tissue has been dissolved away. (B) A slice through the head of the femur, with bone marrow and other soft tissue likewise dissolved away, reveals 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.)

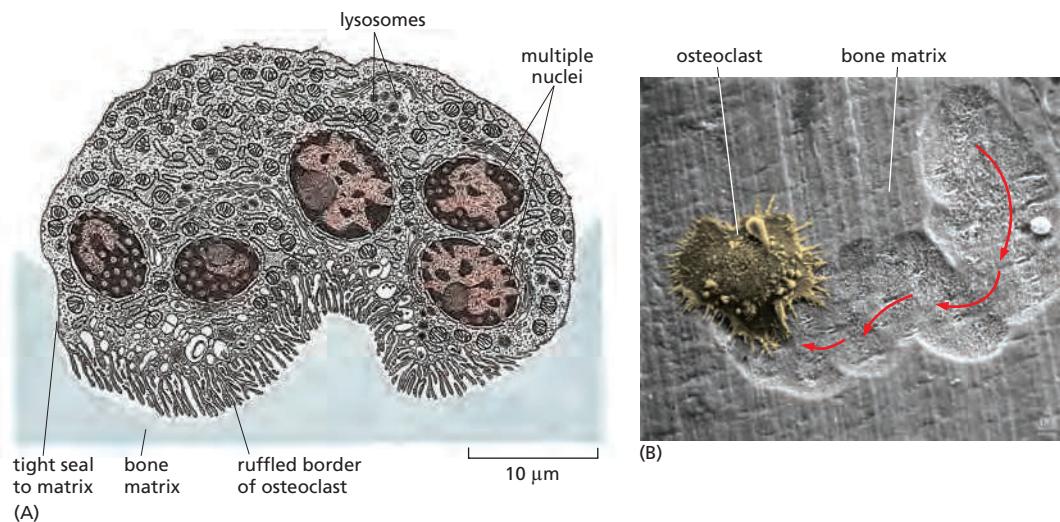


Figure 22-16 Osteoclasts. (A) Drawing of an osteoclast in cross section. This giant, multinucleated cell erodes bone matrix. The “ruffled border” is a site of secretion of acids (to dissolve the bone minerals) and hydrolases (to digest the organic components of the matrix). Osteoclasts vary in shape, 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.)

in this chapter). The precursor cells are released 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, and the walls of the tunnel become lined with a layer of osteoblasts (Figure 22-17). These osteoblasts lay down concentric layers of new matrix, which gradually fill the cavity, leaving only a narrow canal surrounding the new blood vessel. At the same time as some tunnels are filling up with bone, others are being bored by osteoclasts, cutting through older concentric systems.

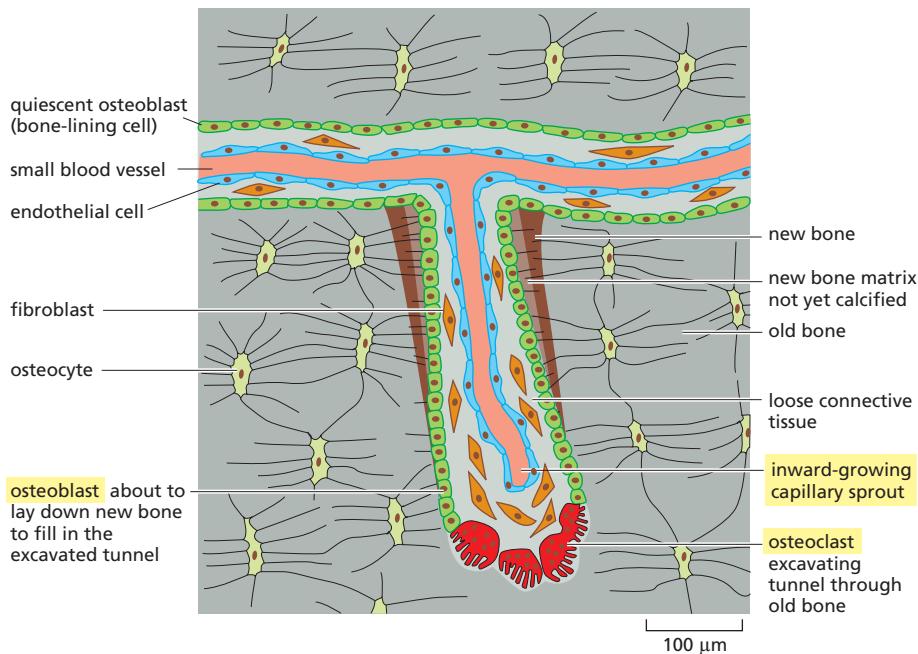


Figure 22-17 The remodeling of compact bone. Osteoclasts acting together in a small group excavate a tunnel through the old bone, advancing at a rate of about 50 μm per day. Osteoblasts enter the tunnel behind them, line its walls, and begin to form new bone, depositing layers of matrix at a rate of 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.)

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. Disturbance of the 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 have powerful effects on this balance. Chronic use of corticosteroid drugs, for example, can cause osteoporosis as a side effect; but this can be treated by other drugs that redress the balance, including agents that block the factors that osteoblasts secrete to recruit osteoclasts.

Local controls 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, by remodeling of the bone tissue ahead of it and behind it.

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 process that resembles the way bones develop in the embryo: 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.

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. While osteoblasts secrete bone matrix, they also produce signals that recruit monocytes from the circulation to become osteoclasts, which degrade bone matrix. Through the activities of these antagonistic classes of cells, bone undergoes a 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.

GENESIS 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 eukaryotic 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 22–18). These differ in function, structure, and development. Although all of them generate contractile forces by using organized filament systems based on actin and myosin II, the actin and myosin molecules employed have somewhat different amino acid sequences, are

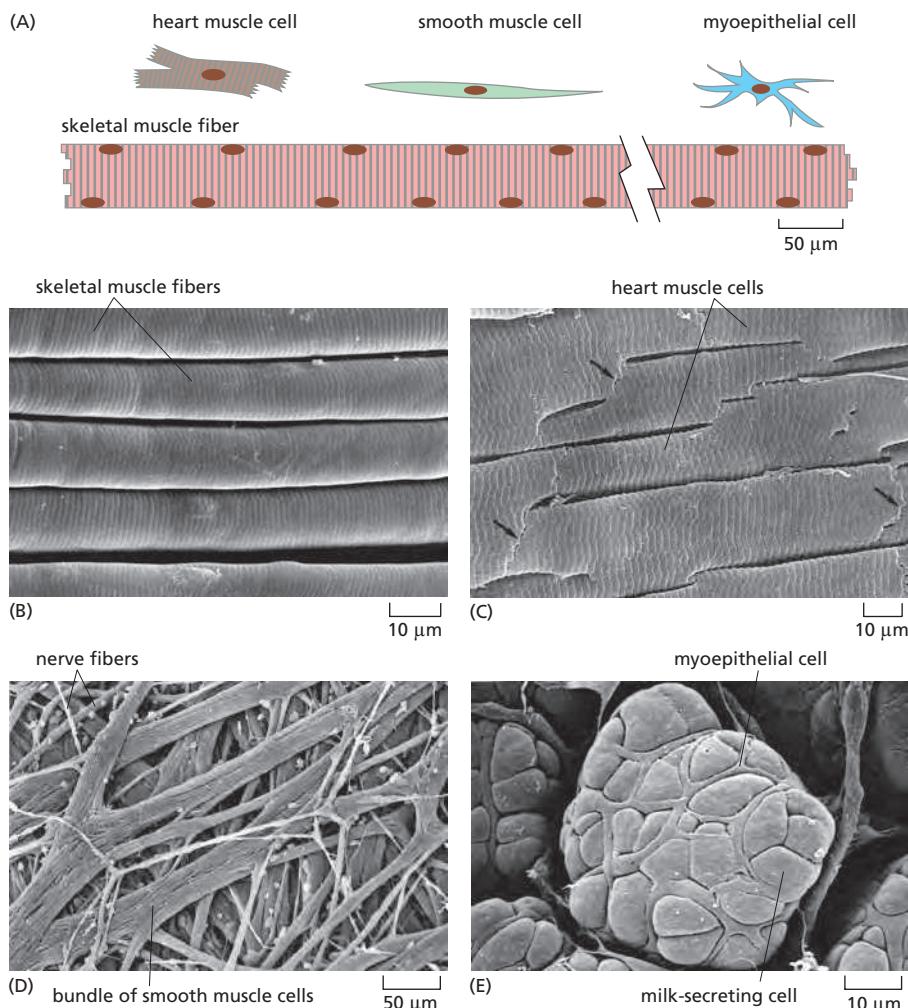


Figure 22-18 The four classes of muscle cells of a mammal. (A) Schematic drawings (to scale). (B–E) Scanning electron micrographs. Skeletal muscle fibers (B, from a hamster) are giant cells with many nuclei and are formed by cell fusion. The other types of muscle cells are more conventional, generally having only one single nucleus. Heart muscle cells (C, from a rat) 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. 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. Smooth muscle cells (D, from the urinary bladder of a guinea-pig) are so named because they do not appear striated; they belong to the connective-tissue family and are closely related to fibroblasts. Note that the smooth muscle is shown here at a lower magnification than the other muscle types. 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 (E, from a secretory alveolus of a lactating rat mammary gland) 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. (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 Tissue Res.* 209:1–10, 1980. With permission from Springer-Verlag.)

differently arranged in the cell, and are associated with different sets of proteins that control contraction.

We focus in this section on skeletal muscle cells, which are responsible for practically all movements that are under voluntary control. These cells can be very large (2–3 cm long and 100 μm 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. In an intact muscle, they are bundled tightly together, with fibroblasts (and some fat cells) in the interstices between them and blood vessels and nerve fibers running through the tissue. The mechanisms of muscle contraction were discussed in Chapter 16. Here we consider the unusual strategy by which the multinucleate skeletal muscle cells are generated and maintained.

Myoblasts Fuse to Form New Skeletal Muscle Fibers

During development, 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. 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 22-19). Once differentiation and cell fusion have occurred, the cells do not divide and the nuclei never again replicate their DNA.

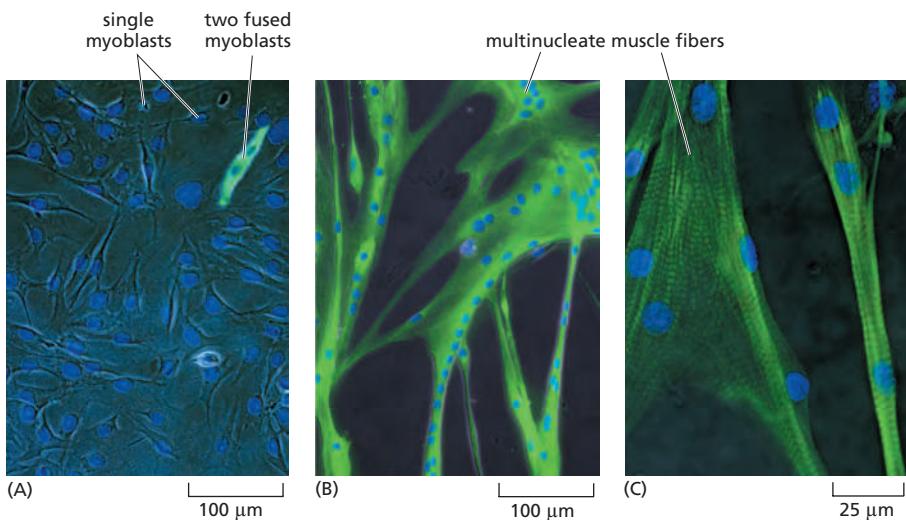


Figure 22–19 Myoblast fusion in culture.

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.)

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 22–20). 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. 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.

The process of muscle repair by means of satellite cells is, however, 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 decline of capacity for repair likewise contributes to the weakening of muscle in the elderly.

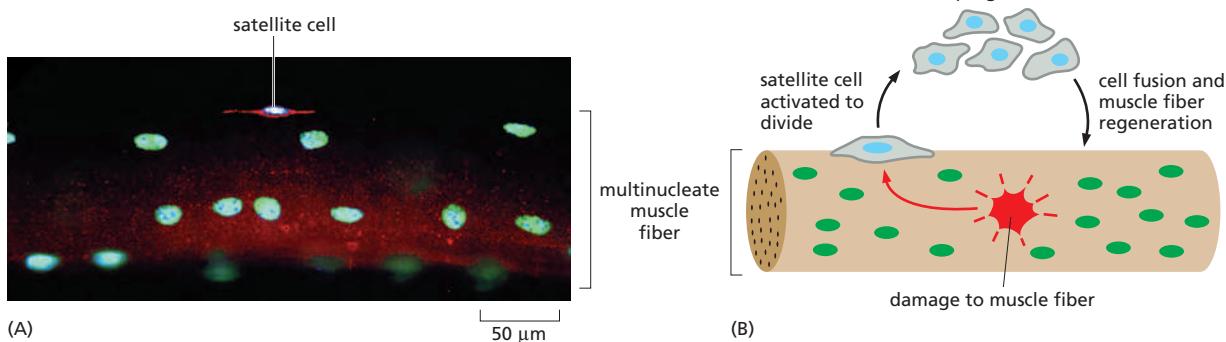


Figure 22–20 Satellite cells repair skeletal muscle fibers. (A) The specimen is stained with an antibody (red) against a muscle cadherin, M-cadherin, which is present on both the satellite cell and the muscle fiber and is concentrated at the site where their membranes are in contact. The nuclei of the muscle fiber are stained green, and the nucleus of the satellite cell is stained blue. (B) Schematic of the repair of a damaged muscle fiber by proliferation and fusion of satellite cells. (A, 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 in order to replace the muscle cells that have been lost. They are the stem cells of skeletal muscle, and exhaustion of their regenerative capacity is responsible for some forms of muscular dystrophy as well as for the decline of muscle mass in old age.

BLOOD VESSELS, LYMPHATICS, AND ENDOTHELIAL CELLS

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, in the hope that it may be possible to block the growth of tumors by attacking the endothelial cells that bring them nourishment.

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 22–21). The inner 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 22–22), together with a few scattered pericytes. Related to vascular smooth muscle cells, pericytes wrap themselves around the small vessels and strengthen them (Figure 22–23).

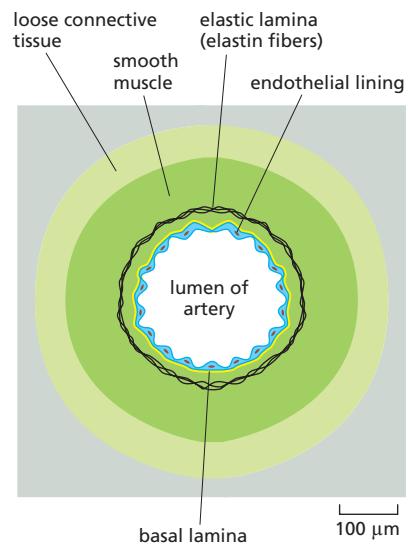


Figure 22–21 Diagram of a small artery in cross section. The endothelial cells form the endothelial lining, which although inconspicuous, is the fundamental component. Compare with the capillary in Figure 22–22.

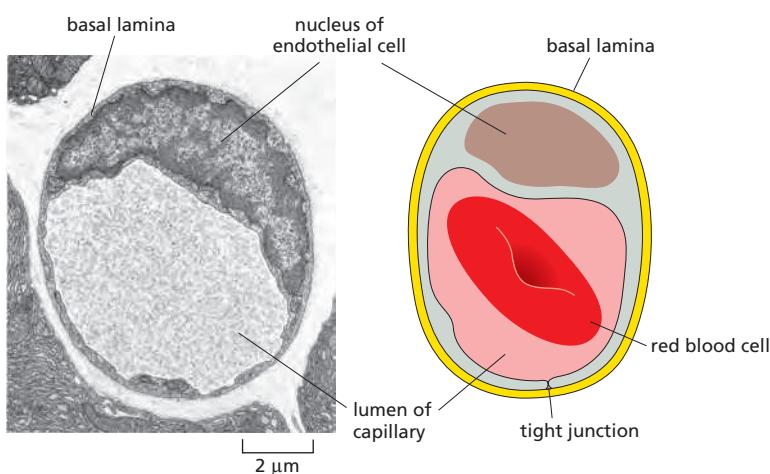


Figure 22–22 Capillaries. Electron micrograph (left) of a cross section of a small capillary in the pancreas. The wall is formed by a single endothelial cell surrounded by a basal lamina, as seen most clearly in the drawing to the right. (From R.P. Bolender, *J. Cell Biol.* 61:287, 1974. With permission from The Rockefeller University Press.)

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 can 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 they control the passage of materials—and the transit of white blood cells—into and out of the bloodstream. Arteries, veins, capillaries, 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 occurs 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 retina, which blood vessels invade according to a predictable timetable, provides 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 22–24**). 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. The tip cell's most striking feature is that it puts out many long filopodia, resembling



Figure 22–23 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 also present 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.)

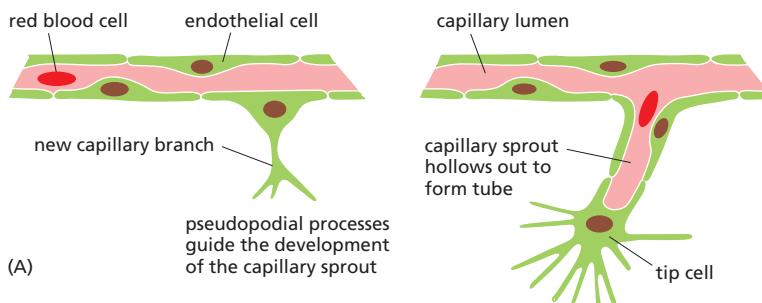


Figure 22–24 Angiogenesis. (A) A new blood capillary forms by the sprouting of an endothelial cell from the wall of an existing small vessel. An endothelial tip cell, with many filopodia, leads the advance of each capillary sprout. The endothelial stalk cells trailing behind the tip cell become hollowed out to form a lumen. (B) Blood capillaries sprouting in the retina of an embryonic mouse that had a red dye injected into the bloodstream, revealing the capillary lumen opening up behind the tip cell (**Movie 22.2**). (B, from H. Gerhardt et al., *J. Cell Biol.* 161:1163–1177, 2003. With permission from the author.)

those of a neuronal growth cone. The column of stalk cells behind it, meanwhile, becomes hollowed out to form a lumen.

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 the netrins, slits, and ephrins mentioned in our account of neural development in the previous chapter. The corresponding receptors 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 chiefly dedicated to the control of vascular development: *vascular endothelial growth factor*, or VEGF.

Tissues Requiring a Blood Supply Release VEGF

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 22–25). 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). 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 level of a transcription factor 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, 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 gradient and move toward its source. As the new vessels form, bringing blood to the tissue, the oxygen concentration rises. The HIF1α activity then declines, VEGF production is shut off, and angiogenesis comes to a halt (Figure 22–26).

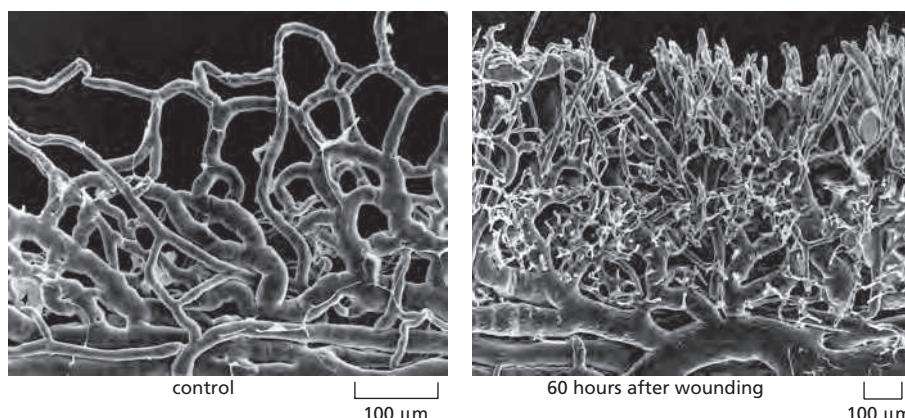


Figure 22–25 New capillary formation in response to wounding. Scanning electron micrographs of casts of the system of blood vessels surrounding the margin of the cornea show the reaction to wounding. The casts are made by injecting a resin into the vessels and letting the resin set; this reveals the 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.)

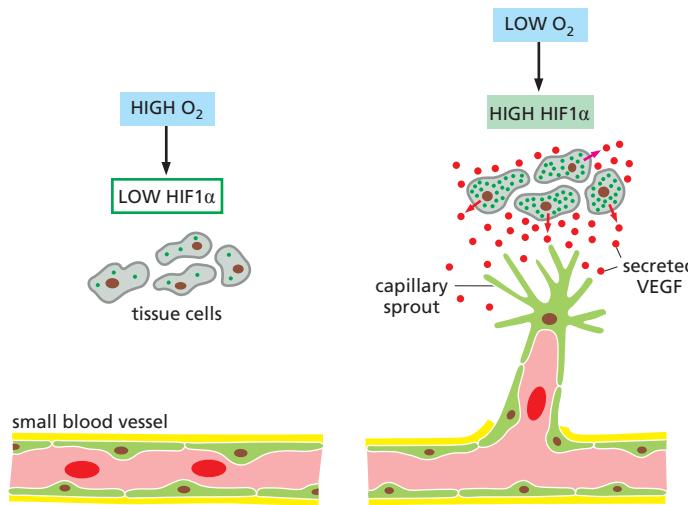


Figure 22–26 The regulatory mechanism controlling blood vessel growth according to a tissue's need for oxygen. Lack of oxygen triggers the secretion of VEGF, which stimulates angiogenesis.

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 22–23) 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 platelet-derived growth factor-B (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 are missing in many regions. 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.

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 in much the same way as the growth cone of a neuron is led. The endothelial stalk cells following behind become hollowed out to form a capillary tube. 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 1α (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. As new vessels enlarge, they recruit increasing numbers of pericytes—cells that cling to the outside of the endothelial tube and mature into the smooth muscle coat that is needed to give the vessel strength.

A HIERARCHICAL STEM-CELL SYSTEM: BLOOD CELL FORMATION

The function of blood vessels is to carry blood, and it is to blood itself that we now turn. 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 throughout the life of the animal. Most remarkably, they are all generated ultimately from a common stem cell, located (in adult humans) in the bone marrow. This *hematopoietic* (blood-making) *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 the osteoclasts in bone, as mentioned earlier. The hematopoietic system is the most complex of the stem-cell systems in the mammalian body, and it is exceptionally important in medical practice.

Red Blood Cells Are All Alike; White Blood Cells Can Be Grouped in Three Main Classes

Blood cells can be classified as red or white. The **red blood cells**, or **erythrocytes**, remain within the blood vessels and transport O₂ and CO₂ bound to hemoglobin. The **white blood cells**, or **leukocytes**, combat infection and in some cases phagocytose and digest debris. Leukocytes, 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.

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 22-27**). The differences in staining reflect major differences of chemistry and function. *Neutrophils* (also called *polymorphonuclear leukocytes* 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 24 (see Movie 16.1). *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 hematopoietic stem cells. *Eosinophils* help to destroy parasites and modulate allergic inflammatory responses.

Once they leave the bloodstream, **monocytes** (see Figure 22-27D) mature into **macrophages**, which, together with neutrophils, are the main “professional phagocytes” in the body. As discussed in Chapter 13, both types of phagocytic 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 (O₂⁻) and hypochlorite (ClO⁻, the active ingredient in bleach), as well as to attack by a concentrated mixture of lysosomal hydrolase enzymes 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.

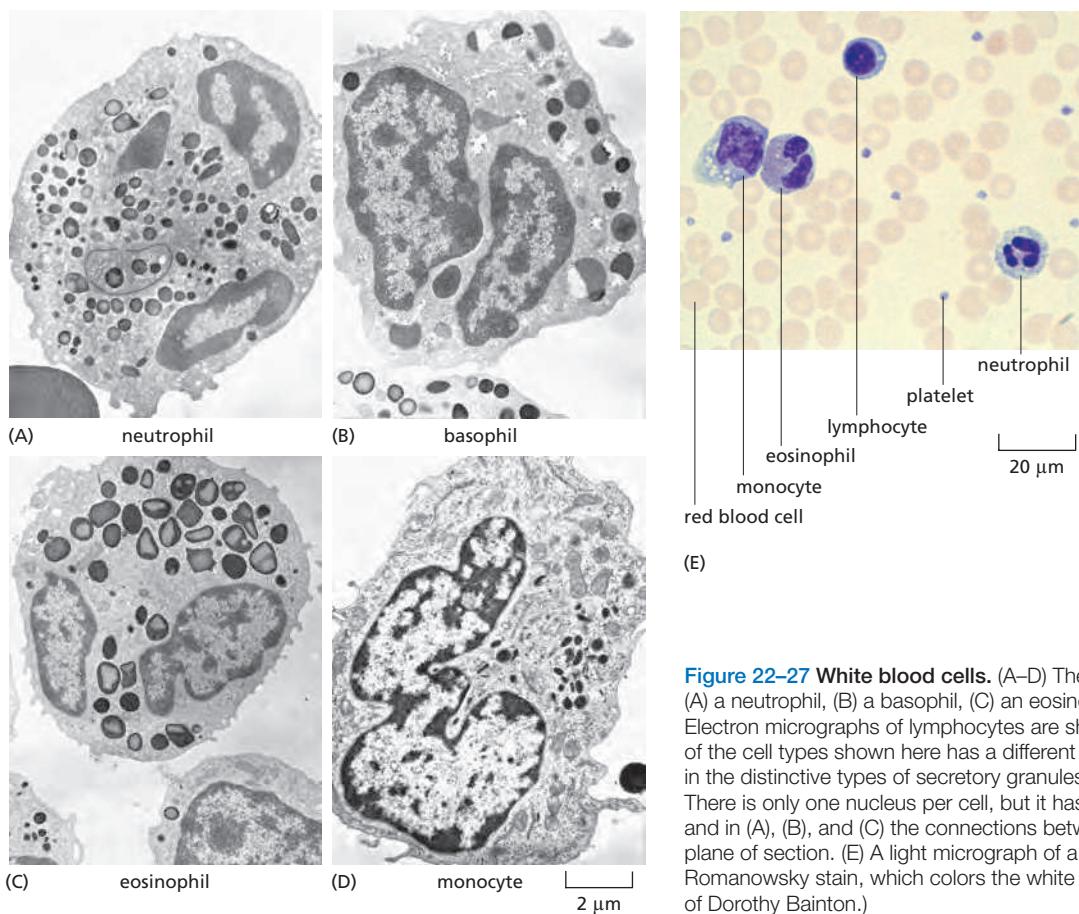


Figure 22–27 White blood cells. (A–D) These electron micrographs show (A) a neutrophil, (B) a basophil, (C) an eosinophil, and (D) a monocyte. Electron micrographs of lymphocytes are shown in Figure 24–14. Each of the cell types shown here has a different function, which is reflected in the distinctive 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. (Courtesy of Dorothy Bainton.)

Monocytes also give rise to *dendritic cells*. 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 instead have a crucial role as presenters of foreign antigens to lymphocytes to trigger an immune response. Dendritic cells in the epidermis (called *Langerhans cells*), for example, ingest foreign antigens and carry these trophies back from the skin 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 lymphocyte-like 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 24. Here we concentrate mainly on the development of the other blood cells, often referred to collectively as **myeloid cells**.

Table 22–1 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 ([Movie 22.3](#)).

The inflammatory response is complex and is governed by many different signal molecules produced locally by mast cells, nerve endings, platelets, and white

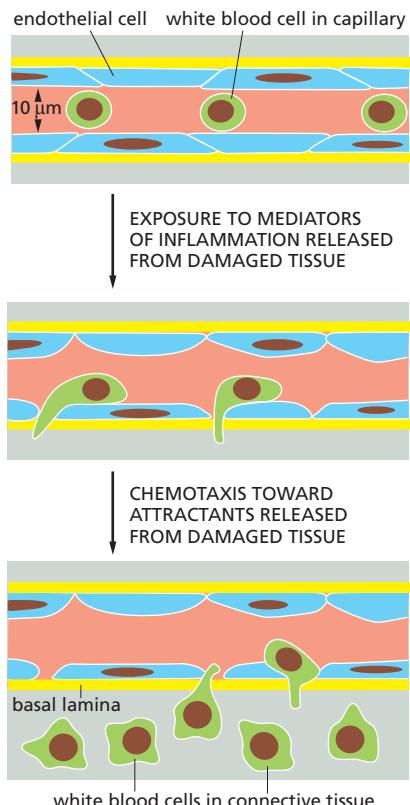
TABLE 22–1 Blood Cells

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 (leukocytes)		
<i>Granulocytes</i>		
Neutrophils (polymorphonuclear leukocytes)	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
<i>Lymphocytes</i>		
B cells	Make antibodies	2×10^9
T cells	Kill virus-infected cells and regulate activities of other leukocytes	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.		

blood cells, as well as by the activation of complement (discussed in Chapter 24). Some of these signal molecules act on the endothelial lining of nearby capillaries, helping white blood cells to first stick and then make an exit from the bloodstream into the tissue where they are needed, as described in Chapter 19 (see Figure 19–28 and Movie 19.2). 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 22–28**).

Other signal molecules produced during an inflammatory response escape into the blood and stimulate the bone marrow to produce more leukocytes and release them into the bloodstream. 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 response to anemia (lack of hemoglobin) due to blood loss, and in the process of acclimatization when one goes to live at high altitude, where oxygen is scarce. Thus, blood cell formation, or *hematopoiesis*, necessarily involves complex controls, which regulate the production of each type of blood cell individually to meet changing needs.

**Figure 22–28** Chemotaxis of white blood cells to damaged tissue.

A chemoattractive signal released from a site of damage, which is toward the bottom of the page, causes white blood cells to exit from the capillary by crawling between adjacent endothelial cells, as shown.

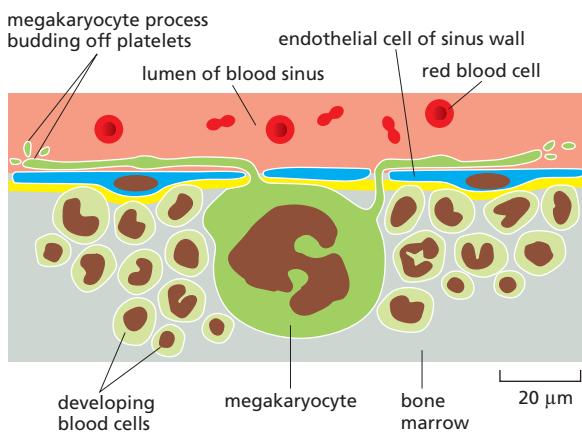


Figure 22–29 A megakaryocyte among other developing blood cells in the bone marrow. The megakaryocyte's enormous size results from its having a highly polyploid nucleus. One megakaryocyte produces about 10,000 platelets, which split off from long processes that extend through holes in the walls of an adjacent blood sinus.

Bone Marrow Contains Multipotent Hematopoietic Stem Cells, Able to Give Rise to All Classes of Blood Cells

In the bone marrow, the developing blood cells and their precursors, including the stem 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 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 22–29 and Movie 22.4).

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. There is no obvious visible characteristic by which we can recognize the ultimate stem cells. In the case of hematopoiesis, the stem cells were first identified by a functional assay that exploited the wandering lifestyle of blood cells and their precursors.

When an animal is exposed to a large dose of x-rays, most of the hematopoietic cells are destroyed and the animal dies within a few days as a result of its 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 hematopoietic tissue (Figure 22–30). Such experiments prove that the marrow contains hematopoietic stem cells. They also show how we can assay for the presence of hematopoietic 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 hematopoietic stem cells. In this way, it has been possible to show that the hematopoietic 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 50,000–100,000; but this is enough. A single such cell injected into a host mouse with defective hematopoiesis is sufficient to reconstitute its entire hematopoietic system, generating a complete set of blood cell types, as well as fresh stem cells. This and other experiments (using artificial lineage markers) show that the individual hematopoietic stem cell is *multipotent* and can give rise to the complete range of blood cell types, both myeloid and lymphoid, as well as to new stem cells like itself (Figure 22–31).

x-irradiation halts blood cell production; mouse would die if no further treatment were given

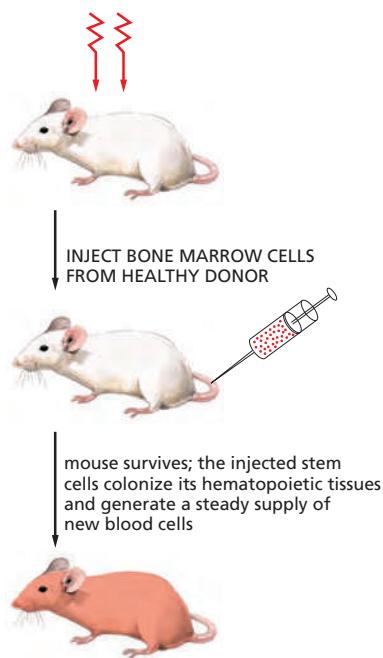


Figure 22–30 Rescue of an irradiated mouse by a transfusion of bone marrow cells. An essentially similar procedure is used in the treatment of leukemia in human patients by bone marrow transplantation.

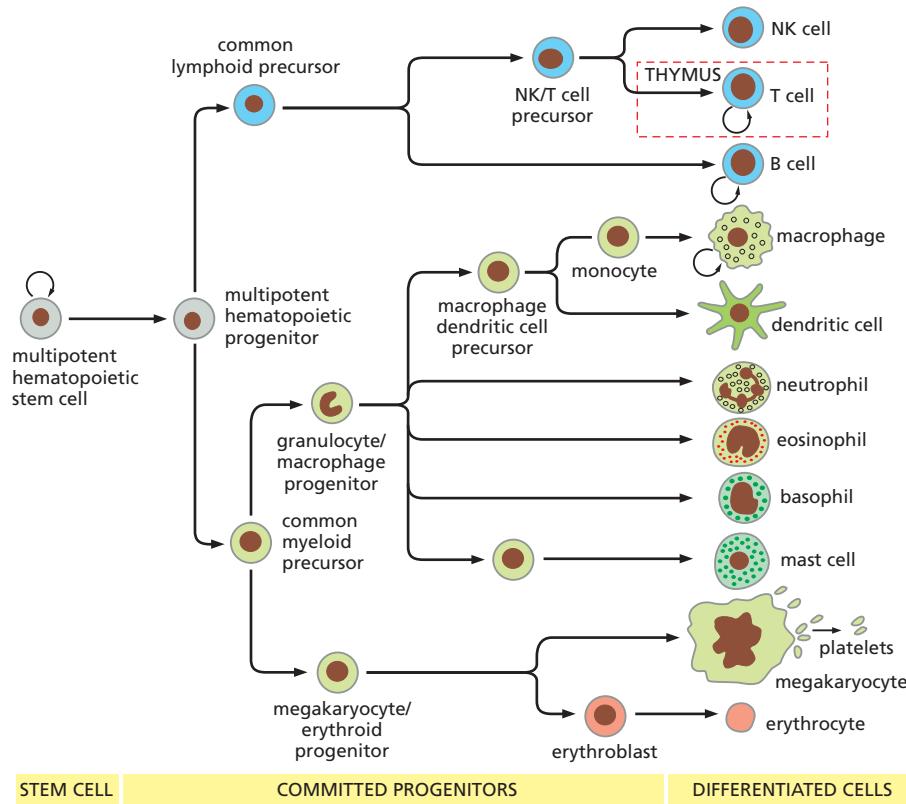


Figure 22–31 A tentative scheme of hematopoiesis. The multipotent stem cell normally divides infrequently to generate either more multipotent stem cells, which are self-renewing, or committed progenitor cells, which are limited in the number of times that they can divide before differentiating to form 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 this cell-lineage diagram. In adult mammals, all of the cells shown develop mainly in the bone marrow—except for T lymphocytes, which as indicated develop in the thymus, and macrophages and osteoclasts, which develop from blood monocytes. Some dendritic cells may also derive from monocytes.

Commitment Is a Stepwise Process

Hematopoietic 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, and the other giving rise to large numbers of all the different types of lymphoid cells. 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 transcription regulators, needed for the production of different subsets of blood cells.

Divisions of Committed Progenitor Cells Amplify the Number of Specialized Blood Cells

Hematopoietic 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 hematopoietic cells. For the same reason, a high rate of blood cell production can be maintained even though the stem-cell division rate is low. The smaller the number of division cycles that the stem cells themselves have to undergo in the course of a lifetime, the lower the risk of generating stem-cell mutations, which would give rise to persistent mutant clones of cells in the body—a particular danger in the hematopoietic system where, as discussed in Chapter 20, a relatively small accumulation of mutations can be sufficient to cause cancer. A low rate of stem-cell division also slows the process of replicative cell senescence (discussed in Chapter 17).

The stepwise nature of commitment means that the hematopoietic 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. Figure 22–31 depicts the hematopoietic family tree. It should be noted, however, that variations are thought to occur: not all stem cells generate the identical patterns of progeny via precisely the same sequence of steps.

Stem Cells Depend on Contact Signals From Stromal Cells

Like the stem cells of other tissues, hematopoietic stem cells depend on signals from their niche, in this case created by the specialized connective tissue of the bone marrow. (This is the site in adult humans; during development, and in non-human mammals such as the mouse, hematopoietic stem cells can also make their home in other tissues—notably liver and spleen.) When they lose contact with their niche, the hematopoietic stem cells tend to lose their stem-cell potential (Figure 22–32). Evidently the loss of potency is not absolute or instantaneous, however, since the stem cells can still survive journeys via the bloodstream to colonize other sites in the body.

Factors That Regulate Hematopoiesis Can Be Analyzed in Culture

While the stem cells depend on contact with bone marrow stromal cells for long-term maintenance, their committed progeny do not, or at least not to the same degree. These cells can thus be dispersed and cultured in a semisolid matrix of dilute agar or methylcellulose, and factors derived from other cells can be added artificially to the medium. The semisolid matrix inhibits migration, so that 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 hematopoiesis and hence to purify them and explore their actions. These substances are glycoproteins and are usually called **colony-stimulating factors (CSFs)**. Some of these factors circulate in the blood and act as hormones, while others act in the bone marrow as secreted local mediators; still others take the form of membrane-bound signals that act through cell-cell contact.

An important example of the latter is a protein called *Steel* or *Stem Cell Factor (SCF)*. This is expressed both in the bone marrow stroma (where it helps to define the stem-cell niche) and along pathways of migration, and it occurs both in a membrane-bound and a soluble form. It binds to a receptor tyrosine kinase called Kit, and it is required during development for guidance and survival not only of hematopoietic cells but also of other migratory cell types—specifically, germ cells and pigment cells.

Erythropoiesis Depends on the Hormone Erythropoietin

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, to which we now turn.

The erythrocyte is by far the most common type of cell in the blood (see Table 22–1). 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 22–33). The erythrocyte therefore cannot grow or divide, and it has 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}

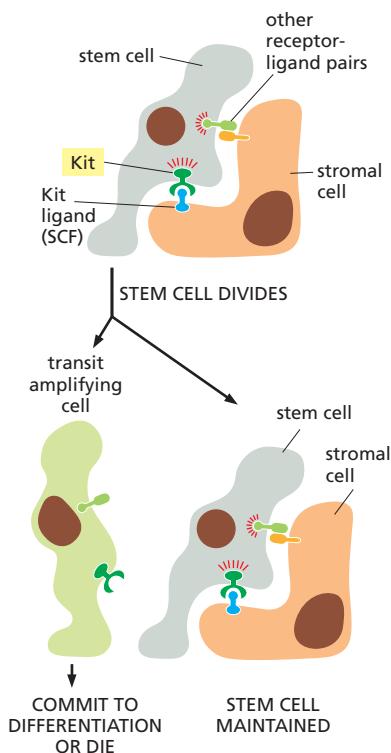


Figure 22–32 Dependence of hematopoietic stem cells on contact with stromal cells. The contact-dependent interaction between the Kit receptor and its ligand is one of several signaling mechanisms thought to be involved in hematopoietic stem-cell maintenance. The real system is certainly more complex. Moreover, the dependence of hematopoietic cells on contact with stromal cells cannot be absolute, since small numbers of the functional stem cells can be found free in the circulation. SCF, stem-cell factor.

Figure 22–33 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.

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, boosts the production of erythrocytes. The effect is rapid: the rate of release of new erythrocytes into the bloodstream rises steeply 1–2 days after an increase in erythropoietin levels in the bloodstream. Clearly, the hormone must act on cells that are 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 whose survival and proliferation are governed by other factors.

Multiple CSFs Influence Neutrophil and Macrophage Production

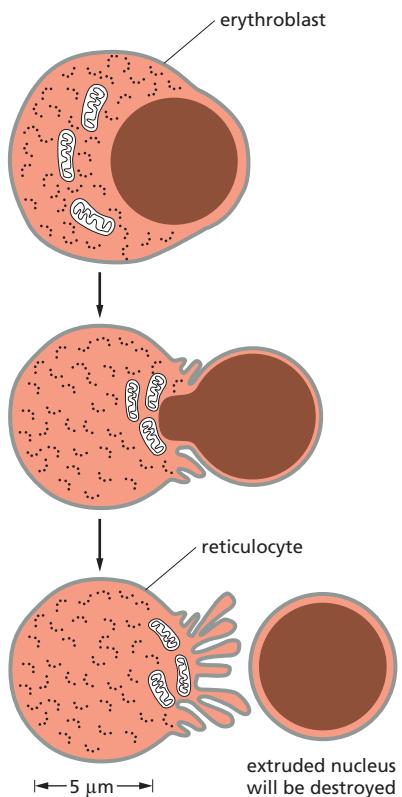
The two classes of cells dedicated to phagocytosis, neutrophils and macrophages, develop from a common progenitor cell called a **granulocyte/macrophage (GM) progenitor cell**. Like the other granulocytes (eosinophils and basophils), neutrophils circulate in the blood for only a few hours before migrating 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.

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. CSFs can be synthesized artificially and are now widely used in human patients to stimulate the regeneration of hematopoietic tissue and to boost resistance to infection.

The Behavior of a Hematopoietic 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 hematopoietic 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 hematopoietic lineage to facilitate differentiation; it might act early to influence commitment; or it might simply increase the probability of cell survival (Figure 22–34). By monitoring the fate of isolated individual hematopoietic cells in culture, it has been possible to show that a single CSF, such as granulocyte/



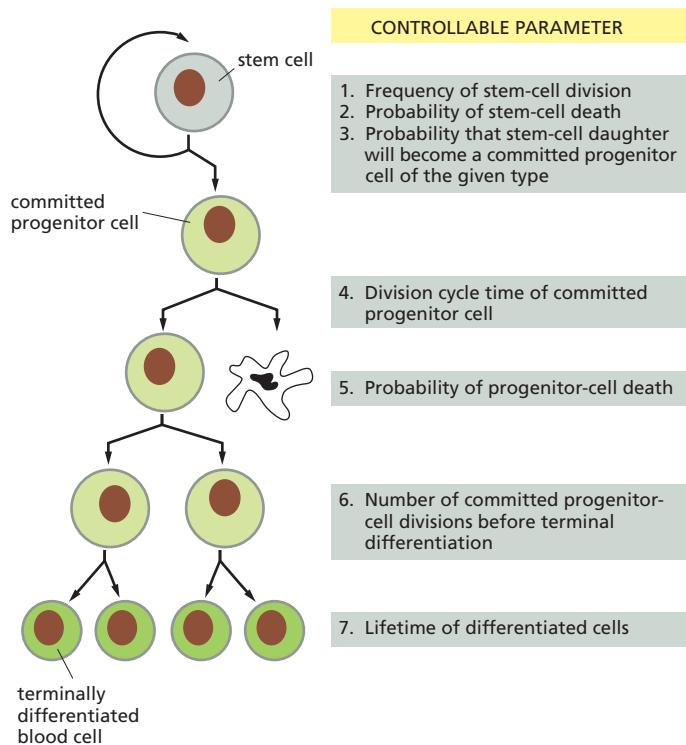


Figure 22–34 Some of the parameters through which the production of blood cells of a specific type might be regulated. Studies in culture suggest that various colony-stimulating factors (CSFs) can affect all of these aspects of hematopoiesis.

macrophage CSF, can exert all these effects, although it is still not clear which are most important *in vivo*.

Studies *in vitro* indicate, moreover, that there is a large element of chance in the way a hematopoietic cell behaves—a reflection, presumably, of “noise” in the genetic control system, as discussed in Chapters 7 and 8. 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 earlier, in Figure 22–31, conveys the impression of a program executed with computer-like logic and precision. Individual cells may be more varied, 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 hematopoietic cells in the absence of CSFs is death by apoptosis (discussed in Chapter 18), and the control of cell survival plays a central part in regulating the numbers of blood cells. The amount of apoptosis in the vertebrate hematopoietic 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. As noted in Chapter 18, 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. Many cases of leukemia arise through mutations that confer this capacity on committed hematopoietic 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, hematopoietic stem cells are found mainly in bone marrow, and they depend on signals from the marrow stromal (connective-tissue) cells to maintain their stem-cell character. The stem cells are few and far between, and they 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 hematopoiesis 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.

REGENERATION AND REPAIR

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 and maintain homeostasis. There are, however, limits to what these natural repair mechanisms can achieve. In most parts of the human brain, for example, nerve cells that die, as in Alzheimer's disease, are not replaced. Likewise, when heart muscle dies for lack of oxygen, as in a heart attack, it is replaced by scar tissue rather than new heart muscle.

Some animals do far better than humans and can regenerate entire organs, such as whole limbs, after amputation. Among the invertebrates, there are some species that can even regenerate all the tissues of the body from a single somatic cell. These phenomena encourage the hope that human cells might be coaxed by artificial measures into similar feats of repair and regeneration, so as 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 1 diabetics, the heart muscle cells that die in a heart attack, and so on. As we learn more about the basic cell biology, these goals, once only a dream, are beginning to seem attainable.

In this section, we start with some examples of the remarkable regenerative abilities of some animal species, as an indication of what is possible in principle. We shall then discuss how we can improve upon the natural repair processes of the human body and treat disease by exploiting the properties of the various types of stem cells found in human tissues. In the final section of the chapter, we shall see how a deeper understanding of the molecular biology of cell differentiation and of stem cells has revealed ways to convert one type of cell into another, opening up radically new possibilities.

Planarian Worms Contain Stem Cells That Can Regenerate a Whole New Body

Schmidtea mediterranea is a small freshwater flatworm, or *planarian*, just under a centimeter long when grown to full size ([Figure 22–35](#)). It has an epidermis, a gut,

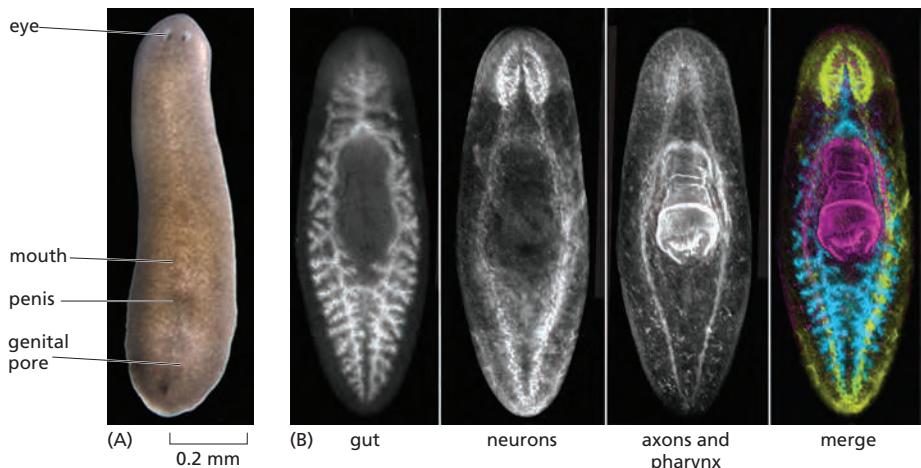


Figure 22–35 The planarian worm, *Schmidtea mediterranea*. (A) External view. (B) Immunostaining with three different antibodies, revealing the internal anatomy. (A, courtesy of A. Sánchez Alvarado; B, from A. Sánchez Alvarado, *BMC Biol.* 10:88, 2012.)

a brain, a pair of primitive eyes, a peripheral nervous system, musculature, and excretory and reproductive organs—most of the basic body parts familiar in other animals, although all relatively simple by vertebrate standards and built from about 20–25 distinct differentiated cell types. For more than a century, planarians such as *Schmidtea* have intrigued biologists because of their extraordinary capacity for regeneration: a small tissue fragment taken from almost any part of the body will reorganize itself and grow to form a complete new animal. This property goes with another: when the animal is starved, it gets smaller and smaller, by reducing its cell numbers while maintaining essentially normal body proportions. This behavior is called *degrowth*, and it can continue until the animal is as little as one-twentieth or even a smaller fraction of its full size. Supplied with food, it will grow back to full size again. Cycles of degrowth and growth can be repeated indefinitely, without impairing survival or fertility.

Underlying this behavior is a process of continual cell turnover. Along with the differentiated cells, which do not divide, there is a population of small, apparently undifferentiated dividing cells called neoblasts. The neoblasts constitute about 20% of the cells in the body and are widely distributed within it; by cell division, they serve as stem cells for the production of new differentiated cells. Differentiated cells, meanwhile, are continually dying by apoptosis, allowing their corpses to be phagocytosed and digested by neighboring cells. Through this cell cannibalism, the constituents of the dying cells can be efficiently recycled. Cell birth continues in a dynamic balance with cell death and cell cannibalism, no matter whether the animal is fed or starved. In conditions of starvation, the balance is evidently tilted toward cell cannibalism, and in conditions of plenty, toward cell birth.

A high dose of x-rays halts all cell division, puts a stop to cell turnover, and destroys the capacity for regeneration. The result is death after a delay of several weeks. The animal can be rescued, however, by injecting into it a single neoblast isolated from an unirradiated donor (Figure 22–36). In a certain proportion of cases, the injected cell divides to form a clone of progeny that eventually repopulate the entire body, creating a healthy regenerative individual with an apparently complete set of differentiated cell types as well as dividing neoblasts. Genetic markers prove that these are all derived from the single neoblast that was injected. It follows that at least some neoblasts are *totipotent* (or at least highly *pluripotent*) stem cells; that is, cells able to give rise to all (or at least almost all) of the cell types that make up the body of a flatworm, including more neoblasts like themselves.

Some Vertebrates Can Regenerate Entire Organs

One might think that such powers of regeneration would be a prerogative of small, simple, primitive animals. But some vertebrates, too, especially fish and amphibians, show remarkable regenerative abilities. A newt, for example, can regenerate

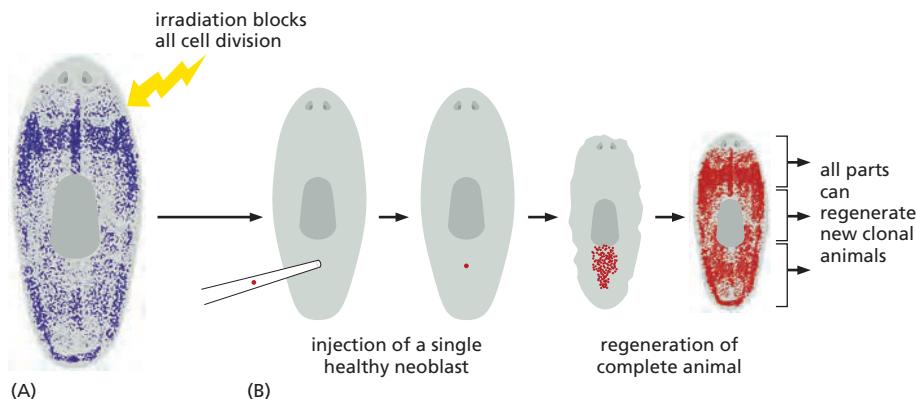


Figure 22–36 Regeneration of a planarian from a single somatic cell.

(A) The distribution of dividing cells (neoblasts, blue) in the adult body. Irradiation blocks all cell division and prevents regeneration, but (B) a single unirradiated neoblast cell injected into the irradiated animal is able to reconstitute all tissues. This eventually produces a complete animal that consists entirely of the progeny of this one cell and can regenerate. (Adapted from E.M. Tanaka and P.W. Reddien, *Dev. Cell* 21:172–185, 2011.)

a whole amputated limb. In this process, differentiated cells seem to revert to an embryonic character by first forming on the amputation stump a *blastema*—a small bud resembling an embryonic limb bud. The blastema then grows and its cells differentiate to form a correctly patterned replacement for the limb that has been lost, in what looks like a recapitulation of embryonic limb development (**Figure 22–37**). A large contribution to the blastema comes from the skeletal muscle cells in the limb stump. These multinucleate cells re-enter the cell cycle, dedifferentiate, and break up into mononucleated cells, which then proliferate within the blastema, before eventually redifferentiating. But do they redifferentiate only into muscle, or do they behave like neoblasts in the planarian and give rise to the full range of cell types needed to reconstruct the missing part of the limb? Careful lineage tracing, using genetic markers, shows (contrary to previous belief) that the cells are restricted according to their origins: muscle-derived cells give rise only to muscle, connective-tissue cells only to connective tissues, epidermal cells only to epidermal cells. The cells in the adult vertebrate body are, after all, less adaptable than the cells of the flatworm: by working in concert, they can replace the lost structure, but each cell type is far from totipotent.

Why a newt can regenerate a whole limb—as well as many other body parts—but a mammal cannot remains a profound mystery.

Stem Cells Can Be Used Artificially to Replace Cells That Are Diseased or Lost: Therapy for Blood and Epidermis

Earlier in this chapter, we saw how mice can be irradiated to kill off their hematopoietic cells, and then rescued by a transfusion of new stem cells, which repopulate the bone marrow and restore blood cell production (see Figure 22–30). In the same way, patients with some forms of leukemia or lymphoma can be irradiated or chemically treated to destroy their cancerous cells along with the rest of their hematopoietic tissue, and then can be rescued by a transfusion of healthy, non-cancerous hematopoietic stem cells. In favorable cases, these can be sorted out from samples of the patient's own hematopoietic tissue before it is ablated. They are then transfused back afterward, avoiding problems of immune rejection.

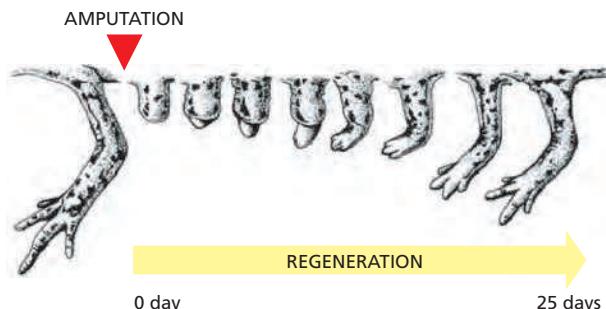
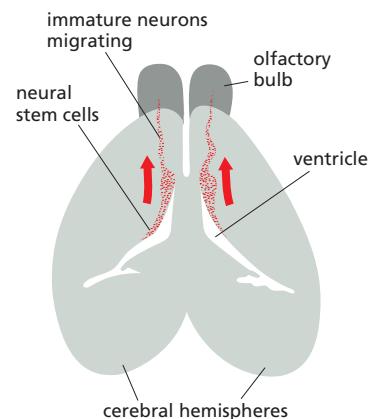


Figure 22–37 Newt limb regeneration.

The time-lapse sequence shows the stages of regeneration after amputation at the level of the humerus. The sequence spans the events of wound healing, dedifferentiation of stump tissues, blastema formation, and redifferentiation. (Courtesy of Susan Bryant and David Gardiner.)

Figure 22–38 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 mentioned earlier. In adult humans, there is a continuing turnover of neurons in the hippocampus, a region specially concerned with learning and memory. (Adapted from B. Barres, *Cell* 97:667–670, 1999. With permission from Elsevier.)



Another 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 (through rather long and complicated procedures) to repopulate the damaged body surface.

Neural Stem Cells Can Be Manipulated in Culture and Used to Repopulate the Central Nervous System

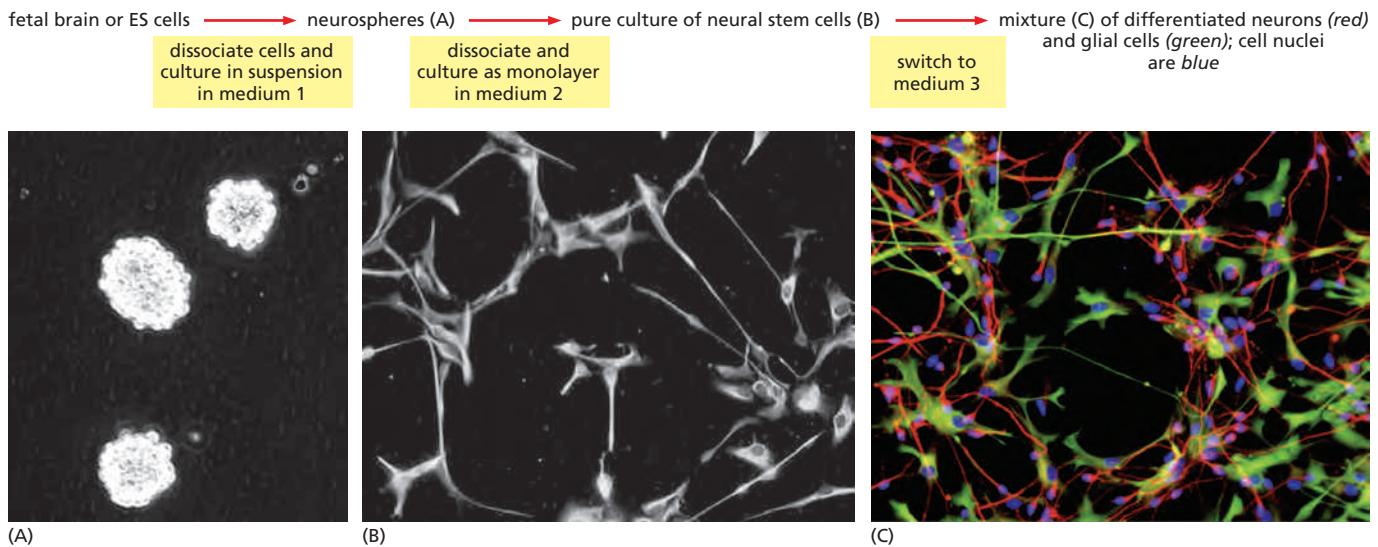
The central nervous system (the CNS) is the most complex tissue in the body, at an opposite extreme from the epidermis. And yet fish and amphibians can regenerate large parts of the brain, spinal cord, and eyes after they have been cut away. In adult mammals, however, these tissues have very little capacity for self-repair, and stem cells capable of generating new neurons are hard to find—so hard to find, indeed, that for many years they were thought to be absent.

We now know, however, that neural stem cells that generate both neurons and glial cells do persist in certain parts of the adult mammalian brain (Figure 22–38). Neuronal turnover occurs on a dramatic scale in certain songbirds' brains, where large numbers of neurons die each year and are replaced by newborn neurons as part of a process by which the birds refine their song for each new breeding season. In the adult human brain, there is a continuing turnover of neurons in the hippocampus, a region specially concerned with learning and memory. Here, plasticity of adult function is associated with turnover of a specific subset of neurons. About 1400 fresh neurons in this class are generated every day, giving a turnover of 1.75% of the population per year.

Fragments taken from self-renewing regions of the adult brain, or from the brain of a fetus, can be dissociated and used to establish cell cultures, where they give rise to 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 either a mixture of neurons and glial cells (Figure 22–39), or just one of these two cell types, according to the composition of the culture medium.

Neural stem cells, whether derived as above or from pluripotent stem cells as described in the next section, can be grafted into an adult brain. Once there, they show a remarkable ability to adjust their behavior to match their new location. Stem cells from the mouse hippocampus, for example, when implanted in the mouse olfactory-bulb-precursor pathway (see Figure 22–38), give rise to neurons that become correctly incorporated into the olfactory bulb. This capacity of neural stem cells and their progeny to adapt to a new environment in animals suggests applications in the treatment for diseases where neurons degenerate, and



for injuries of the central nervous system. For example, might it be possible to use injected neural stem cells to replace the neurons that die in Parkinson's disease or to repair accidents that sever the spinal cord?

Summary

Animals vary in their capacity for regeneration. At one extreme, planarian worms contain stem cells (neoblasts) that support continual turnover of all cell types, and an entire worm can be regenerated from practically any small body fragment or even from a single neoblast cell. Newts can regenerate limbs and other large body parts after amputation, but the cells remain restricted according to their origins: muscle cells in the regenerate derive from muscle, epidermis from epidermis, and so on. In mammals, regeneration is more limited. Nevertheless, it is becoming possible to go beyond the natural limits of wound healing by exploiting stem-cell biology. Thus, certain regions of the nervous system contain stem cells that support production of neurons in these sites throughout life. Neural stem cells can be obtained from these sites or from fetal brains, grown in culture, and then grafted back into other sites in the brain, where they are able to generate neurons appropriate to the new location.

Figure 22–39 Neural stem cells. Shown are 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 embryonic stem (ES) or induced pluripotent stem (iPS) cells (discussed later in this chapter). (Micrographs from L. Conti et al., *PLoS Biol.* 3:1594–1606, 2005.)

CELL REPROGRAMMING AND PLURIPOTENT STEM CELLS

When cells are transplanted from one site in the mammalian body to another or are removed from the body and maintained in culture, they remain largely faithful to their origins. 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, and the whole pattern of self-renewing and differentiated cells in the adult body is amazingly stable. What, at a fundamental molecular level, is the nature of these stable differences between cell types? Is there any way to override the cell memory mechanisms and force a switch from one state to another that is radically different?

We have already discussed these fundamental questions from a general standpoint in Chapter 7. Here we consider them more closely in the context of stem-cell biology, where there has been a recent revolution in our understanding and in our ability to manipulate states of cell differentiation. With further research, these advances would seem to have important practical consequences.

Nuclei Can Be Reprogrammed by Transplantation into Foreign Cytoplasm

If we cannot switch the basic character of a specialized cell by changing its environment, can we do so by interfering with its inner workings in a more direct and drastic way? An extreme treatment of this sort is to take the nucleus of the cell and transplant it into the cytoplasm of a large cell of a different type. If the specialized character is defined and maintained by cytoplasmic factors, the transplanted nucleus should switch its pattern of gene expression to conform with that of the host cell. In Chapter 7, we described a famous experiment of this sort, using the frog *Xenopus*. In this experiment, the nucleus of a differentiated cell (a cell from the lining of a tadpole's gut) was used to replace the nucleus of an oocyte (an egg-cell precursor arrested in prophase of the first meiotic division, in readiness for fertilization). The resulting hybrid cell went on, in a certain fraction of cases, to develop into a complete normal frog (see Figure 7-2A). This was crucial evidence for what is now a central principle of developmental biology: the cell nucleus, even that of a differentiated cell, contains a complete genome, capable of supporting development of all normal cell types. At the same time, the experiment showed that cytoplasmic factors can indeed reprogram a nucleus: the oocyte cytoplasm can drive the gut cell nucleus back to an early embryonic state, from which it can then step through the changing patterns of gene expression that lead all the way to a complete adult organism.

The full story, however, is not quite so simple. First, the reprogramming in such experiments is not perfect. When the transplanted nucleus is taken from a gut cell, for example, a gene that is normally specific to the gut is found to be expressed persistently, even in the muscle cells of the final animal. Second, the experiment succeeds in only a limited proportion of cases, and this success rate becomes lower and lower, the more mature the animal from which the transplanted nucleus is taken: very large numbers of transplantations must be done to score a single success if the nucleus comes from a differentiated cell of an adult frog.

Nuclear transplantation can be done in mammals too, with basically similar results. Thus, a nucleus taken from a differentiated cell in the mammary gland of an adult sheep and transplanted into an enucleated sheep's egg was able to support development of an apparently normal sheep—the famous Dolly. Again, the success rate is low: many transplantations have to be done to obtain one such individual.

Reprogramming of a Transplanted Nucleus Involves Drastic Epigenetic Changes

In a typical fully differentiated cell, there seem to be mechanisms maintaining the pattern of gene expression that cytoplasmic factors cannot easily override. An obvious possibility is that the stability of the pattern of gene expression in an adult cell may depend, in part at least, on self-perpetuating modifications of chromatin, as discussed in Chapter 4. As explained in Chapter 7, the phenomenon of X-inactivation in mammals provides a clear example of such epigenetic control. Two X chromosomes exist side by side in each female cell, exposed to the same chemical environment, but while one remains active, the other persists from one cell generation to the next in a condensed inactive state; cytoplasmic factors cannot be responsible for the difference, which must instead reflect mechanisms intrinsic to the individual chromosome. Elsewhere in the genome also, controls at the level of chromatin act in combination with other forms of regulation to govern the expression of each gene. Genes can be shut down completely, or switched on constitutively, or maintained in a labile state where they can be readily switched on or off according to changing circumstances.

The reprogramming of a nucleus transplanted into an oocyte involves dramatic changes in chromatin. The nucleus swells, increasing its volume 50-fold as the chromosomes decondense; there is a wholesale alteration in patterns of methylation of DNA and histones; the standard histone H1 (the histone that links

adjacent nucleosomes) is replaced by a variant form that is peculiar to the oocyte and early embryo; and the preexisting type of histone H3 is also replaced at many sites by a distinct isoform. Evidently, the egg contains factors that reset the state of the chromatin in the nucleus, wiping out old histone modifications on chromatin and imposing new ones. Reprogrammed in this way, the genome becomes competent once again to initiate embryonic development and to give rise to the full range of differentiated cell types.

Embryonic Stem (ES) Cells Can Generate Any Part of the Body

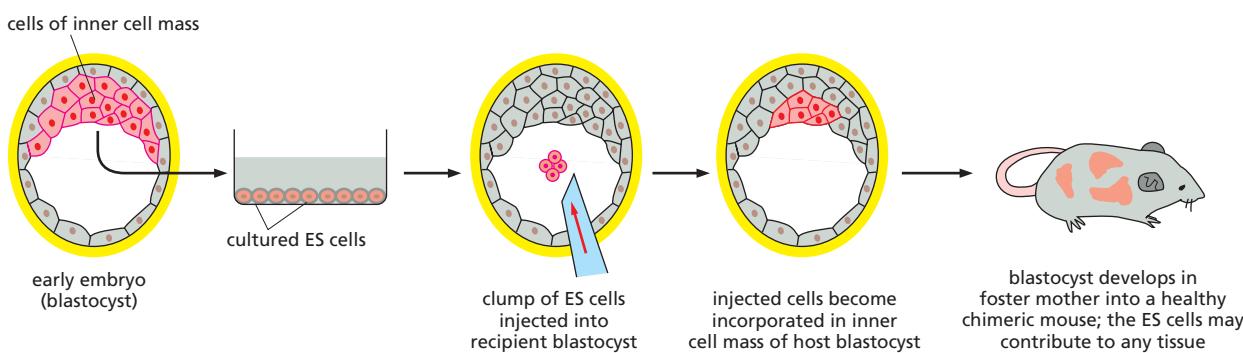
A fertilized egg, or an equivalent cell produced by nuclear transplantation, is a remarkable thing: it can generate a whole new multicellular individual, and that means that it can give rise to every normal type of specialized cell, including even egg or sperm cells for production of the next generation. A cell in such a state is said to be **totipotent**, and it is said to be **pluripotent** if it can give rise to most cell types but not absolutely all. Nevertheless, such a progenitor is not a stem cell: it is not self-renewing, but is instead dedicated to a program of progressive differentiation. If it were the only available starting point for study and exploitation of pluripotent cells, the enterprise would require a continual supply of fresh fertilized eggs or fresh nuclear transplantation procedures—an awkward requirement for studies in experimental animals, and unacceptable for practical applications in humans.

Here, however, nature has been unexpectedly kind to scientists. 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**. These originate from the inner cell mass of the early embryo (the cluster of cells that give rise to the body of the embryo proper, as opposed to extraembryonic structures), and they have an extraordinary property: given suitable culture conditions, they will continue proliferating indefinitely and yet retain an unrestricted developmental potential. Their only limitation is that they do not give rise to extraembryonic tissues such as those of the placenta. Thus they are classified as pluripotent, rather than totipotent. But this is a minor restriction. 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, integrating perfectly into whatever site they may come to occupy, and adopting the character and behavior that normal cells would show at that site. They can even give rise to germ cells, from which a new generation of animals can be derived (**Figure 22–40**).

ES cells let us move between cell culture, where we can use powerful techniques for genetic transformation and selection, and the intact organism, where we can discover how such genetic manipulations affect development and physiology. Thus, ES cells have opened the way to efficient genetic engineering in mammals, leading to a revolution in our understanding of mammalian molecular biology.

Cells with properties similar to those of mouse ES cells can now be derived from early human embryos and from human fetal germ cells, and even, as we shall explain below, from differentiated cells taken from adult mammalian tissues. In this way, one can obtain a potentially inexhaustible supply of pluripotent

Figure 22–40 Production and pluripotency of ES cells. ES cells are derived from the inner cell mass (ICM) of the early embryo. The ICM cells are transferred to a culture dish containing an appropriate medium, where they become converted to ES cells and can be kept proliferating indefinitely without differentiating. The ES cells can be taken at any time—after genetic manipulation, if desired—and injected back into the inner cell mass of another early embryo. There they take part in formation of a well-formed chimeric animal that is a mixture of ordinary and ES-derived cells. The ES-derived cells can differentiate into any of the cell types in the body, including germ cells from which a new generation of mice can be produced. These next-generation progeny are no longer chimeric, but consist of cells that all inherit half their genes from the cultured ES cell line.



cells. Grown in culture, these can be manipulated, by suitable choice of culture conditions, to give rise to large quantities of almost any type of differentiated cell, opening the way to important practical applications. Before discussing them, however, we consider the underlying biology.

A Core Set of Transcription Regulators Defines and Maintains the ES Cell State

What is it that gives ES cells and related types of pluripotent stem cells their extraordinary capabilities? What can they tell us about the fundamental mechanisms underlying stemness, cell differentiation, and the stability of the differentiated state?

For some attributes, the answer is simple. For example, an essential feature of ES cells is that they must avoid senescence. As discussed in Chapter 17, this is the fate of fibroblasts and many other types of somatic cells: they are limited in the number of times they will divide, in part at least because they lack telomerase activity, with the result that their telomeres become progressively eroded in each division cycle, leading eventually to cell-cycle arrest. ES cells, by contrast, express high levels of active telomerase, allowing them to escape senescence and continue dividing indefinitely. This is a property shared with other, more specialized types of stem cells, such as those of the adult intestine, which similarly can carry on dividing for hundreds or thousands of cycles.

The deeper problem is to explain how the whole complex pattern of gene expression in an ES cell is organized and maintained. As a first step, one can look for genes expressed specifically in ES cells or in the corresponding pluripotent cells of the early embryo. This approach identifies a relatively small number of candidate ES-critical genes; that is, genes that seem to be essential in one way or another for the peculiar character of ES cells. A gene called *Oct4*, for example, is exclusively expressed in ES cells and in related classes of cells in the intact organism—specifically, in the germ-cell lineage and in the inner cell mass and its precursors. *Oct4* codes for a transcription regulator. When it is lost from ES cells, they lose their ES cell character; and when it is missing in an embryo, the cells that should specialize as inner cell mass are diverted into an extraembryonic pathway of differentiation and their development is aborted.

Fibroblasts Can Be Reprogrammed to Create Induced Pluripotent Stem Cells (iPS Cells)

In Chapter 7, we saw that fibroblasts and some other cell types can be driven to switch their character and differentiate as muscle cells if the master muscle-specific transcription regulator *MyoD* is artificially expressed in them. Could the same technique be used to convert adult cell types into ES cells, through forced expression of factors such as *Oct4*? This question was tackled by transfecting fibroblasts with retroviral vectors carrying genes that one might hope to have such an effect. A total of 24 candidate ES-critical genes were tested in this way. None of them was able by itself to cause the conversion; but in certain combinations they could do so. In 2006, the first breakthrough experiments whittled down the requirement to a core set of four factors, all of them transcription regulators: *Oct4*, *Sox2*, *Klf4*, and *Myc*, known as the OSKM factors for short. When coexpressed, these could reprogram mouse fibroblasts, permanently converting them into cells closely similar to ES cells (Figure 22–41). ES-like cells created in this way are called **induced pluripotent stem cells**, or **iPS cells**. Like ES cells, iPS cells can continue dividing indefinitely in culture, and when incorporated into a mouse blastocyst they can participate in creation of a perfectly formed chimeric animal. In this animal, they can contribute to the development of any tissue and can turn into any differentiated cell type, including functional germ cells from which a new generation of mice can be raised (see Figure 22–40).

iPS cells can now be derived from adult human cells and from various other differentiated cell types besides fibroblasts. Numerous methods can be used to drive

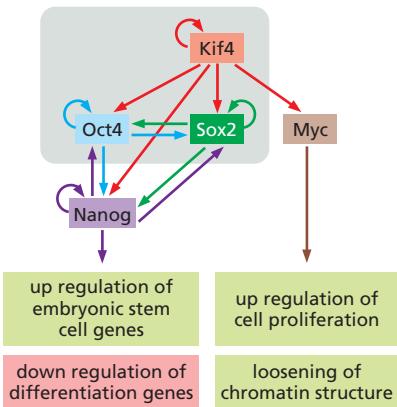


Figure 22–41 Reprogramming fibroblasts to iPS cells with the OSKM factors. As indicated, the master gene regulator proteins Oct4, Sox2, and Klf4 (OSK) induce both their own and each other's synthesis (gray shading). This generates a self-sustaining feedback loop that helps to maintain cells in an embryonic stem cell-like state, even after all of the experimentally added OSKM initiators have been removed. Myc overexpression speeds up early stages of the reprogramming process through the mechanisms shown (see Figure 17–61). Stable reprogramming also involves the permanently induced expression of the *Nanog* gene, which produces an additional master transcription regulator. (Adapted from J. Kim et al., *Cell* 132:1049–1061, 2008.)

expression of the transforming OSKM factors, including methods that leave no trace of foreign DNA in the reprogrammed cell. Variations of the original cocktail of transcription regulators can drive the conversion, with different specialized cell types having somewhat different requirements. Myc overexpression, for example, turns out not to be absolutely necessary, although it enhances the efficiency of the process. And differentiated cell types may express some of the required factors as part of their normal phenotype. For example, cells of the dermal papilla of hair follicles already express Sox2, Klf4, and Myc; to convert them into iPS cells, it is enough to force them artificially to express Oct4. Oct4, indeed, seems to have a central role and to be generally indispensable for the creation of iPS cells.

Reprogramming Involves a Massive Upheaval of the Gene Control System

Converting a differentiated cell into an iPS cell is not like flicking a switch on some predictable, precisely engineered piece of machinery. Only a few of the cells that receive the OSKM factors will actually become iPS cells—one in several thousand in the original experiments, and still only a small minority with more recent, improved techniques. In fact, the success of the original experiments depended on clever selection to pick out those few cells where the conversion had occurred ([Figure 22–42](#)).

Conversion to an iPS character by the OSKM factors is not only inefficient but also slow: fibroblasts take ten days or more from introduction of the conversion factors before they begin to express markers of the iPS state. This suggests that the transformation involves a long cascade of changes. These changes are being extensively studied, and they affect both the expression of individual genes and the state of the chromatin. The results of one such study are outlined in [Figure 22–43](#). The process begins with a Myc-induced cell proliferation and loosening of chromatin structure that promotes the binding of the other three master regulators to many hundreds of different sites in the genome. At a large proportion of these sites, Oct4, Sox2, and Klf4 all bind in concert. The binding sites include the endogenous *Oct4*, *Sox2*, and *Klf4* genes themselves, which eventually creates the types of positive feedback loops just described that makes expression of these genes self-sustaining (see [Figure 22–41](#)). But self-induction of *Oct4*, *Sox2*, and *Klf4* is only a small part of the transformation that occurs. The three core factors activate some target genes and repress others, producing a cascade of effects that reorganize the gene control system globally and at every level, changing the patterns of histone modification, DNA methylation, and chromatin compaction, as well as the expression of innumerable proteins and noncoding RNAs. By the end of this complex process, the resulting iPS cell is no longer dependent on the artificially generated factors that triggered the change: it has settled into a stable, self-sustaining state of coordinated gene expression, making its own *Oct4*, *Sox2*, *Klf4*, and Myc (and all the other essential ingredients of a pluripotent stem cell) from its own endogenous copies of the genes.

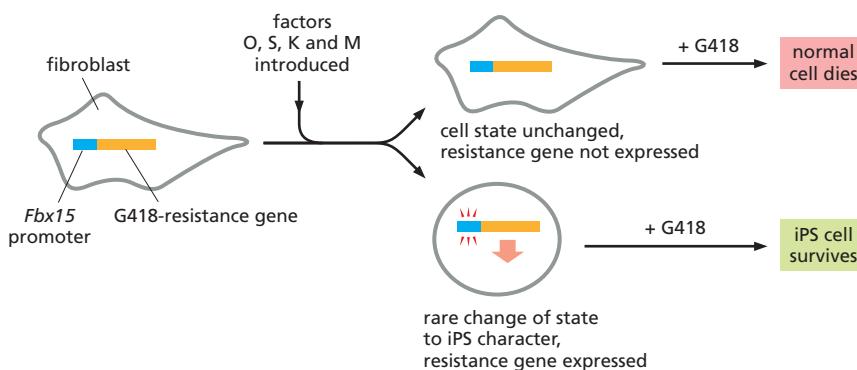


Figure 22–42 A strategy used to select cells that have converted to an iPS character. The experiment makes use of a gene (*Fbx15*) that is present in all cells but is normally expressed only in ES and early embryonic cells (although not required for their survival). A fibroblast cell line is genetically engineered to contain a gene that produces an enzyme that degrades G418 under the control of the *Fbx15* regulatory sequence. G418 is an aminoglycoside antibiotic that blocks protein synthesis in both bacteria and eukaryotic cells. When the OSKM factors are artificially expressed in this cell line, a small proportion of the cells undergo a change of state and activate the *Fbx15* regulatory sequence, driving expression of the G418-resistance gene. When G418 is added to the culture medium, these are the only cells that survive and proliferate. When tested, they turn out to have an iPS character.

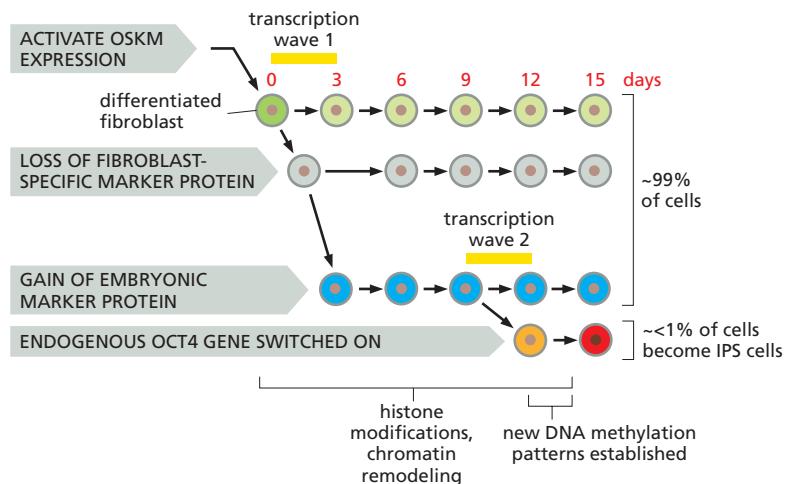


Figure 22–43 A summary of some of the major events that accompany the reprogramming of mouse fibroblasts to iPS cells. By sorting cells at various times after the OSKM induction shown, one can carry out detailed biochemical analyses on the different cell populations shown. This led to the discovery that two major waves of new gene transcription are induced, but that the second wave occurs only in the subset of cells expressing an embryonic marker protein. Some 1500 genes are found to be differentially expressed in these cells, compared to the large majority of cells that fail to progress toward iPS cells. As indicated, major DNA methylation changes are observed only after the alteration of chromatin structures.

In the first transcription wave, among the genes prominently induced are those for cell proliferation, metabolism, and cytoskeletal organization; in contrast, genes associated with fibroblast development are repressed. In the second transcription wave, genes required for embryonic development and for stem cell maintenance are induced. (Adapted from J.M. Polo et al., *Cell* 151:1617–1632, 2012.)

An Experimental Manipulation of Factors that Modify Chromatin Can Increase Reprogramming Efficiencies

The low efficiency and slow rate of conversion suggest that there is some barrier blocking the switch from the differentiated state to the iPS state in these experiments, and that overcoming this barrier is a difficult process that involves a large element of chance. Likewise, the outcome is variable, with significant differences between the individual lines of transformed cells that are generated, even when the initial differentiated cells are genetically and phenotypically identical. Only some of the candidate iPS lines pass all the tests of pluripotency. At a molecular level, there are differences even among the fully validated iPS cells: although they share many features, they vary in details of their gene expression patterns and, for example, in their patterns of DNA methylation.

Overcoming these difficulties will be critical for improving our understanding of how cell specialization is controlled and organized in multicellular organisms; it should also facilitate many medical advances. Thus, intensive research is being carried out on the reprogramming process. One approach aims at obtaining a much clearer picture of the role that chromatin structures play in gene regulation in eukaryotes.

From our discussion of nuclear transplantation, one might expect that any reprogramming of a differentiated cell would require a radical and widespread change in the chromatin structure of selected genes. Not only are such changes observed, but a large number of different experiments reveal that the efficiency of the reprogramming process can be substantially increased by altering the activity of proteins that affect chromatin structure. **Figure 22–44** categorizes some of the factors that have been shown to enhance the transformation of fibroblasts to iPS cells; those in the top three rows—chromatin remodelers, histone modifying enzymes, and histone variants—are especially well known to have profound effects on the organization of nucleosomes in chromatin (discussed in Chapter 4).

We can only touch briefly here on the massive amounts of data that have been accumulating in this exciting research area. The major challenge that remains is to obtain a systems-level model for the complex set of biochemical changes that are involved in reprogramming. For example, which chromatin changes come first, and which then follow? How can these be triggered by the master transcription regulators through their binding to specific DNA sequences, and why do many cells in a population appear resistant to these effects?

ES and iPS Cells Can Be Guided to Generate Specific Adult Cell Types and Even Whole Organs

We can think of embryonic development in terms of a series of choices presented to cells as they follow a road that leads from the fertilized egg to terminal

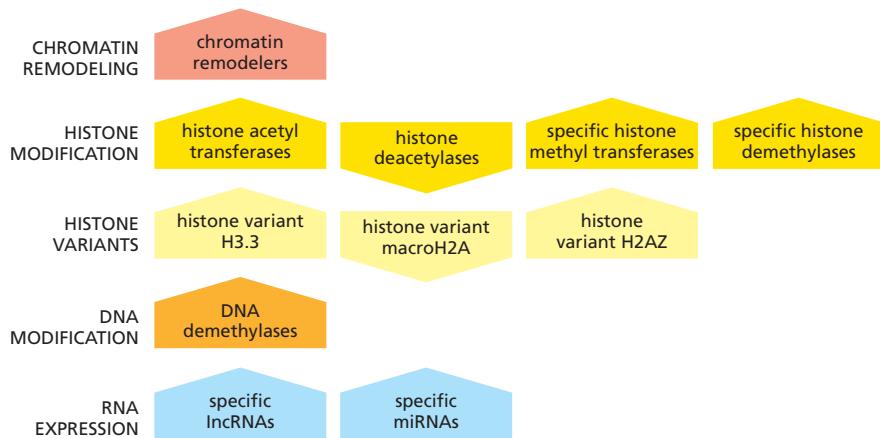
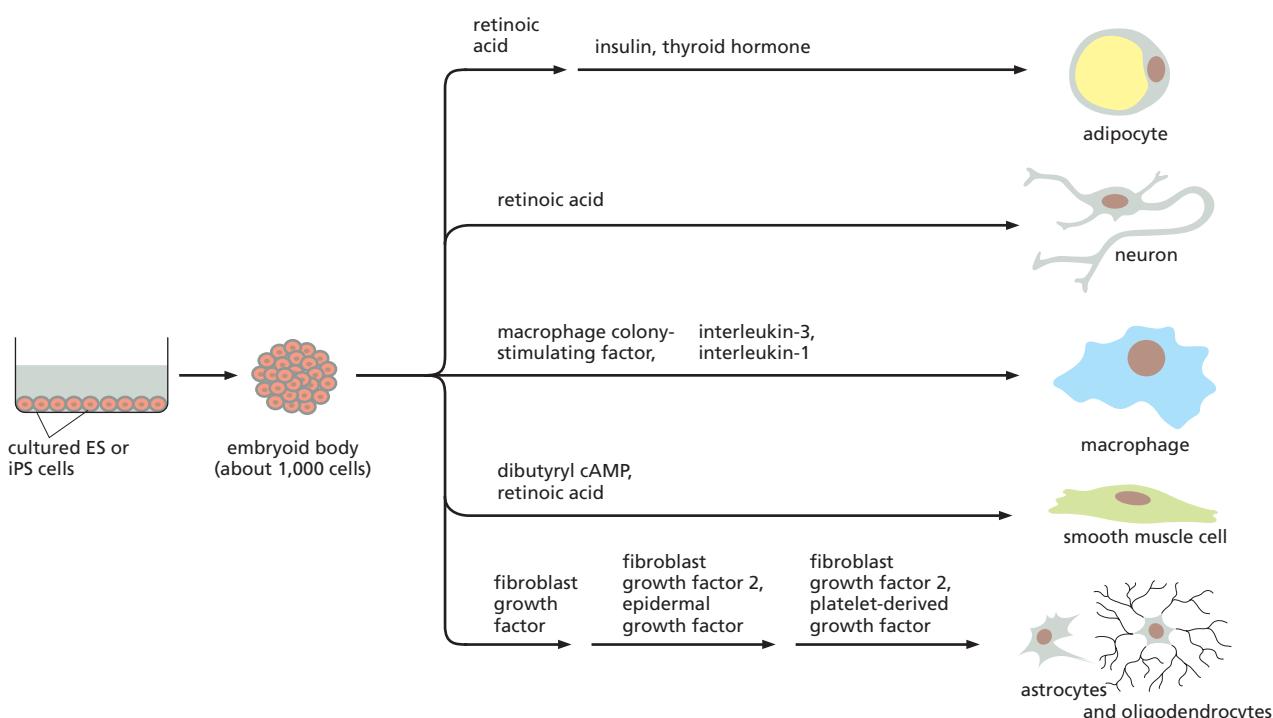


Figure 22–44 Factors that have been observed to enhance reprogramming efficiency. Emphasized here are those factors that can alter chromatin states, with those in the top three rows having the most direct effects. An up arrow indicates that reprogramming is increased when the activity of the indicated factor is increased; a down arrow indicates that reprogramming is increased when the activity of the indicated factor is decreased. Thus, for example, increased activity of histone acetyl transferases and increased activity of histone deacetylases have opposite effects, as expected from their biochemical activities (see p. 196).

differentiation. After their long sojourn in culture, the ES cells or iPS cells and their progeny can still read the signs at each branch in the highway and respond as normal embryonic cells would. If ES or iPS 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 of the type known as a *teratoma*, containing a mixture of cell types inappropriate to the site in the body.

In culture, by exposing the ES or iPS cell to an appropriate sequence of signal proteins and growth factors, delivered with the right timing, it is possible to guide the cell along a pathway that approximates a normal developmental pathway, so as to convert it into one of the standard specialized adult cell types (Figure 22–45 and Movie 22.5). Success requires trial and error, but has now been achieved for many different final specialized states, including neuronal, muscular, and intestinal cell types. In a few cases, it has even been possible, by careful manipulation of the culture conditions, to get ES or iPS cells to interact with one another so as to construct an entire organ, albeit on a small scale (Figure 22–46).

Figure 22–45 Production of differentiated cells from ES or iPS cells in culture. These cells can be cultured indefinitely as pluripotent cells when attached as a monolayer to a dish. Alternatively they can be detached and allowed to form aggregates called embryoid bodies, which causes the cells to 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.)



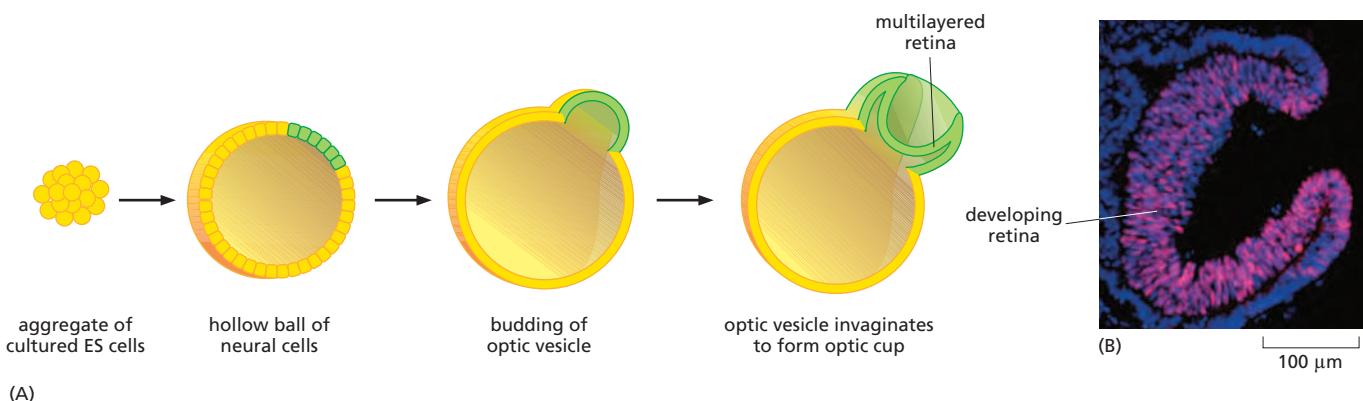


Figure 22–46 Cultured ES cells can give rise to a three-dimensional organ. (A) Remarkably, under appropriate conditions, mouse ES cells in culture can proliferate, differentiate, and interact to form a three-dimensional, eye-like structure, which includes a multilayered retina similar in organization to the one that forms *in vivo*. (B) Fluorescent micrograph of an optic cup formed by ES cells in culture. The structure includes a developing retina, containing multiple layers of neural cells, which produce a protein (pink) that serves as a marker for retinal tissue. (B, from M. Eiraku, N. Takata, H. Ishibashi et al., *Nature* 472:51–56, 2011. With permission from Macmillan Publishers Ltd.)

Cells of One Specialized Type Can Be Forced to Transdifferentiate Directly Into Another

The route we have just described, from one mode of differentiation to another via conversion to an iPS cell, seems needlessly roundabout. Could we not convert cell type A into cell type B directly, without backtracking to the embryonic-like iPS state? For many years, it has been known that such *transdifferentiation* can be achieved in a few special cases, such as the conversion of fibroblasts into skeletal muscle cells by forced expression of MyoD (see p. 396). But now, with the insights that have come from the study of ES and iPS cells, ways are being found to bring about such interconversions in a much wider range of cases.

An elegant example comes from studies of the heart. By forcing expression of an appropriate combination of factors—not Oct4, Sox2, Klf4, and Myc, but Gata4, Mef2c, and Tbx5—it is possible to convert heart fibroblasts directly into heart muscle cells. This has been done in the living mouse, using retroviral vectors, and the transformation occurs with high efficiency when the vectors carrying the transgenes are injected directly into the heart muscle tissue itself. Although they occupy only a small fraction of the tissue volume, the fibroblasts in the heart outnumber the heart muscle cells, and they survive in large numbers even where the heart muscle cells have died. Thus, in a typical nonfatal heart attack, where heart muscle cells have died for lack of oxygen, the fibroblasts proliferate and make collagenous matrix so as to replace the lost muscle with a fibrous scar. This is a poor sort of repair. By forcing expression of the appropriate factors in the heart, as described above, it has proved possible, in the mouse at least, to do better than nature and regenerate lost heart muscle by transdifferentiation of heart fibroblasts.

We are still a long way from putting this technique into practice as a treatment for heart attacks in humans, but it shows what the future may hold—not only for this medical problem, but for many others.

ES and iPS Cells Are Useful for Drug Discovery and Analysis of Disease

A large part of the excitement surrounding ES and iPS cells and the technology of transdifferentiation comes from the prospect of using the artificially generated cells for tissue repair. It begins to seem that virtually any type of tissue might be replaceable, allowing treatment of degenerative diseases that have previously had no cure. Research in this area is moving rapidly, but there are many difficulties to be overcome.

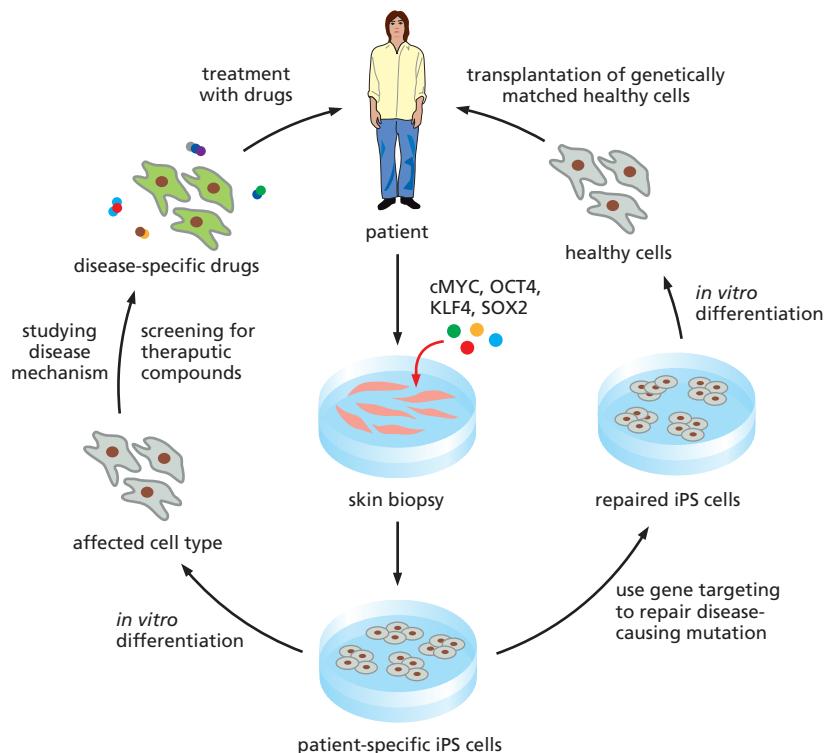


Figure 22–47 Use of iPS cells for drug discovery and for analysis and treatment of genetic disease. The left side of the diagram shows how differentiated cells that are generated from iPS cells derived from a patient with a genetic disease can be used for analysis of the disease mechanism and for discovery of therapeutic drugs. The right side of the diagram shows how the genetic defect might be repaired in the iPS cells, which could then be induced to differentiate in an appropriate way and grafted back into the patient without danger of immune rejection. (Based on D.A. Robinton and G.Q. Daley, *Nature* 481:295–305, 2012).

With the advent of iPS cells and direct transdifferentiation, at least one major hurdle has been surmounted, in principle at least: the problem of immune rejection. ES cells, because they are created from early embryos that generally come from unrelated donors, will never be genetically identical to the cells of the patient receiving the transplant. The transplanted cells and their progeny are therefore liable to rejection by the immune system. Both iPS and transdifferentiated cells, in contrast, can be generated from a small sample of the patient's own tissue and so should escape immune attack when transplanted back into the same individual.

Tissue repair by transplantation, however, is not the only application for which ES, iPS, and transdifferentiated cells can be used: there are other ways in which they promise to be more immediately valuable. In particular, they can be used to generate large, homogeneous populations of specialized cells of any chosen type in culture; and these can serve for investigation of disease mechanisms and in the search for new drugs acting on a specific cell type (Figure 22–47).

Where a disease has a genetic cause, we can derive iPS cells from sufferers and use these cells to produce the specific cell types that malfunction, to investigate how the malfunction occurs, and to screen for drugs that might help to put it right. *Timothy syndrome* provides an example. In this rare genetic condition, there is a severe, life-threatening disorder in the rhythm of the heart beat (as well as several other abnormalities), as a result of a mutation in a specific type of Ca^{2+} channel. To study the underlying pathology, researchers took skin fibroblasts from patients with the disorder, generated iPS cells from the fibroblasts, and drove the iPS cells to differentiate into heart muscle cells. These cells, when compared with heart muscle cells prepared similarly from normal control individuals, showed irregular contractions and abnormal patterns of Ca^{2+} influx and electrical activity that could be characterized in detail. From this finding, it is a small step to development of an *in vitro* assay for drugs that might correct the misbehavior of the heart muscle cells.

This approach to drug discovery—where iPS cells are prepared from the individual patient, differentiated into the relevant cell type, and used to test candidate drugs *in vitro*—would seem to represent a huge advance on the slow, costly traditional methods that involve administration of test compounds to large numbers of people.

Summary

In the adult mammalian body, the various types of stem cells are highly specialized, each giving rise to a limited range of differentiated cell types. Cells become restricted to specific pathways of differentiation during embryonic development. One way to force a return to a pluripotent or totipotent state is by nuclear transplantation: the nucleus of a differentiated cell can be injected into an enucleated oocyte, whose cytoplasm reprograms the genome back to an approximation of an early embryonic state. This allows production of an entire new individual. The reversion of the genome to this state involves radical, genome-wide changes in chromatin structure and DNA methylation.

Remarkably, cells taken from the inner cell mass of an early mammalian embryo can be propagated in culture indefinitely in a pluripotent state. When transplanted back into a host early embryo, these embryonic stem (ES) cells can contribute cells to any tissue, including the germ line. ES cells have been invaluable for genetic engineering in mice. Cells with similar properties, called induced pluripotent stem cells (iPS cells), can be generated from adult differentiated cells such as fibroblasts by forced expression of a cocktail of key transcription regulators. A similar method can be used to reprogram adult cells directly from one specialized state to another. In principle, iPS cells generated from cells biopsied from an adult human patient could be used for tissue repair in that same individual, avoiding the problem of immune rejection. More immediately, they provide a source of specialized cells that can be used to analyze in vitro the effects of mutations affecting human cells and for screening for drugs for treatment of genetic diseases.

WHAT WE DON'T KNOW

- What determines tissue and organ size? How do the cells in each tissue know when to terminate their growth and division, so as to limit the size of an organ or tissue appropriately?
- What is the fundamental molecular difference that distinguishes a stem cell?
- How is the correct balance between stem cells, progenitor cells, and differentiated cells maintained in a tissue or organ?
- What role does chromatin structure play in cell memory and in cell reprogramming?
- How are molecules inherited asymmetrically during cell division?
- How do germ cells avoid aging?

PROBLEMS

Which statements are true? Explain why or why not.

22–1 In the small intestine, stem cells in the crypts divide asymmetrically to maintain the population of cells that make up the villi; after each division, one daughter remains a stem cell and the other begins to divide rapidly to produce differentiated progeny.

22–2 Stem cells, being stem cells, are by definition the same in all tissues.

22–3 Every tissue that can be renewed is renewed from a tissue-specific population of stem cells.

22–4 Disturbance of the balance in the activities of osteoblasts and osteoclasts in favor of osteoclasts can give rise to the condition known as osteoporosis, the brittle-bone syndrome of the elderly.

Discuss the following problems.

22–5 In the 1950s, scientists fed ^3H -thymidine to rats to label cells that were synthesizing DNA, and then followed the fates of labeled cells for periods of up to a year. They found three patterns of cell labeling in different tissues. Cells in some tissues such as neurons in the central nervous system and the retina did not get labeled. Muscle, kidney, and liver, by contrast, each showed a small number of labeled cells that retained their label, apparently without further division or loss. Finally, cells such as those in the squamous epithelia of the tongue and esophagus were

labeled in fairly large numbers, with radioactive pairs of nuclei visible in 12 hours; however, the labeled cells disappeared over time. Which of these three patterns of labeling would you expect to see if the labeled cells were generated by stem cells? Explain your answer.

22–6 At any given time, intestinal crypts of mice comprise about 15 stem cells and 10 Paneth cells. After cell division, which occurs about once a day, the daughter cells remain stem cells only if they maintain contact with a Paneth cell. This constant competition for Paneth-cell contact raises the possibility that crypts might become monoclonal over time; that is, the crypt cells at one point in time might derive from only 1 of the 15 stem cells that existed at some earlier time. To test this possibility, you use the so-called confetti marker that upon activation expresses any one of three fluorescent proteins in the stem cells of the crypt. You then examine crypts at various times to determine whether they contain cells with multiple colors or only one color (**Figure Q22–1**). Do the crypts become monoclonal over time or not? How can you tell?

22–7 The origin of new β cells of the pancreas—from stem cells or from preexisting β cells—was not resolved until a decade ago, when the technique of lineage tracing was used to decide the issue. Using transgenic mice that expressed a tamoxifen-activated form of Cre recombinase under the control of the insulin promoter, which is active only in β cells, investigators could remove an inhibitory segment of DNA and thereby allow expression of human placental alkaline phosphatase (HPAP), which can be detected by histochemical staining. After a pulse of tamoxifen that converted about 30% of β cells in young mice to

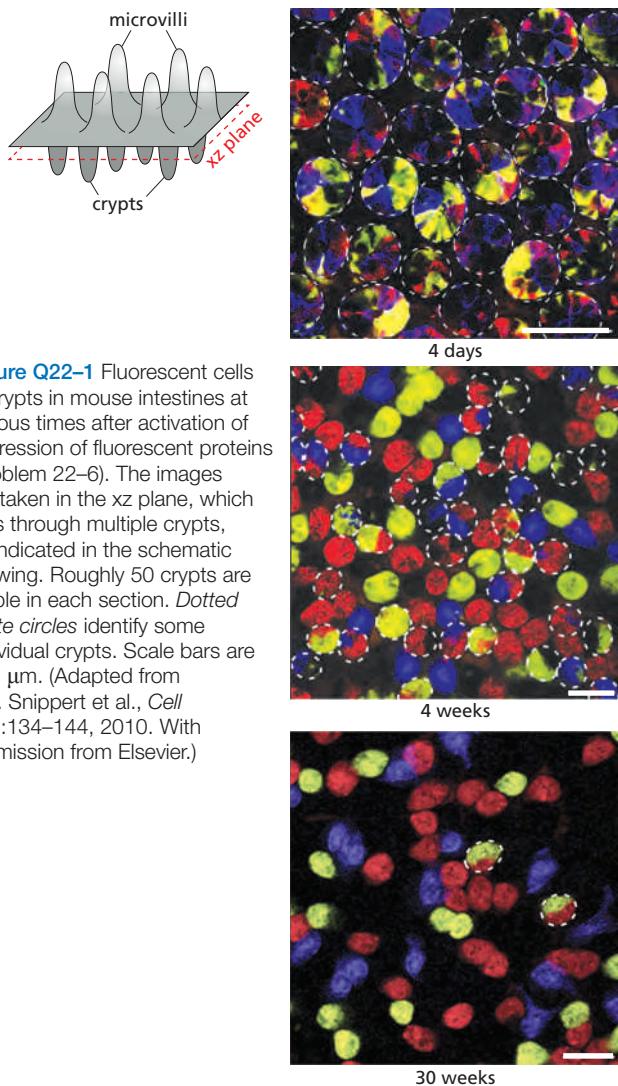


Figure Q22-1 Fluorescent cells in crypts in mouse intestines at various times after activation of expression of fluorescent proteins (Problem 22–6). The images are taken in the xz plane, which cuts through multiple crypts, as indicated in the schematic drawing. Roughly 50 crypts are visible in each section. Dotted white circles identify some individual crypts. Scale bars are 100 μm . (Adapted from H.J. Snippert et al., *Cell* 143:134–144, 2010. With permission from Elsevier.)

cells that express HPAP, the investigators followed the percentage of labeled β cells for a year, during which time the total number of β cells in the pancreas increased by 6.5-fold. How do you suppose the percentage of β cells would change over time if new β cells were derived from stem cells? What if new β cells were derived from preexisting β cells? Which hypothesis do the results in **Figure Q22-2** support?

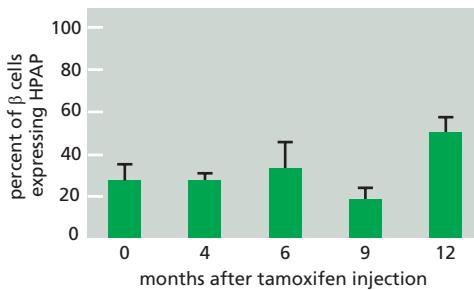


Figure Q22-2 Percentage of labeled β cells in pancreatic islets of mice at different ages (Problem 22–7). All mice were injected with a pulse of tamoxifen at 6 to 8 weeks of age and then stained for human placental alkaline phosphatase (HPAP) at various times afterward. Error bars represent standard deviations.

22–8 One of the earliest assays for hematopoietic stem cells made use of their ability to form colonies in the spleens of heavily irradiated mice. By varying the amounts of transplanted bone marrow cells, investigators showed that the number of spleen colonies varied linearly with dose and that the curve passed through the origin, suggesting that single cells were capable of forming individual colonies. However, because colony formation was rare relative to the numbers of transplanted cells, it was possible that undispersed clumps of two or more cells were the actual initiators.

A classic paper resolved this issue by exploiting rare, cytologically visible genome rearrangements generated by irradiation. Recipient mice were first irradiated to deplete bone marrow cells, and then they were irradiated a second time after transplantation to generate rare genome rearrangements in the transplanted cell population. Spleen colonies were then screened to find ones that carried genome rearrangements. How do you suppose this experiment distinguishes between colonization by single cells versus cellular aggregates?

22–9 It is possible to purify hematopoietic stem cells using a combination of antibodies directed against cell-surface targets. By removing cells that expressed surface markers characteristic of specific lineages such as B cells, granulocytes, myelomonocytic cells, and T cells, investigators generated a population of cells enriched for stem cells. They further enriched this population for putative stem cells by positively selecting for cells that expressed suspected stem-cell surface markers. Spleen colony formation in irradiated mice by these putative stem cells and the unfractionated bone marrow cells is shown in **Figure Q22-3**. Given that only about 1 in 10 cells lodges in the spleen, do these results support the idea that the enriched population consists mostly of hematopoietic stem cells? What additional information would you need to have to feel confident that the enriched cells are true stem cells? What proportion of bone marrow cells are hematopoietic stem cells?

22–10 Generation of induced pluripotent stem (iPS) cells was first accomplished using retroviral vectors to carry the OSKM (Oct4, Sox2, Klf4, and Myc) set of transcription regulators into cells. The efficiency of fibroblast reprogramming was typically low (0.01%), in part because large numbers of retroviruses must integrate to bring about reprogramming and each integration event carries with it the risk of inappropriately disrupting or activating a critical gene. In what other ways, or other forms, do you suppose you might deliver the OSKM transcription regulators so as to avoid these problems?

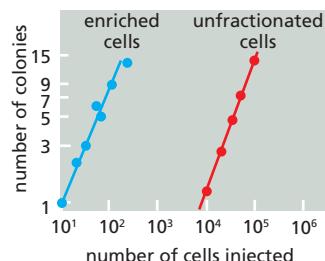


Figure Q22-3 Spleen colony formation by cells enriched for stem cells and by unfractionated bone marrow cells (Problem 22–9).

REFERENCES

General

- Fawcett DW & Jersh R (2002) Bloom and Fawcett's Concise Histology, 2nd ed. New York/London: Arnold.
- Gurdon JB & Melton DA (2008) Nuclear reprogramming in cells. *Science* 322, 1811–1815.
- Li L & Xie T (2005) Stem cell niche: structure and function. *Annu. Rev. Cell Dev. Biol.* 21, 605–631.
- Losick VP, Morris LX, Fox DT & Spradling A (2011) *Drosophila* stem cell niches: a decade of discovery suggests a unified view of stem cell regulation. *Dev. Cell* 21, 159–171.
- Young B, Woodford P & O'Dowd G (2014) Wheater's Functional Histology: A Text and Colour Atlas, 6th ed. Edinburgh: Churchill Livingstone/Elsevier.

Stem Cells and Renewal in Epithelial Tissues

- Barker N, van Es JH, Kuipers J et al. (2007) Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* 449, 1003–1007.
- Bianpaint C & Fuchs E (2014) Plasticity of epithelial stem cells in tissue regeneration. *Science* 344, 1242281.
- Crosnier C, Stamataki D & Lewis J (2006) Organizing cell renewal in the intestine: stem cells, signals and combinatorial control. *Nat. Rev. Genet.* 7, 349–359.
- Sato T, van Es JH, Snippert HJ et al. (2011) Paneth cells constitute the niche for *Lgr5* stem cells in intestinal crypts. *Nature* 469, 415–418.
- Sato T & Clevers H (2013) Growing self-organizing mini-guts from a single intestinal stem cell: mechanism and applications. *Science* 340, 1190–1194.
- 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.
- Taub R (2004) Liver regeneration: from myth to mechanism. *Nat. Rev. Mol. Cell Biol.* 5, 836–847.
- Watt FM & Huck WTS (2013) Role of the extracellular matrix in regulating stem cell fate. *Nat. Rev. Mol. Cell Biol.* 14, 467–473.

Fibroblasts and Their Transformations: The Connective-Tissue Cell Family

- Cooper KL, Oh S, Sung Y et al. (2013) Multiple phases of chondrocyte enlargement underlie differences in skeletal proportions. *Nature* 495, 375–378.
- Karsenty G & Wagner EF (2002) Reaching a genetic and molecular understanding of skeletal development. *Dev. Cell* 2, 389–406.
- Rinn JL, Bondre C, Gladstone HB et al. (2006) Anatomic demarcation by positional variation in fibroblast gene expression programs. *PLoS Genet.* 2, e119.
- 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.

Genesis and Regeneration of Skeletal Muscle

- 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.

Blood Vessels, Lymphatics, and Endothelial Cells

- 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.
- Gerhardt H, Golding M, Fruttiger M et al. (2003) VEGF guides angiogenic sprouting utilizing endothelial tip cell filopodia. *J. Cell Biol.* 161, 1163–1177.
- Lawson ND & Weinstein BM (2002) *In vivo* imaging of embryonic vascular development using transgenic zebrafish. *Dev. Biol.* 248, 307–318.
- Pugh CW & Ratcliffe PJ (2003) Regulation of angiogenesis by hypoxia: role of the HIF system. *Nat. Med.* 9, 677–684.
- Tammela T & Alitalo K (2010) Lymphangiogenesis: molecular mechanisms and future promise. *Cell* 140, 460–476.

A Hierarchical Stem-Cell System: Blood Cell Formation

- Orkin SH & Zon LI (2008) Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* 132, 631–644.
- 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.

Regeneration and Repair

- Brockes JP & Kumar A (2008) Comparative aspects of animal regeneration. *Annu. Rev. Cell Dev. Biol.* 24, 525–549.
- Tanaka EM & Reddien PW (2011) The cellular basis for animal regeneration. *Dev. Cell* 21, 172–185.
- Wagner DE, Wang IE & Reddien PW (2011) Clonogenic neoblasts are pluripotent adult stem cells that underlie planarian regeneration. *Science* 332, 811–816.

Cell Reprogramming and Pluripotent Stem Cells

- Apostolou E & Hochedlinger K (2013) Chromatin dynamics during cellular reprogramming. *Nature* 502, 462–471.
- Egawa N, Kitaoka S, Tsukita K et al. (2012) Drug screening for ALS using patient-specific induced pluripotent stem cells. *Sci. Transl. Med.* 4, 145ra104.
- Eggan K, Baldwin K, Tackett M et al. (2004) Mice cloned from olfactory sensory neurons. *Nature* 428, 44–49.
- Fox IJ, Daley GQ, Goldman SA et al. (2014) Use of differentiated pluripotent stem cells as replacement therapy for treating disease. *Science* 345, 1247391.
- Inoue H, Nagata N, Kurokawa H & Yamanaka S (2014) iPS cells: a game changer for future medicine. *EMBO J.* 33, 409–417.
- Kim J, Chu J, Shen X et al. (2008) An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* 132, 1049–1061.
- Orkin SH & Hochedlinger K (2011) Chromatin connections to pluripotency and cellular reprogramming. *Cell* 145, 835–850.
- Polo JM, Anderssen E, Walsh RM et al. (2012) A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* 151, 1617–1632.
- Radziszewska A & Silver JCR (2014) Do all roads lead to Oct4? The emerging concepts of induced pluripotency. *Trends Cell Biol.* 24, 275–284.
- Sasaki Y, Eiraku M & Suga H (2012) *In vitro* organogenesis in three dimensions: self-organising stem cells. *Development* 139, 4111–4121.
- Soza-Ried J & Fisher AG (2012) Reprogramming somatic cells towards pluripotency by cellular fusion. *Curr. Opin. Genet. Dev.* 22, 459–465.
- Takahashi K & Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663–676.
- Theunissen TW & Jaenisch R (2014) Molecular control of induced pluripotency. *Cell Stem Cell* 14, 720–734.
- Watanabe A, Yamada Y & Yamanaka S (2013) Epigenetic regulation in pluripotent stem cells: a key to breaking the epigenetic barrier. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368, 20120292.
- Yamanaka S (2013) The winding road to pluripotency (Nobel Lecture). *Angew. Chem. Int. Ed. Engl.* 52, 13900–13909.

Pathogens and Infection

CHAPTER
23

Infectious diseases currently cause about one-quarter of all human deaths worldwide, more than all forms of cancer combined and second only to cardiovascular diseases. In addition to the continuing heavy burden of ancient diseases such as tuberculosis and malaria, newer infectious diseases continually emerge. The current pandemic (worldwide epidemic) of AIDS (*acquired immune deficiency syndrome*), was first clinically observed in 1981 and has since caused more than 35 million deaths worldwide. Moreover, some diseases long thought to result from other causes are now recognized to be associated with infections. Most gastric ulcers, for example, are caused not by stress or spicy food, but by infection of the stomach lining by the bacterium *Helicobacter pylori*.

The burden of infectious diseases is not spread equally across the planet. Poorer countries and communities suffer disproportionately, often due to poor public sanitation and health systems. Some infectious diseases, however, occur primarily or exclusively in industrialized communities: Legionnaire's disease, for example, a bacterial infection of the lungs, commonly spreads through air-conditioning systems.

Since the mid-1800s, physicians and scientists have struggled to identify the agents—collectively called **pathogens**—that are capable of causing infectious diseases. 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 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.

Although pathogens are understandably a focus of attention, only a relatively small fraction of the microbial species we encounter are pathogens. Much of the biomass of the Earth is made up of microbes, and they produce everything from the oxygen we breathe to the soil nutrients we use to grow food. Even those species of microbes that colonize the human body do not generally cause disease. The collective of microorganisms that reside in or on an organism is called the **microbiota**. Many of these microbes have a beneficial effect on the health of the organism, assisting its normal development and physiology.

In this chapter, we give an overview of the different kinds of pathogens, as well as those microorganisms that colonize our body without causing trouble. We then discuss the cell biology of infection—the molecular interactions between pathogens and their host. In Chapter 24, we consider how our innate and adaptive immune systems collaborate to defend us against pathogens.

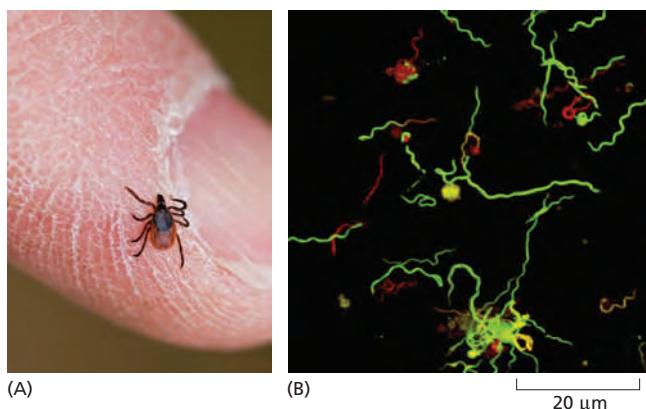
INTRODUCTION TO PATHOGENS AND THE HUMAN MICROBIOTA

We normally think of pathogens as hostile invaders, but a pathogen, like any other organism, is simply exploiting an available niche in which to live and procreate. Living on or in a host organism is a very effective strategy, and it is possible that every organism on Earth is subject to some type of infection (Figure 23-1). A human host is a nutrient-rich, warm, and moist environment, which remains at a

IN THIS CHAPTER

INTRODUCTION TO PATHOGENS
AND THE HUMAN MICROBIOTA

CELL BIOLOGY OF INFECTION

**Figure 23–1** Parasitism at many levels.

(A) Most animals harbor parasites, an example being the blacklegged tick or deer tick (*Ixodes scapularis*), shown here on a human finger. Although ticks of this species thrive on white-tailed deer and other wild mammals, they can also live on humans. (B) Ticks themselves harbor their own parasites including the bacterium *Borrelia burgdorferi*, stained here with a vital dye that labels living bacteria green and dead bacteria red. These spiral-shaped bacteria live in deer ticks and can be transmitted to humans during a tick's blood meal. *Borrelia burgdorferi* causes Lyme disease, which is characterized by a bull's-eye-shaped skin rash and fever; if the infection is left untreated, various complications can result, including arthritis and neurological abnormalities. The idea that parasites have their own parasites was noted by Jonathan Swift 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, from Acorn, White-Footed Mice and Tick Cycle Augment Risks of Lyme Disease in 2012. March 14, 2012. Reprinted with permission of Anita Sil; B, courtesy of M. Embers.)

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 colonize the human body or cause disease, and we explore the wide variety of organisms that are known to cause disease.

The Human Microbiota Is a Complex Ecological System That Is Important for Our Development and Health

The human body contains about 10^{13} human cells, as well as a microbiota consisting of approximately 10^{14} bacterial, fungal, and protozoan cells, which represent thousands of microbial species—the so-called **normal flora**. The combined genomes of the various species of the human microbiota, called the **microbiome**, contain more than 5×10^6 genes—more than 100 times greater than the number of genes in the human genome itself. A consequence of this genomic diversity is that the microbiota expands the range of biochemical and metabolic activities available to the humans.

The microbiota is usually confined to the skin, mouth, digestive tract, and vagina. With the exception of microbes colonizing the skin, it consists primarily of anaerobic bacteria, with distinct communities of species inhabiting each body part. These communities vary considerably between individual humans, even between close relatives or identical twins. Although the microbiota of an individual is generally consistent over time, it is influenced by a variety of factors, including age, diet, health status, and antibiotic use.

There are various ecological relationships that these microbes have with their host. In **mutualism**, both the microbe and host benefit. The anaerobic bacteria that inhabit our intestines, for example, gain shelter and a nutrient supply but also contribute to the digestion of our food, produce important nutrients for us, and are essential for the normal development of our gastrointestinal tract and innate and adaptive immune systems. In **commensalism**, the microbe benefits but offers no benefit and causes no harm: for example, we are infected with many viruses that have no noticeable effect on our health. In **parasitism**, the microbe benefits to the detriment of the host, as is often the case for pathogens.

Many infectious diseases are caused by a single pathogen. There is increasing evidence, however, that an imbalance in the community of microbes that constitute the microbiota can contribute to some diseases, including autoimmune and allergic diseases, obesity, inflammatory bowel disease, and diabetes. Remarkably, in such cases of microbiota imbalance (referred to as *dysbiosis*), the transfer of the microbiota from a healthy individual to someone suffering from the disease can be beneficial and sometimes curative, as in the case of *Clostridium difficile* colitis caused by overgrowth of the bacterium.

Pathogens Interact with Their Hosts in Different Ways

If it is normal for us to live with a community of microbes, why are some of them capable of causing us illness or death? Although the ability of a particular

microorganism to cause disease depends on many factors, it requires that the pathogen possess specialized pathogenic characteristics that allow it to live in humans.

Primary pathogens can cause overt disease in most healthy people. Some primary pathogens cause acute, life-threatening epidemic infections and spread rapidly from one sick or dying host to another; historically important examples include the bacterium *Vibrio cholerae*, which causes cholera, and the variola and influenza viruses, which cause smallpox and flu, respectively. Others may persistently infect a single individual for years without causing overt disease; examples include the bacterium *Mycobacterium tuberculosis* (which can cause the life-threatening lung infection tuberculosis) and the intestinal worm *Ascaris*. Although these potential primary pathogens can make some people critically ill, billions of people carry these foreign organisms in an asymptomatic way, often unaware that they are infected. It is sometimes difficult to draw a line between the asymptomatic presence of such pathogens and the normal microbiota. Some microbes of the normal flora can act as **opportunistic pathogens**, in that they cause disease only if our immune systems are weakened or if they gain access to a normally sterile part of the body.

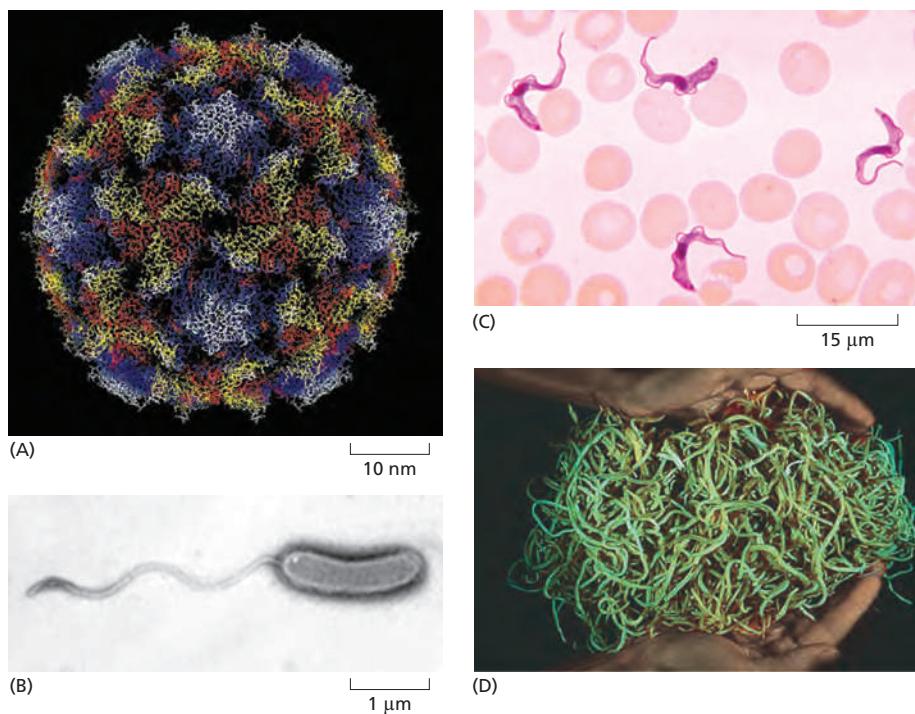
In order to survive and multiply, a successful pathogen must be able to: (1) enter the host (usually by breaking an epithelial barrier); (2) find a nutritionally compatible niche in the host's body; (3) avoid, subvert, or circumvent the host's innate and adaptive immune responses; (4) replicate, using host resources; and (5) exit one host and spread to another. Pathogens have evolved various mechanisms that maximally exploit the biology of their host organisms to help accomplish these tasks. For some pathogens, these mechanisms are adapted to a unique host species, whereas for others the mechanisms are sufficiently general to permit invasion, survival, and replication in a wide variety of hosts. Because pathogens have evolved the ability to interface directly with the molecular machinery of host cells, we have learned a great deal about cell biological principles by studying them.

Our constant exposure to pathogens has strongly influenced human evolution. In modern times, humans have learned how to limit the ability of pathogens to infect us through improvements in public health measures and childhood nutrition, vaccines, antimicrobial drugs, and routine testing of blood used for transfusions. As we learn more about the mechanisms by which pathogens cause disease (called *pathogenesis*), our creativity and resourcefullness will continue to serve as an important addition to our immune systems in fighting infectious diseases.

Pathogens Can Contribute to Cancer, Cardiovascular Disease, and Other Chronic Illnesses

Some viral and bacterial pathogens can cause or contribute to chronic, life-threatening illnesses that are not normally classified as infectious diseases. An important example is cancer. As discussed in Chapter 20, the oncogene concept—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 viral genes encodes an overactive homolog of the host tyrosine kinase Src (see Figure 3–63), which has been implicated in many kinds of cancer. Several human cancers are also known to have a viral origin. *Human papillomavirus*, for example, which causes genital warts, is responsible for more than 90% of cervical cancers (see Figure 20–40). The recent development of a vaccine against the most abundant cancer-associated strains of human papillomavirus promises to prevent many of these cancers in the future. In other cases, chronic tissue damage caused by infection can increase the likelihood of cancer. Inflammation caused by the stomach-dwelling bacterium *H. pylori* can be a major contributor to stomach cancer, as well as to gastric ulcers.

The major causes of death in wealthy industrialized nations are cardiovascular diseases. They frequently result from *atherosclerosis*, the accumulation in blood

**Figure 23–2** 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 greatly reduced by widespread vaccination. (B) The bacterium *Vibrio cholerae*, the causative agent of the epidemic, diarrheal disease cholera. (C) The protozoan parasite *Trypanosoma brucei* (purple) in a field of erythrocytes (red blood cells; pink). This parasite causes African sleeping sickness, a potentially fatal disease of the central nervous system. (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, photograph courtesy of John Mekalanos; C, CDC, Department of Health and Human Services; D, from J.K. Baird et al., *Am. J. Trop. Med. Hyg.* 35:314–318, 1986. Photograph by Daniel H. Connor.)

vessel walls of fatty deposits that can block blood flow and cause heart attacks and strokes. A hallmark of early atherosclerosis is the appearance in blood vessel walls of clumps of macrophages called foam cells, which recruit other white blood cells into the forming *atherosclerotic plaque*. Foam cells in atherosclerotic plaques often contain the bacterial pathogen *Chlamydia pneumoniae*, which commonly causes pneumonia in humans and is a significant risk factor for atherosclerosis in humans and animal models. Other bacterial species are also implicated in atherosclerosis, including bacteria usually associated with teeth and gums, such as *Porphyromonas gingivalis*. As we learn more about the interactions between pathogens and the human body, it seems likely that more chronic conditions will be found to have a link to an infectious agent.

Pathogens Can Be Viruses, Bacteria, or Eukaryotes

Many types of pathogens cause disease in humans. The most familiar are viruses and bacteria. Viruses cause diseases ranging from AIDS and smallpox to the common cold. Viruses are essentially fragments of nucleic acid (DNA or RNA) that generally encode a relatively small number of gene products, wrapped in a protective shell of proteins (Figure 23–2A) and (in some cases) an outer membrane envelope (see Figure 5–62). Much larger and more complex than viruses, bacteria are prokaryotic cells, which perform most of their basic metabolic functions themselves, relying on the host primarily for nutrition (Figure 23–2B).

Some other infectious agents are eukaryotic organisms. These range from single-celled fungi and protozoa (Figure 23–2C) to large, complex metazoa such as parasitic worms. One of the most common human parasites, shared by about a billion people at present, is the nematode worm *Ascaris lumbricoides*, which infects the gut (Figure 23–2D). It closely resembles its harmless nematode cousin *Caenorhabditis elegans*, which is used as a model organism for genetic and developmental biological research (see Figure 1–39). *C. elegans*, however, is only about 1 mm in length, whereas *Ascaris* can reach 30 cm.

We now introduce the basic features of each of the major types of pathogens, before we examine the mechanisms that pathogens use to infect their hosts.

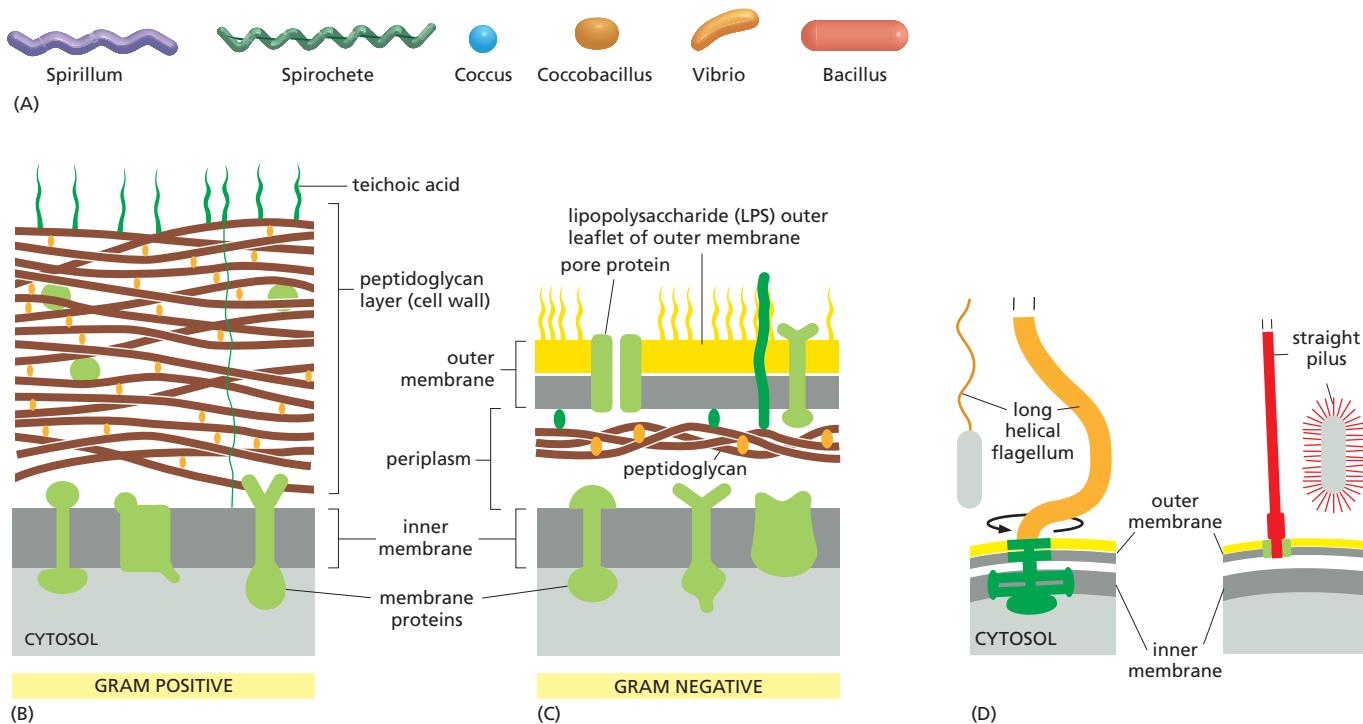


Figure 23–3 Bacterial shapes and cell-surface structures. (A) Bacteria are traditionally classified by shape. (B and C) They are also classified as **Gram positive** or **Gram negative**. (B) Gram-positive bacteria such as *Streptococcus* and *Staphylococcus* have a single membrane and a thick cell wall made of cross-linked **peptidoglycan**. They are called Gram positive because they retain the violet dye used in the Gram-staining procedure. (C) Gram-negative bacteria such as *Escherichia coli* (*E. coli*) and *Salmonella* have two membranes, separated by the **periplasm** (see Figure 11–17). The peptidoglycan cell wall of these organisms is located in the periplasm and is thinner than in Gram-positive bacteria; they therefore fail to retain the dye in the Gram-staining procedure. The inner membrane of both Gram-positive and Gram-negative bacteria is a phospholipid bilayer. The inner leaflet of the outer membrane of Gram-negative bacteria is also made primarily of phospholipids, whereas the outer leaflet of the outer membrane is composed of a unique glycosylated lipid called **lipopolysaccharide (LPS)**. (D) Cell-surface appendages are important for bacterial behavior. Many bacteria swim using the rotation of helical flagella. The bacterium illustrated has only a single flagellum at one pole; however, many 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 along surfaces.

Bacteria Are Diverse and Occupy a Remarkable Variety of Ecological Niches

Although **bacteria** generally lack internal membranes, they are highly sophisticated cells whose organization and behaviors have attracted the attention of many scientists. Bacteria are classified broadly by their shape—as rods, spheres (*cocci*), or spirals (Figure 23–3A)—as well as by their so-called **Gram-staining** properties, which reflect differences in the structure of the bacterial cell wall. **Gram-positive** bacteria have a thick layer of peptidoglycan cell wall outside their inner (plasma) membrane (Figure 23–3B), whereas **Gram-negative** bacteria have a thinner peptidoglycan cell wall. In both cases, the cell wall protects against lysis by osmotic swelling, and it is a target of host antibacterial proteins such as lysozyme and antibiotics such as penicillin. Gram-negative bacteria are also covered outside the cell wall by an outer membrane containing **lipopolysaccharide (LPS)** (Figure 23–3C). Both peptidoglycan and LPS are unique to bacteria and are recognized as *pathogen-associated molecular patterns (PAMPs)* by the host innate immune system, as discussed in Chapter 24. The surface of bacterial cells can also display an array of appendages, including flagella and pili, which enable bacteria to swim or adhere to desirable surfaces, respectively (Figure 23–3D). Apart from cell shape and structure, differences in ribosomal RNA and genomic DNA sequence are also used for phylogenetic classification. Because bacterial genomes are small—typically between 1,000,000 and 5,000,000 nucleotide pairs (compared to more than

3,000,000,000 for humans)—they are now simple to sequence, making this an important new classification tool.

Bacteria also exhibit extraordinary molecular, metabolic, and ecological diversity. At the molecular level, bacteria are far more diverse than eukaryotes, and they can occupy ecological niches having extremes of temperature, salt concentrations, and nutrient limitation. Some bacteria 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**. Others can only replicate inside the body of their host and are therefore called **obligate pathogens**. Bacteria also differ in the range of hosts they will infect. *Shigella flexneri*, for example, which causes epidemic dysentery (bloody diarrhea), 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 other vertebrates, including chickens and turtles. A champion generalist is the opportunistic pathogen *Pseudomonas aeruginosa*, which can cause disease in a wide variety of plants and animals.

Bacterial Pathogens Carry Specialized Virulence Genes

Pathogenic bacteria and their closest nonpathogenic relatives often differ in a relatively small number of genes. 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**. Such virulence genes are often clustered together on the bacterial chromosome; large clusters are called *pathogenicity islands*. Virulence genes can also be carried on *bacteriophages* (bacterial viruses) or *transposons* (see Table 5–4), both of which integrate into the bacterial chromosome, or on extrachromosomal *virulence plasmids* (Figure 23–4A).

Pathogenic bacteria are thought to emerge when groups of virulence genes are transferred together into a previously avirulent bacterium by a process called **horizontal gene transfer** (to distinguish it from vertical gene transfer from parent to offspring). Horizontal transfer can occur by one of three mechanisms: natural *transformation* by released naked DNA, *transduction* by bacteriophages, or sexual exchange by *conjugation* (Figure 23–4B and Movie 23.1). Sequencing the genomes of large numbers of pathogenic and nonpathogenic bacteria has indicated that horizontal gene transfer has made important contributions to bacterial evolution, enabling species to inhabit new ecological and nutritional niches, as well as to cause disease. Even within a single bacterial species, the amount of chromosomal

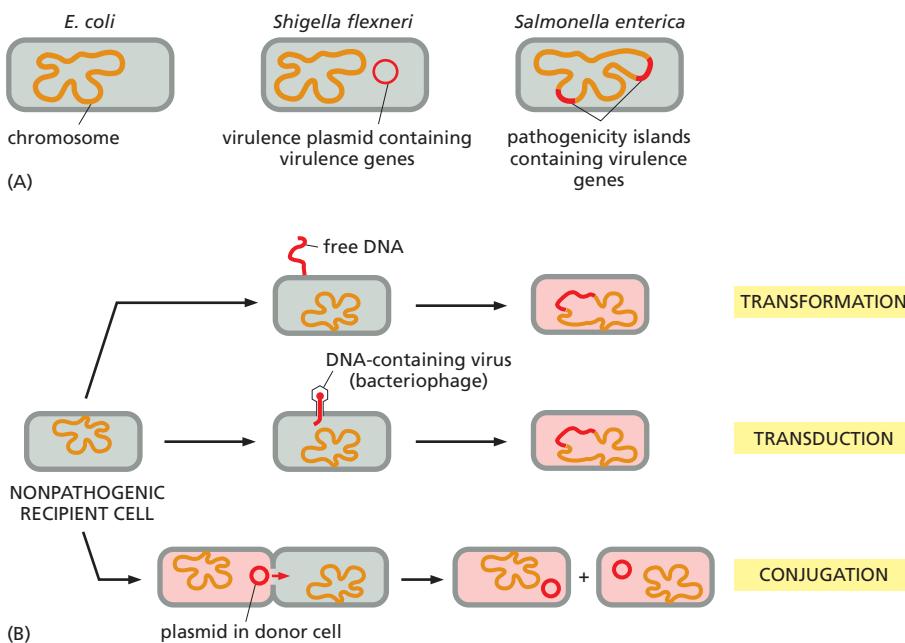


Figure 23–4 Genetic differences between pathogenic and nonpathogenic bacteria. (A) Genetic differences between nonpathogenic *E. coli* and two closely related food-borne pathogens—*Shigella flexneri*, which causes dysentery, and *Salmonella enterica*, a common cause of food poisoning. Nonpathogenic *E. coli* has a single circular chromosome. The chromosome of *S. flexneri* differs from that of *E. coli* in a limited number of locations; 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. (B) Bacterial pathogens evolve by horizontal gene transfer. This can occur by three mechanisms: natural transformation, in which naked DNA is taken in by competent bacteria; transduction, in which bacterial viruses (*bacteriophages*) transfer DNA from one bacterium into another; and conjugation, during which plasmid DNA, and even chromosomal DNA, is transferred from a donor to a recipient bacterium.

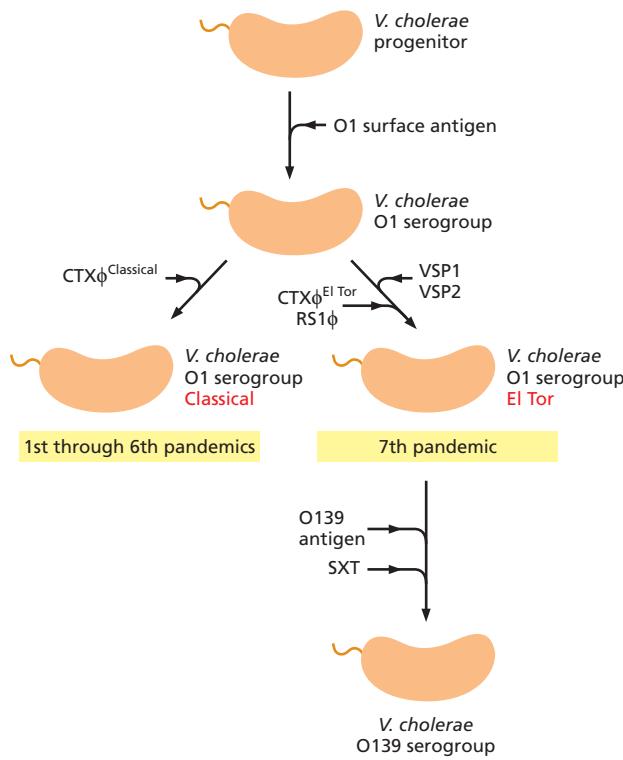


Figure 23–5 Comparative-genomics-based model for the evolution of pathogenic *Vibrio cholerae* strains.

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 Figure 23–3C). Incorporation of the $\text{CTX}\phi$ 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 $\text{CTX}\phi$ bacteriophage again, along with an associated bacteriophage RS1 ϕ and two pathogenicity islands (VSP1 and VSP2), creating the El Tor strain that emerged as the seventh worldwide pandemic in 1961. In 1992, an El Tor strain was isolated that had picked up a new DNA cassette, enabling it to produce the O139 antigen type of carbohydrate chain rather than the O1 type. This altered the bacterium's interaction with the human immune system, without diminishing its virulence; this bacterium also picked up a new pathogenicity island (SXT). An electron micrograph of *Vibrio cholerae* (*V. cholerae*) is shown in Figure 23–2B.

variation is astonishing; the genomes of different strains of *Escherichia coli* can differ by as much as 25%. Such variation has led to the concept that a bacterial species has both a *core genome* common to all isolates within the species and a larger *pan-genome* consisting of all genes present in the full spectrum of isolates.

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. Of the hundreds of strains of *Vibrio cholerae*, the only ones that cause pandemic human disease are those infected with a mobile bacteriophage (CTX ϕ) containing genes encoding the two subunits of the toxin that causes the diarrhea. As summarized in Figure 23–5, seven pandemics of *V. cholerae* have arisen since 1817. The first six were caused by the periodic reemergence of so-called Classical strains. In addition to the toxin-encoding bacteriophage, these Classical strains shared a similar O1 surface antigen, part of the LPS in the outer membrane (see Figure 23–3C). In 1961, the seventh pandemic began, caused by a new strain named “El Tor,” which arose when an O1-expressing strain acquired two bacteriophages and at least two new pathogenicity islands. El Tor eventually displaced the Classical strains. In 1992, a new strain emerged in which O1 was replaced with another O-antigen variant called O139, which was not recognized by antibodies present in the blood of survivors of previous cholera epidemics. The O139 strain also contains a transposon-like element that encodes antibiotic resistance. As this example makes clear, the rapid evolution of bacterial pathogens can be likened to an arms race which pits the survival of a bacterium against our immune systems and the tools of modern medicine. Similar struggles for survival take place between all pathogens and humans, and understanding these conflicts provides key insights into the evolution of pathogens and greatly informs us how we treat new outbreaks of infectious diseases.

Bacterial Virulence Genes Encode Effector Proteins and Secretion Systems to Deliver Effector Proteins to Host Cells

What are the gene products that enable a bacterium to cause disease in a healthy host? For pathogenic bacteria that live outside of host cells, called *extracellular*

bacterial pathogens, virulence genes often encode secreted toxic proteins (*toxins*) that interact with host cell structural or signaling proteins to elicit a response that is beneficial to the pathogen. Several of these bacterial toxins are among the most potent of known human poisons. Bacterial toxins are often composed of two protein components—an A subunit with enzymatic activity, and a B subunit that binds to specific receptors on the host cell surface and directs the trafficking of the A subunit to the cytosol by various routes (Figure 23–6). The *Vibrio cholerae* phage, for example, encodes the two subunits of **cholera toxin** (Movie 23.2). The A subunit catalyzes the transfer of an ADP-ribose moiety from NAD⁺ to the trimeric G protein G_s (see Figure 15–23), which activates adenylyl cyclase to make cyclic AMP (see Figure 15–25). ADP-ribosylation prevents inactivation of the G protein and results in the overaccumulation of intracellular cyclic AMP and the release of ions and water into the intestinal lumen, leading to the watery diarrhea associated with cholera. The infection then spreads to new hosts via released bacteria, which can contaminate food and water.

Some pathogenic bacteria secrete multiple toxins, each of which targets a different signaling pathway in host cells. Anthrax, for example, is an acute infectious disease of sheep, cattle, and occasionally humans. It is caused by contact with spores of the Gram-positive bacterium *Bacillus anthracis*. Dormant spores can survive in soil for long periods. If inhaled, ingested, or rubbed into breaks in the skin, spores can germinate and the bacteria replicate. The bacteria secrete two toxins with identical B subunits but different A subunits. The B subunits bind to a host cell-surface receptor protein to transfer the two different A subunits into

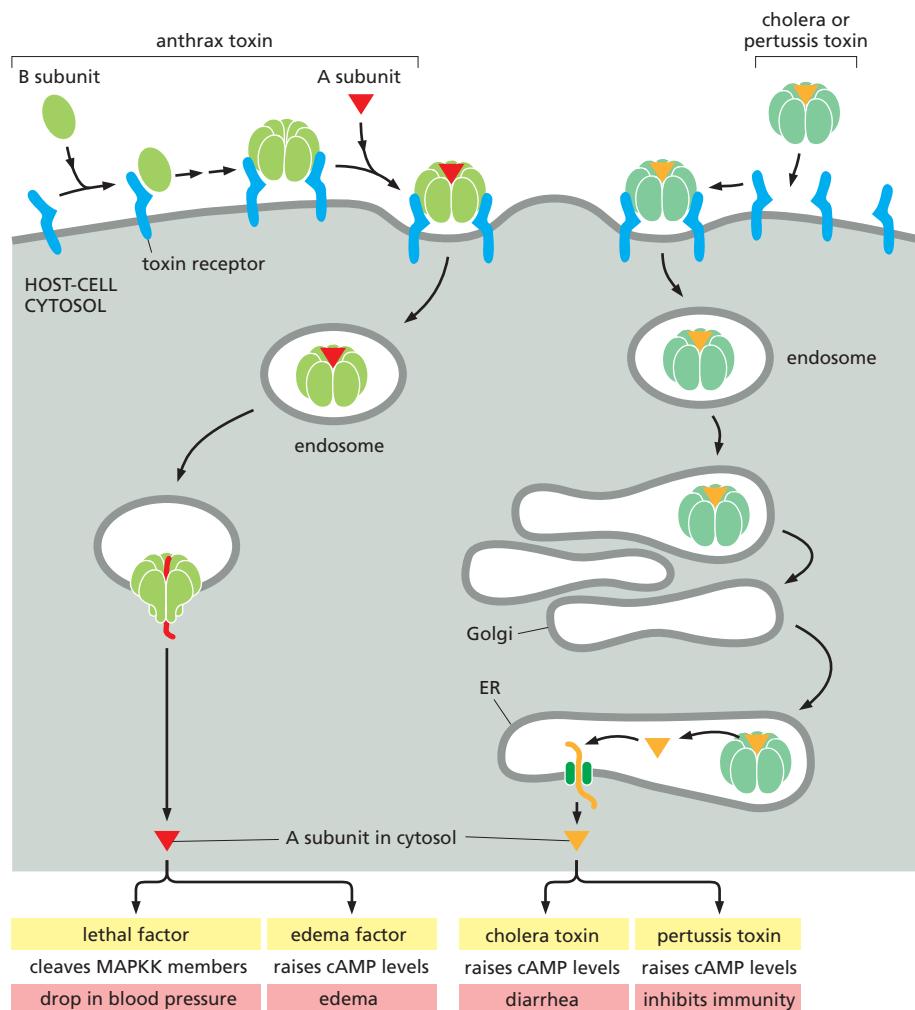


Figure 23–6 Bacterial toxin entry into host cells. Bacterial toxins are often composed of A and B protein subunits. The B (binding) subunit of the toxin interacts with host-cell toxin receptors, enabling endocytosis and intracellular trafficking of B subunit as well as its associated and enzymatically active A subunit(s). In the case of *Bacillus anthracis*, the B subunit changes conformation in the low pH environment of the endosome to form a pore through which two different A subunits, lethal factor and edema factor, are transported across the membrane of the endosome in an unfolded conformation. In the cases of *Vibrio cholerae* toxin and *Bordetella pertussis* toxin, the B and A subunits are transported to the Golgi apparatus and then to the endoplasmic reticulum (ER), where the A subunits are then translocated into the cytosol in an unfolded conformation through a protein translocation channel.

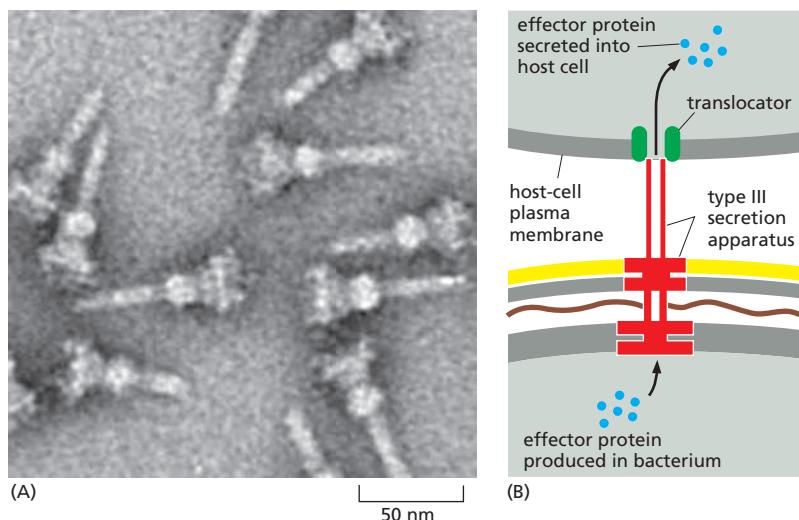


Figure 23–7 Type III secretion systems that can deliver effector proteins into the cytosol of a host cell. (A) Electron micrograph of purified type III secretion systems, each of which consists of over two dozen proteins. (B) The large lower ring is embedded in the bacterial inner membrane, and the smaller upper ring is embedded in the bacterial outer membrane. During infection, docking of the tip of the hollow needle at a host-cell plasma membrane results in the secretion of bacterial translocator proteins (green), which form a pore in the host membrane, through which bacterial effector proteins are then secreted into the host cell. (A, from O. Schraadt et al., *PLoS Pathog.* 6(4):e1000824, 2010.)

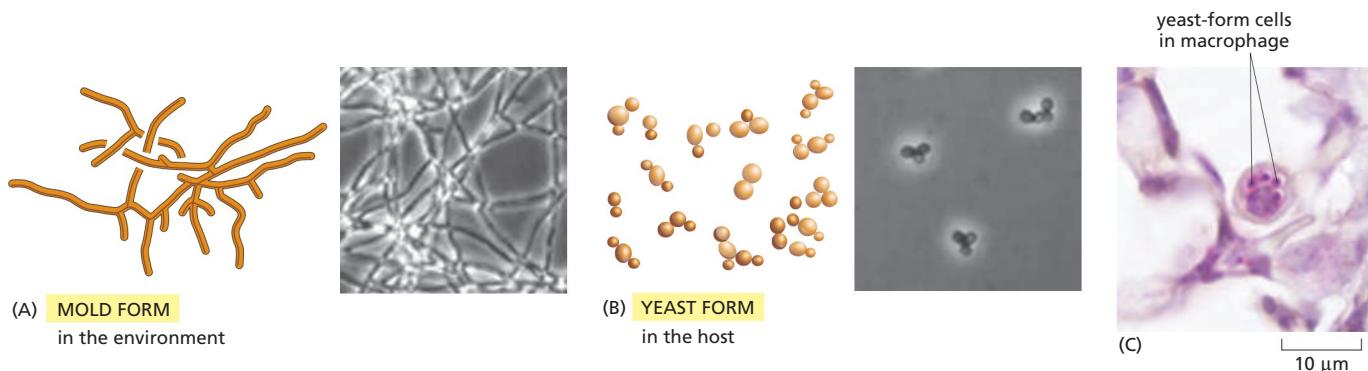
host cells (see Figure 23–6). The A subunits are called **lethal factor** and **edema factor**. The A subunit of edema toxin is an adenylyl cyclase that catalyzes the production of cyclic AMP (see Figure 15–25), leading to an ion imbalance that can cause an accumulation of extracellular fluid (*edema*) in the skin or lung. The A subunit of lethal toxin is a protease that cleaves several activated members of the mitogen-activated protein kinase kinase (MAP kinase kinase) family (see Figure 15–49), disrupting intracellular signaling and leading to immune cell dysfunction and cell death. Injection of lethal toxin into the bloodstream of an animal causes shock (a large fall in blood pressure) and death.

Apart from toxins, bacteria use specialized **secretion systems** to secrete many other *effector proteins* that interact with host cells. Gram-negative bacteria have a *general secretion system* and several classes of *accessory secretion systems* (types I–VI). A subset of these accessory secretion systems, called *contact-dependent secretion systems*, is present in many bacteria that contact or live inside host cells. The **type III secretion system** (Figure 23–7), for example, injects into the host-cell cytoplasm *effector proteins* that 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. 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. Other types of delivery systems used by bacterial pathogens appear to have evolved independently. For example, *type IV secretion systems* are closely related to the conjugation apparatus that many bacteria use to exchange genetic material.

Fungal and Protozoan Parasites Have Complex Life Cycles Involving Multiple Forms

Pathogenic fungi and protozoan parasites are eukaryotes, as are their hosts. Consequently, antifungal and antiparasitic drugs are often less effective and more toxic to the host than are antibiotics that target bacteria. A second characteristic of fungal and parasitic infections that makes them difficult to treat is the tendency of the pathogens to switch among several different forms during their life cycles. A drug that is effective at killing one form can be ineffective at killing another form; therefore the population can survive the treatment.

Fungi include both unicellular *yeasts* (such as *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, which are used to bake bread and brew beer, and as model organisms for cell biology research) 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 23–8).

Protozoan parasites are single-celled eukaryotes with more elaborate life cycles than fungi, and they frequently require more than one host. **Malaria** is the most devastating protozoal disease, infecting more than 200 million people every year and killing upward of 500,000. It is caused by four species of *Plasmodium*, which are transmitted to humans by the bite of the female *Anopheles* mosquito.

Figure 23–8 Dimorphism in the pathogenic fungus *Histoplasma capsulatum*. (A) At low temperature in the soil, *H. capsulatum* grows as a multicellular filamentous mold consisting of many individual cells connected together. (B) After it is inhaled into the lung of a mammal, the increase in temperature causes a switch to a yeast form consisting of small clumps of round cells. (C) A stained histologic section of a mouse lung infected with *H. capsulatum*, showing a macrophage containing yeast forms of the pathogen. (A and B, courtesy of Sinem Beyhan and Anita Sil; C, courtesy of Davina Hocking Murray and Anita Sil.)

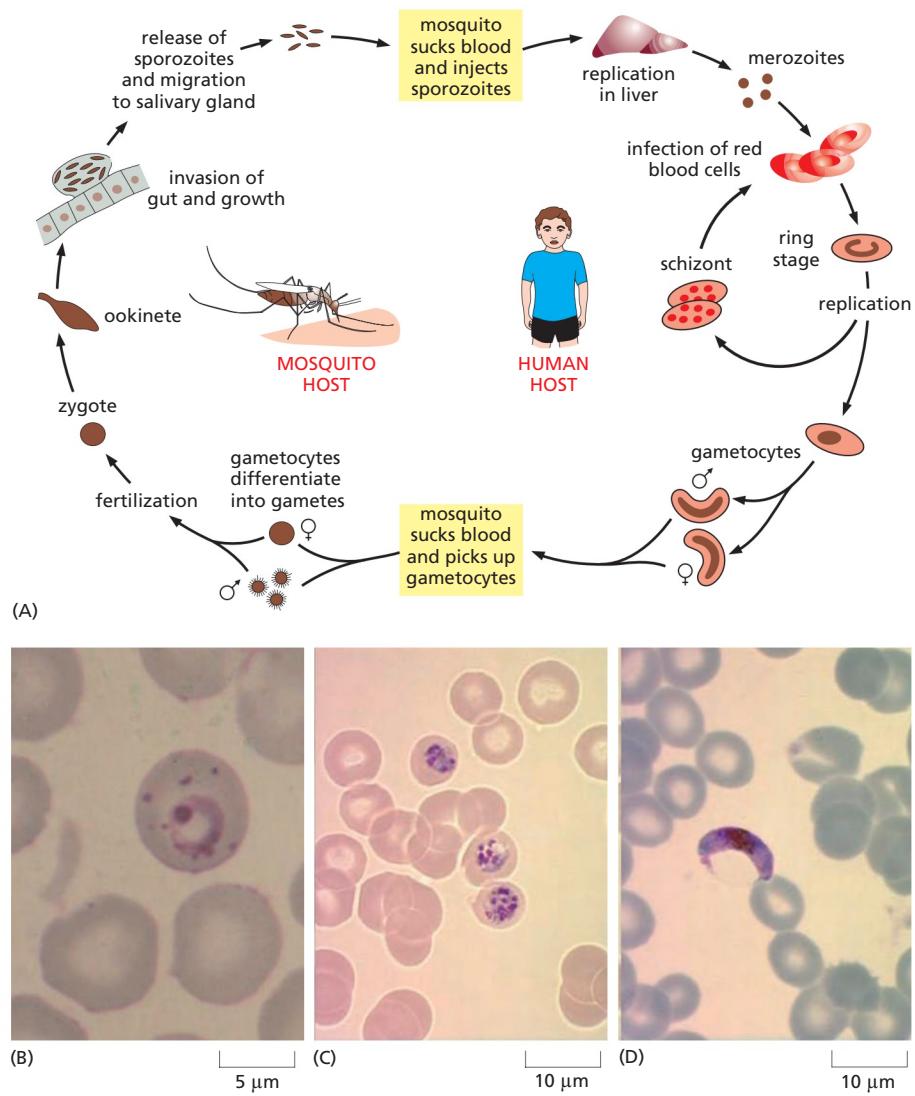


Figure 23–9 The complex life cycle of malaria parasites. (A) The sexual cycle of *Plasmodium falciparum* requires passage between a human host and an insect host (Movie 23.3). (B)–(D) Blood smears from people with malaria, showing three different forms of the parasite that appear in red blood cells: (B) ring stage; (C) schizont; and (D) gametocyte. (B–D, courtesy of the Centers for Disease Control, Division of Parasitic Diseases, DPDx.)

Plasmodium falciparum causes the most serious form of malaria and is the most intensively studied of the malaria-causing parasites. It exists in many distinct forms, and it requires both the human and mosquito hosts to complete its sexual cycle (Figure 23–9). Several of these forms are highly specialized to invade and replicate in specific tissues—the lining of the insect gut, 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 23–9B–D).

All Aspects of Viral Propagation Depend on Host Cell Machinery

Bacteria, fungal, and protozoan pathogens are living cells themselves. They use their own machinery for DNA replication, transcription, and translation, and, for the most part, 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. Most clinically important human viruses have small genomes consisting of double-stranded DNA or single-stranded RNA (Table 23–1), and we now have complete genome sequences of almost all of them.

Viral genomes typically encode three types of protein: proteins for replicating the genome, proteins for packaging the genome and delivering it to more host cells, and proteins for modifying the structure or function of the host cell to enhance the replication of the virus (see Figure 7–62). In general, viral replication involves (1) entry into the host cell, (2) disassembly of the infectious virus particle, (3) replication of the viral genome, (4) transcription of viral genes and synthesis of viral proteins, (5) assembly of these viral components into progeny virus particles,

TABLE 23–1 Viruses That Cause Human Disease

Virus	Genome type	Disease
Herpes simplex virus 1	Double-stranded DNA	Recurrent cold sores
Epstein–Barr virus (EBV)	Double-stranded DNA	Infectious mononucleosis
Varicella-zoster virus	Double-stranded DNA	Chickenpox and shingles
Smallpox virus (Variola)	Double-stranded DNA	Smallpox
Human papillomavirus	Double-stranded DNA	Warts, cancer
Adenovirus	Double-stranded DNA	Respiratory disease
Hepatitis-B virus	Part single-, part double-stranded DNA	Hepatitis B
Human immunodeficiency virus (HIV-1)	Single-stranded RNA [+] strand	Acquired immune deficiency syndrome (AIDS)
Poliovirus	Single-stranded RNA [+] strand	Poliomyelitis
Rhinovirus	Single-stranded RNA [+] strand	Common cold
Hepatitis-A virus	Single-stranded RNA [+] strand	Hepatitis A
Hepatitis-C virus	Single-stranded RNA [+] strand	Hepatitis C
Yellow fever virus	Single-stranded RNA [+] strand	Yellow fever
Coronavirus	Single-stranded RNA [+] strand	Common cold, respiratory disease
Rabies virus	Single-stranded RNA [-] strand	Rabies
Mumps virus	Single-stranded RNA [-] strand	Mumps
Measles virus	Single-stranded RNA [-] strand	Measles
Influenza virus type A	Single-stranded RNA [-] strand	Respiratory disease (flu)

Figure 23–10 A simple viral life cycle. The hypothetical simple virus shown here consists of a small double-stranded DNA molecule that codes for only a single viral capsid protein. To reproduce, the viral genome must first enter a host cell, where it is replicated to produce multiple copies, which are transcribed and translated to produce the viral coat protein. The viral genomes can then assemble spontaneously with the coat protein to form a new virus particle, which escapes from the host cell. No known virus is this simple.

and (6) release of progeny virions (Figure 23–10). A single virus particle (a *virion*) that infects a single host cell can produce thousands of progeny.

Virions come in a wide variety of shapes and sizes (Figure 23–11), and although most have relatively small genomes, genome size can vary considerably. The recently discovered giant viruses of amoebae, called *pandoraviruses*, are the largest known viruses, with 700 nm particles and double-stranded DNA genomes of over 2,000,000 nucleotide pairs. The virions of *poxvirus* are also large: they are 250–350 nm long and enclose a genome of double-stranded DNA of about 270,000 nucleotide pairs. At the other end of the size scale are the virions of *parvovirus*, which are less than 30 nm in diameter and have a single-stranded DNA genome of fewer than 5000 nucleotides.

Viral genomes are packaged in a protein coat, called a **capsid**, which in some viruses is further enclosed by a lipid bilayer membrane, or envelope. The capsid is made of one or several proteins, arranged in regular arrays that generally produce structures with either helical symmetry, which results in a cylindrical structure (for example, influenza, measles, and bunyavirus), or icosahedral symmetry (for example, poliovirus and herpesvirus; see Figure 23–11). Some viruses instead produce capsids with more complicated structures (for example, poxviruses). When the capsid is packaged with the viral genome, the structure is called a *nucleocapsid*. The nucleocapsids of *nonenveloped viruses* usually leave an infected cell by lysing it. For *enveloped viruses*, by contrast, the nucleocapsid is enclosed within a lipid bilayer membrane that the virus acquires in the process of budding from the host-cell plasma membrane, which it does without disrupting the membrane or killing the cell (Figure 23–12). Enveloped viruses can cause persistent infections that may last for years, often without noticeable deleterious effects on the host.

Because the host cell performs most of the critical steps in viral replication, the identification of effective antiviral drugs that do not harm the host can be difficult. Probably the most effective strategy for containing viral diseases is through vaccinating of potential hosts. Highly successful vaccination programs have effectively

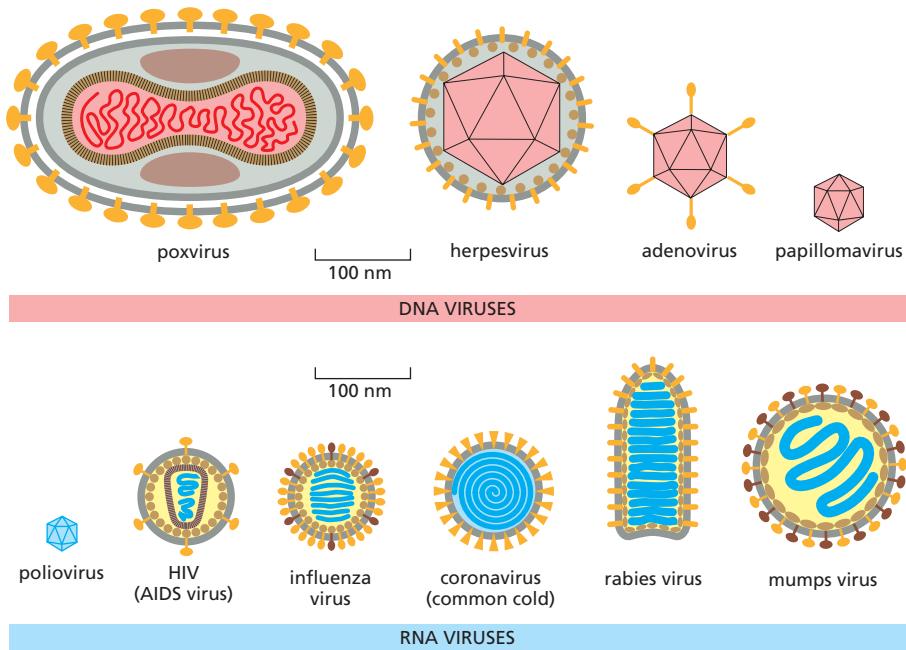
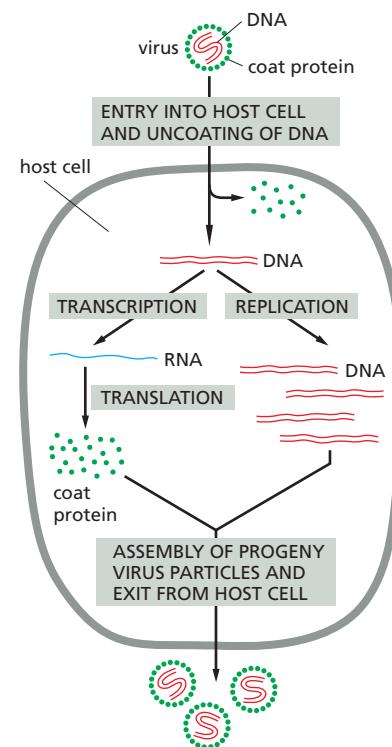


Figure 23–11 Examples of viral morphology. As shown, both DNA and RNA viruses vary greatly in both size and shape.

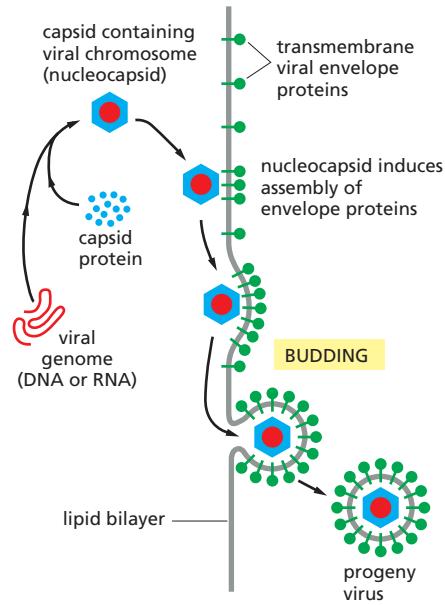
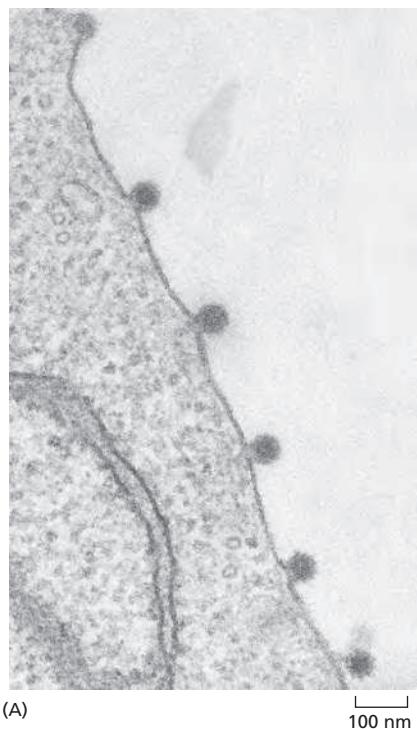


Figure 23–12 Acquisition of a viral envelope. (A) Electron micrograph of an animal cell from which six copies of an enveloped virus (*Semliki forest virus*) are budding. (B) Schematic drawing of the envelope assembly and budding processes. The lipid bilayer that surrounds the viral capsid is derived 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.)

eliminated smallpox infection from the planet, and the eradication of poliomyelitis is approaching completion (Figure 23–13).

Summary

Infectious diseases are caused by pathogens, which include viruses, bacteria, and fungi, as well as protozoan and metazoan parasites. All pathogens must have mechanisms for entering their host and for evading immediate destruction by the host. The great majority of bacteria are not pathogenic to humans. Those that are pathogenic produce specific virulence factors that mediate the bacteria's interactions with the host; these proteins change the behavior of host cells in ways that promote the replication and spread of the bacteria. Eukaryotic pathogens such as fungi and protozoan parasites typically pass through several different forms during the course of infection; the ability to switch among these forms is usually required for these pathogens 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 eukaryotic 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 on subverting the machinery of the host cell.

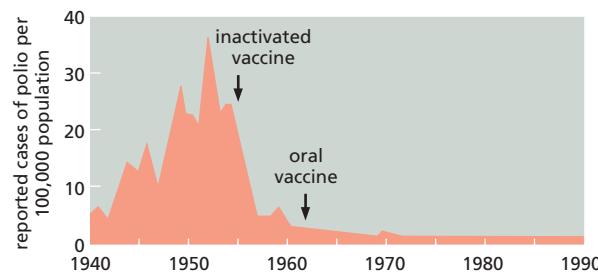


Figure 23–13 Effective control of a viral disease through vaccination. The graph shows the number of cases of poliomyelitis reported per year in the United States. The arrows indicate the timing of the introduction of the Salk vaccine (inactivated virus given by injection) and the Sabin vaccine (live attenuated virus given orally).

CELL BIOLOGY OF INFECTION

The mechanisms through which pathogens cause disease are as diverse as the pathogens themselves. Nonetheless, all pathogens must carry out certain common tasks: they must gain access to 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 Overcome Epithelial Barriers to Infect the Host

The first step in infection is for the pathogen to gain access to the host. A thick 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, the barrier is just a single monolayer of epithelial cells. Nonetheless, all these epithelia serve as barriers to infection.

Wounds in barrier epithelia allow pathogens direct access to unoccupied niches within otherwise sterile host tissues. 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. *Staphylococci* from the skin and nose, or *Streptococci* from the throat and mouth, are two examples of opportunistic bacterial pathogens that are responsible for many serious infections resulting from breaches in epithelial barriers. The recent emergence of bacterial strains of *Staphylococcus* that are resistant to the antibiotics commonly used for treatment (for example, methicillin-resistant *Staphylococcus aureus*, or MRSA, which infects up to 50,000,000 people worldwide) is of particular concern. Papillomaviruses, which cause warts and cervical cancer, also take advantage of breaches in epithelial barriers.

Primary pathogens, however, need not wait for a wound to gain access to their host. One efficient way for such a pathogen to cross the skin is to catch a ride in the saliva of a biting arthropod. A diverse group of bacteria, viruses, and protozoa has developed the ability to survive in insects 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 others that are specialized for survival in a mosquito (see Figure 23–9). Viruses that are spread by insect bites cause yellow fever and Dengue fever, as well as many kinds of viral encephalitis (inflammation of the brain). These viruses replicate in both insect cells and mammalian cells, as required for their transmission by an insect vector.

The efficient spread of a pathogen via an insect vector requires that an individual insect consumes a blood meal from an infected host and transfers the pathogen to a naive host. In a few striking cases, the pathogen alters the behavior of the insect so that its transmission to a new host is more likely. An example is the bacterium *Yersinia pestis*, which causes bubonic plague. It multiplies in the flea's foregut to form aggregated masses that physically block the digestive tract; during each repeated, but futile, attempt at feeding, some of the bacteria in the foregut are flushed into the bite site, thus transmitting plague to a new host (Figure 23–14).

Pathogens That Colonize an Epithelium Must Overcome Its Protective Mechanisms

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 and the bladder, are normally kept nearly sterile. How do these epithelia avoid bacterial colonization? A layer of protective mucus covers the respiratory epithelium, and the coordinated beating of cilia sweeps the mucus and trapped bacteria up and out of the lung. The epithelial lining of the bladder and the upper gastrointestinal tract also has a thick layer of mucus, and

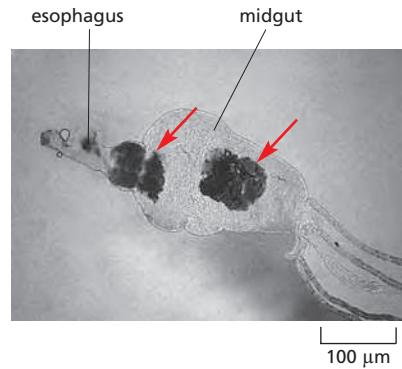
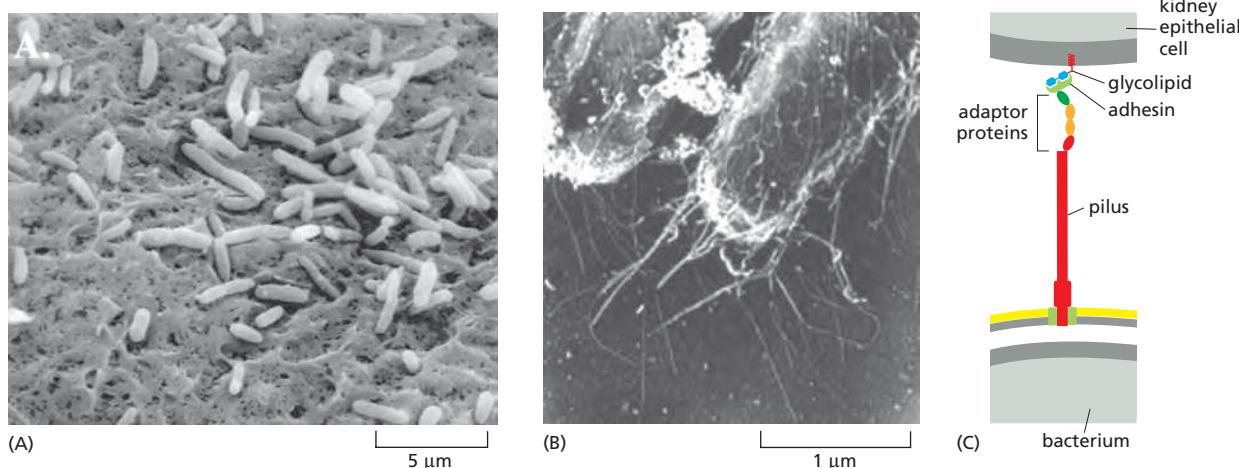


Figure 23–14 Plague bacteria within a flea. This light micrograph shows the digestive tract dissected from a flea that had dined about two weeks previously on the blood of an animal infected with the plague bacterium, *Yersinia pestis*. The bacteria multiplied in the flea gut to produce large cohesive 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. Schwan, *J. Infect. Dis.* 178:1406–1415, 1998.)



these organs are periodically flushed by urination and by peristalsis, respectively, which washes away most microbes.

Pathogenic bacteria and eukaryotic parasites that infect these epithelial surfaces have evolved specific mechanisms for overcoming these protective mechanisms. Those that infect the urinary tract, for example, adhere tightly to the epithelial lining via specific **adhesins**, which are proteins or protein complexes that recognize and bind to cell-surface molecules on the epithelium. An important group of adhesins in *E. coli* strains that infect the kidney are components of the *pili*—surface projections that can be several micrometers long and thus able to span the thickness of the protective mucus layer; at the tip of each pilus is an adhesin protein that binds tightly to the D-galactose-D-galactose disaccharide on glycolipids on the surface of kidney cells (Figure 23–15). Strains of *E. coli* that infect the bladder rather than the kidney express a second kind of pilus with a different adhesin protein that binds to bladder epithelial cells. It is the specificity of the adhesin proteins on the tips of the two types of pili that is responsible for the bacteria's colonizing of the different parts of the urinary tract.

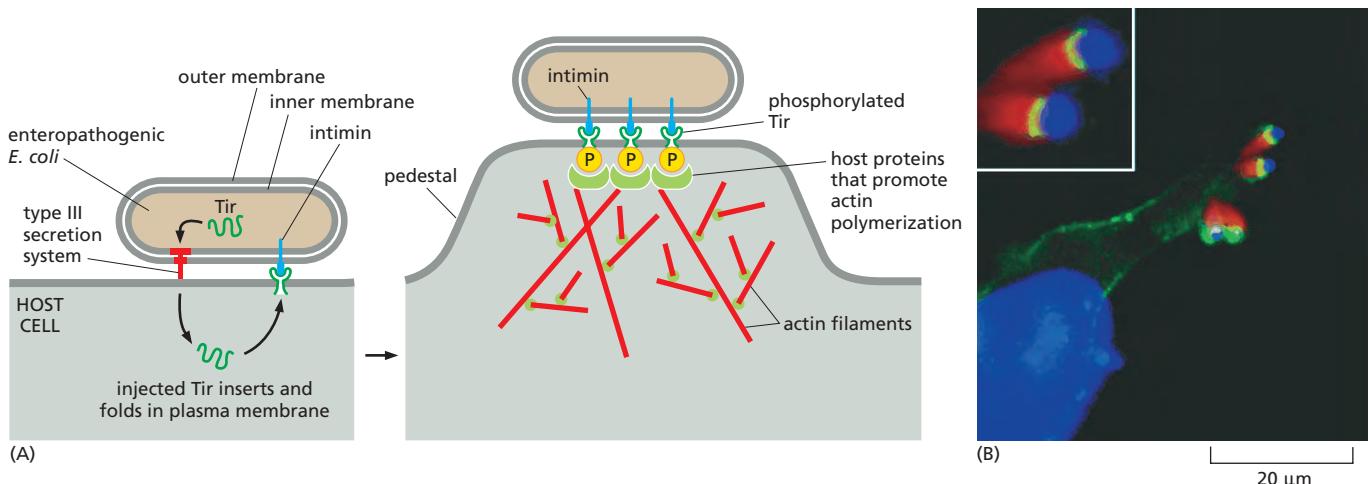
The stomach is an especially hostile environment for pathogens. Besides the thick layer of mucus and peristaltic washing, it is filled with acid (average pH ≈ 2), which is lethal to almost all bacteria ingested in food. Yet, it is home to a microbiota of hundreds of resident species, including the bacterium *H. pylori*, which, as we discussed earlier, is the major cause of stomach ulcers and some stomach cancers. The hypothesis that a persistent bacterial infection could cause stomach ulcers was initially met with skepticism. The young Australian doctor who made the initial discovery finally proved the point: he drank a pure culture of *H. pylori* and developed inflammation of the stomach, which often precedes the development of ulcers. A short course of antibiotics can now effectively cure a patient of recurrent stomach ulcers. Remarkably, *H. pylori* is able to persist for life as a commensal in most humans. One way in which it survives in the stomach is by producing the enzyme *urease*, which converts urea to ammonia that neutralizes the acid in its immediate vicinity. The bacterium also uses its flagellum for chemotactic motility, allowing it to seek out the more neutral pH near the surface of gastric epithelial cells. *H. pylori* virulence proteins that target both epithelial and immune cells help *H. pylori* persist in the stomach, but they can also induce chronic inflammation, alteration in host gene expression, changes in cell proliferation and apoptosis, and disruption of cell-cell junctions, all of which are predisposing factors for stomach cancer.

Extracellular Pathogens Disturb Host Cells Without Entering Them

Extracellular pathogens can cause serious disease without entering host cells. *Bordetella pertussis*, the bacterium that causes whooping cough, for example, colonizes the respiratory epithelium and circumvents the normal mechanism that

Figure 23–15 Pathogenic *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. (C) An *E. coli* pilus has adaptor proteins on its tip that bind to glycolipids on the surface of kidney cells. (A, from G.E. Soto and S.J. Hultgren, *J. Bacteriol.* 181:1059–1071, 1999. With permission from the American Society for Microbiology; B, courtesy of D.G. Thanassi and S.J. Hultgren, *Methods* 20:111–126, 2000. With permission from Academic Press.)



clears the respiratory tract by expressing adhesins that bind ciliated epithelial cells. The adherent bacteria produce 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 the cholera toxin discussed above, has an A subunit that 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 (see Figure 23–6). This toxin also interferes with the chemotactic pathway that neutrophils use to seek out and destroy invading bacteria (see Figures 16–3 and 16–86). *B. pertussis* colonization of the respiratory tract causes severe coughing, which helps spread the infection.

Not all extracellular pathogens that colonize an epithelium exert their effect through toxins. Enteropathogenic *E. coli* (EPEC), which causes diarrhea in young children, uses a type III secretion system (see Figure 23–7) to deliver its own special receptor protein (called *Tir*) into the plasma membrane of a host intestinal epithelial cell (Figure 23–16). The extracellular domain of *Tir* binds to the bacterial surface protein *intimin*, triggering actin polymerization in the host cell that results in the formation of a unique cell-surface protrusion called a *pedestal*; this pushes the tightly adherent bacteria up about 10 μm from the host-cell membrane, thereby promoting bacterial movement along the cell surface. A similar strategy is used by vaccinia virus (the virus that was used as a vaccine to eradicate smallpox) to form mobile pedestals, which promote spread of the virus from cell to cell. The study of how EPEC and vaccinia virus promote actin polymerization has been of major importance in understanding how intracellular signaling pathways regulate the cytoskeleton in normal, uninfected cells (discussed in Chapter 16). Although pedestal formation promotes the spread of these pathogens, the symptoms of EPEC infection (severe diarrhea) are caused by the loss of absorptive microvilli and disruption of signaling pathways in epithelial cells, which are triggered by *Tir* and other effector proteins.

Intracellular Pathogens Have Mechanisms for Both Entering and Leaving Host Cells

Many pathogens have to enter host cells to cause disease. These **intracellular pathogens** include all viruses and many bacteria and protozoa. Each of these has a preferred niche for replication and survival within host cells. Bacteria and protozoa replicate either in the cytosol or within a membrane-enclosed compartment. While most RNA viruses replicate within the cytosol, most DNA viruses replicate in the nucleus. Life inside a host cell has several advantages. The pathogens are not accessible to *antibodies*, nor are they easy targets for phagocytic cells (discussed in Chapter 24); furthermore, intracellular bacteria and protozoa are bathed in a rich source of nutrients, and viruses have access to the host cell's biosynthetic

Figure 23–16 Interaction of enteropathogenic *E. coli* (EPEC) with host intestinal epithelial cells. (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*. Next, a host-cell protein tyrosine kinase phosphorylates the intracellular domain of *Tir* on tyrosines. Phosphorylated *Tir* recruits host-cell proteins (including an adaptor protein, a WASp protein, and the Arp 2/3 complex) that trigger actin polymerization (see Figure 16–16). Consequently, a branched network of actin filaments assembles underneath the bacterium, forming an actin pedestal (**Movie 23.4**). (B) EPEC on a pedestal. In this fluorescence micrograph, the DNA of the EPEC and host cell is 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. (B, from D. Goosney et al., *Annu. Rev. Cell Dev. Biol.* 16:173–189, 2000. With permission from Annual Reviews.)

machinery for their reproduction. This lifestyle, however, requires that the pathogen have 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. Below we consider some of the myriad ways that individual intracellular pathogens exploit and modify host cell biology to satisfy these requirements.

Viruses Bind to Virus Receptors at 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 by the binding of viral surface proteins to **virus receptors** displayed on the host cell. The first virus receptor identified was an *E. coli* surface protein that is recognized by the bacteriophage lambda; the protein normally functions to transport the sugar maltose from outside the bacterium to the inside where it is used as an energy source. Receptors need not be proteins, however: an envelope protein of herpes simplex virus, for example, binds to heparan sulfate proteoglycans (discussed in Chapter 19) on the surface of certain vertebrate host cells, and simian virus 40 (SV40) binds to a glycolipid. The specificity of virus-receptor interactions often serves as a barrier preventing the spread of a virus from one species to another. Acquiring the ability to bind to a new receptor often requires multiple changes in a virus, but it can be crucial in allowing the cross-species transmission that can result in new disease outbreaks.

Viruses that infect animal cells generally exploit cell-surface receptor molecules that are either ubiquitous (such as the sialic-acid-containing oligosaccharides used by the influenza virus) or found uniquely on those cell types in which the virus replicates (such as the neuron-specific proteins used by rabies virus). Although a virus usually uses a single type of host-cell receptor, some viruses use more than one type. An important example is HIV-1, which requires two types of receptors to enter a host cell. Its primary receptor is CD4, a cell-surface protein on helper T cells and macrophages that is involved in immune recognition (discussed in Chapter 24). It also requires a co-receptor, which is either CCR5 (a receptor for β -chemokines) or CXCR4 (a receptor for α -chemokines), depending on the particular variant of the virus; macrophages are susceptible only to HIV variants that use CCR5 for entry, whereas helper T cells are most efficiently infected by variants that use CXCR4 (Figure 23–17). The viruses that are found within the first few months after HIV infection almost invariably use CCR5, which explains why individuals that carry a defective *CCR5* gene are less susceptible to HIV infection. In the later stages of infection, viruses often either switch to use CXCR4 or adapt to use both co-receptors through the accumulation of mutations; in this way, the virus can change the cell types it infects as the disease progresses. It may seem paradoxical that viruses would infect immune cells, as we might expect that virus binding would trigger an immune response; but invasion of an immune cell can be a useful way for a virus to weaken the immune response and travel around the body to infect other immune cells.

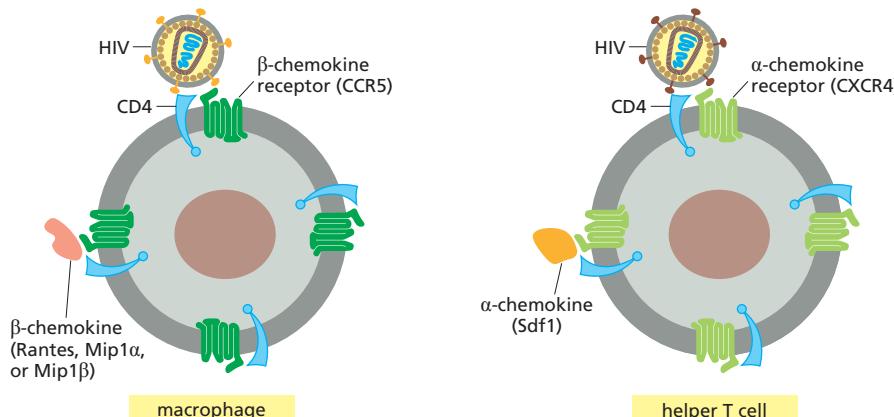


Figure 23–17 Receptor and co-receptors for HIV. All strains of HIV require the CD4 protein as a primary receptor. Early in an infection, most of the viruses use CCR5 as a co-receptor, allowing them to infect macrophages and their precursors, monocytes. As the infection progresses, mutant variants arise 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, Mip1 α , or Mip1 β for CCR5) blocks co-receptor function and prevents viral invasion.

Viruses Enter Host Cells by Membrane Fusion, Pore Formation, or Membrane Disruption

After recognition and attachment to the host cell surface, the virus must enter the cell to replicate. Some **enveloped viruses** enter the host cell by fusing their envelope membrane with the plasma membrane. Most viruses, whether enveloped or nonenveloped, activate signaling pathways in the cell that induce endocytosis, commonly via clathrin-coated pits (see Figure 13–7), leading to internalization into endosomes. Large viruses that do not fit into clathrin-coated vesicles, such as poxviruses, often enter cells by *macropinocytosis*, a process by which membrane ruffles fold over and entrap fluid into macropinosomes (see Figure 13–50). Once inside endosomes, fusion of the viral envelope occurs from the luminal side of the endosome membrane. The mechanism of membrane fusion mediated by viral spike glycoproteins has similarities with SNARE-mediated membrane fusion during normal vesicular trafficking (discussed in Chapter 13).

Enveloped viruses regulate fusion both to ensure that they fuse only with the appropriate host cell membrane and to prevent fusion with one another. For viruses such as HIV-1 that fuse at neutral pH with the plasma membrane (Figure 23–18A), binding to receptors or co-receptors usually triggers a conformational change in a viral envelope protein that exposes a normally buried fusion peptide (see Figure 13–21). Other enveloped viruses, such as influenza A virus, only fuse with a host cell membrane after endocytosis (Figure 23–18B); in this case, it is frequently the acid environment in the late endosome that triggers the conformational change in a viral surface protein that exposes the fusion peptide. The H⁺

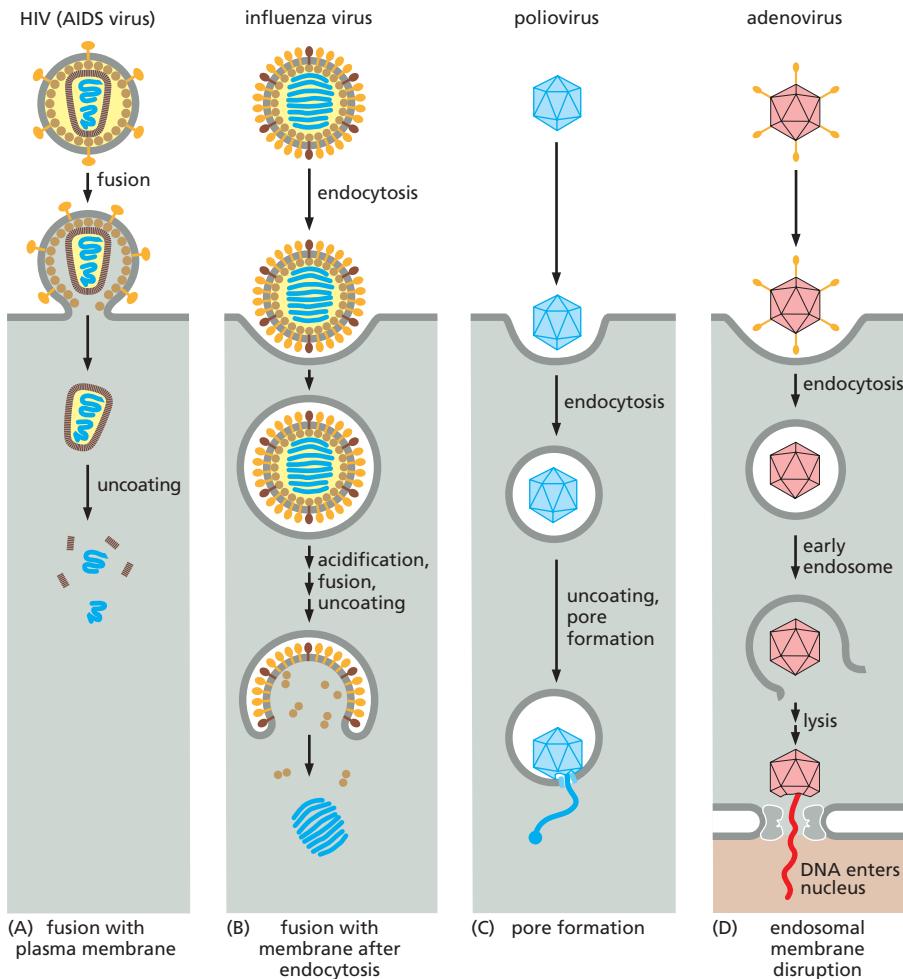


Figure 23–18 Four virus entry strategies.

(A) Some enveloped viruses, such as HIV, fuse directly with the host-cell plasma membrane to release their RNA genome (blue) and capsid proteins (brown) 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 RNA genome (blue) and capsid proteins (brown) into the cytosol. (C) Poliovirus, a nonenveloped virus, induces receptor-mediated endocytosis, and then forms a pore in the endosomal membrane to extrude its RNA genome (blue) into the cytosol. (D) Adenovirus, another nonenveloped virus, uses a more complicated strategy: it induces receptor-mediated endocytosis and then disrupts the endosomal membrane, releasing 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 (**Movie 23.5**).

pumped into the early endosome also has another effect; it enters the influenza virion through an ion channel in the viral envelope and triggers changes in the viral capsid. These priming steps allow the capsids to disassemble once released into the cytosol after virus fusion with the late endosomal membrane.

Nonenveloped viruses use different strategies to enter host cells—strategies that do not rely on membrane fusion. *Poliovirus*, which causes poliomyelitis, binds to a cell-surface receptor, triggering both receptor-mediated endocytosis (see Figure 13–52) and a conformational change in the viral particle. The conformational change exposes a hydrophobic projection on one of the capsid proteins, which inserts into the endosomal membrane to form a pore. The viral RNA genome then enters the cytosol through the pore, leaving the capsid in the endosome (Figure 23–18C). Other nonenveloped viruses such as *adenovirus* disrupt the endosomal membrane after they are taken up by receptor-mediated endocytosis. One of the proteins released from the capsid lyses the endosomal membrane, releasing the remainder of the virus into the cytosol. During endosomal trafficking and subsequent transport within the cytosol, adenoviruses undergo multiple uncoating steps, which sequentially remove structural proteins and ready the virus particles to release their DNA into the nucleus through nuclear pore complexes (Figure 23–18D).

Bacteria Enter Host Cells by Phagocytosis

Bacteria are much larger than viruses—too large to be taken up either through pores or by receptor-mediated endocytosis. Instead, they enter host cells by phagocytosis, which is a normal function of phagocytes such as neutrophils, macrophages, and dendritic cells (discussed in Chapter 24). These phagocytes patrol the tissues of the body and ingest and destroy microbes; however, some intracellular bacterial pathogens such as *M. tuberculosis* use this to their advantage and have evolved to survive and multiply inside macrophages.

Some bacterial pathogens can invade host cells that are normally nonphagocytic. One way they do so is by expressing an invasion protein that binds with high affinity to a host-cell receptor, which is often a cell-cell or cell-matrix adhesion protein (discussed in Chapter 19). For example, *Yersinia pseudotuberculosis* (a bacterium that causes diarrhea and is a close relative of the plague bacterium *Y. pestis*) expresses a protein called invasin that has an RGD motif that is similar to fibronectin's and likewise is recognized by host-cell β_1 integrins (see Figure 19–55). *Listeria monocytogenes*, which causes a rare but serious form of food poisoning, invades host cells by expressing a protein that binds to the cell-cell adhesion protein E-cadherin (see Figure 19–6). For both these bacterial species, binding of the bacterial invasion proteins to the host cell adhesion proteins stimulates signaling through members of the Rho family of small GTPases (discussed in Chapter 16). This in turn activates proteins in the WASp family and the Arp 2/3 complex, leading to actin polymerization at the site of bacterial attachment. Actin polymerization, together with the assembly of a clathrin coat (see Figure 13–6), drives the advancement of the host cell's plasma membrane over the adhesive surface of the microbe, resulting in the phagocytosis of the bacterium—a process known as the *zipper mechanism* of invasion (Figure 23–19A).

A second pathway by which bacteria can invade nonphagocytic cells is known as the *trigger mechanism* (Figure 23–19B). It is used by various pathogens that cause food poisoning, including *Salmonella enterica*, and it is initiated when the bacterium injects a set of effector molecules into the host-cell cytosol through a type III secretion system (see Figure 23–7). Some of these effector molecules activate Rho family proteins, which in turn stimulate actin polymerization, as just discussed. Other bacterial effector proteins interact with host-cell cytoskeletal elements more directly, nucleating and stabilizing actin filaments and causing the rearrangement of actin cross-linking proteins. The overall effect is to cause the formation of localized ruffles on the surface of the host cell (Figure 23–19C and D), which fold over and engulf the bacteria by a process that resembles macropinocytosis. The appearance of cells being invaded by use of the trigger mechanism

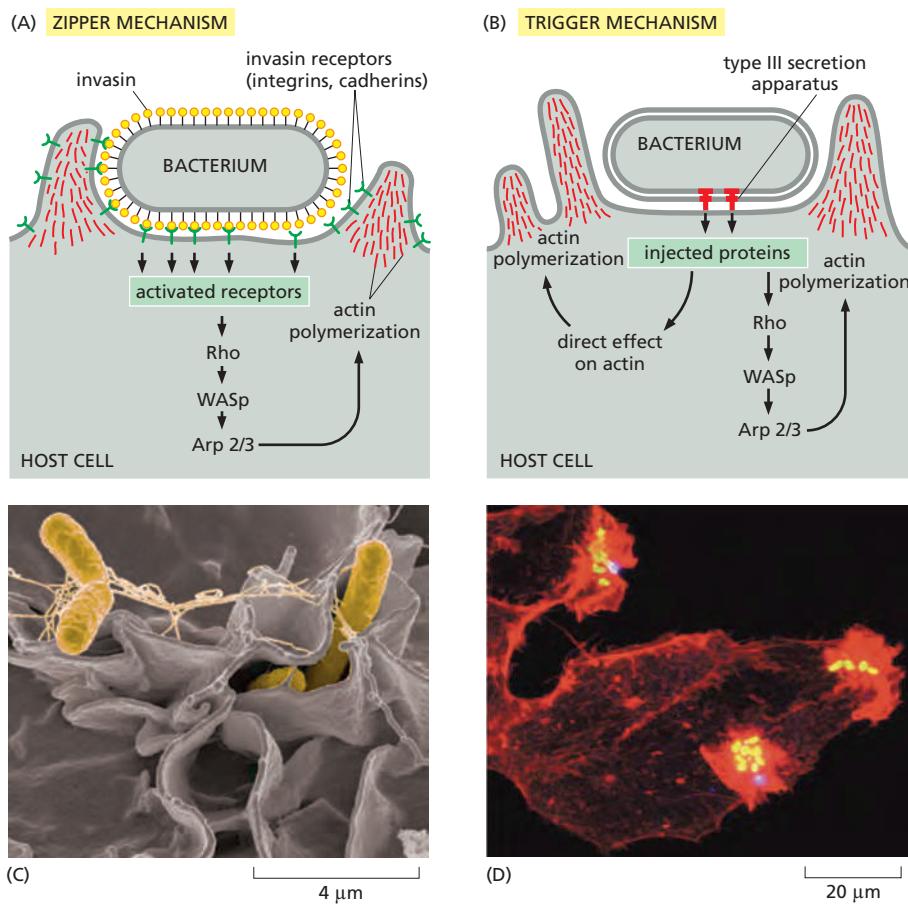


Figure 23–19 Mechanisms used by bacteria to induce phagocytosis by host cells that are normally nonphagocytic. (A) In the zipper mechanism, bacteria express an invasion protein that binds with high affinity to a host-cell receptor, which is often a cell-cell or cell-matrix adhesion protein. (B) In the trigger mechanism, bacteria inject a set of effector molecules into the host-cell cytosol through a type III secretion system called SPI1 (*Salmonella* pathogenicity island 1), inducing membrane ruffling. Both the zipper and trigger mechanisms cause the polymerization of actin at the site of bacterial attachment by activating Rho family small GTPases and the Arp 2/3 complex. (C) A scanning electron micrograph showing a very early stage of *Salmonella enterica* invasion by the trigger mechanism. Bacteria (pseudocolored yellow) are shown 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. (C, from Rocky Mountain Laboratories, NIAID, NIH; D, from J.E. Galán, *Annu. Rev. Cell Dev. Biol.* 17:53–86, 2001. With permission from Annual Reviews.)

is similar to the ruffling induced by some extracellular growth factors, suggesting that the bacteria exploit normal intracellular signaling pathways.

Intracellular Eukaryotic Parasites Actively Invade Host Cells

The uptake of viruses and bacteria into host cells is carried out largely by the host, with the pathogen being a relatively passive participant. In contrast, intracellular eukaryotic parasites, which are typically much larger than other types of intracellular pathogens, invade host cells through a variety of complex pathways that usually require energy expenditure by the parasite.

Toxoplasma gondii, a 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 facilitates entry into the host cell (Figure 23–20). The energy for invasion seems to come from actin polymerization in the parasite rather than host cytoskeleton, and invasion also requires at least one unusual parasite myosin motor protein (Class XIV; see Figure 16–40). At the point of contact, the parasite discharges effector proteins from secretory organelles into the host cell, and these proteins target various host pathways to enable invasion, to block an innate immune response, and promote survival. As the parasite moves into the host cell, a membrane derived from the host-cell plasma membrane surrounds it. Remarkably, the parasite removes host transmembrane proteins from the surrounding membrane as it forms, so that the parasite is protected in a membrane-enclosed compartment that does not fuse with lysosomes and does not participate in host-cell membrane trafficking processes (see Figure 23–20). The specialized membrane is selectively porous: it allows the parasite to take up small metabolic intermediates and nutrients from the host cell's cytosol but excludes macromolecules. Malaria parasites invade human red blood cells using a similar mechanism.

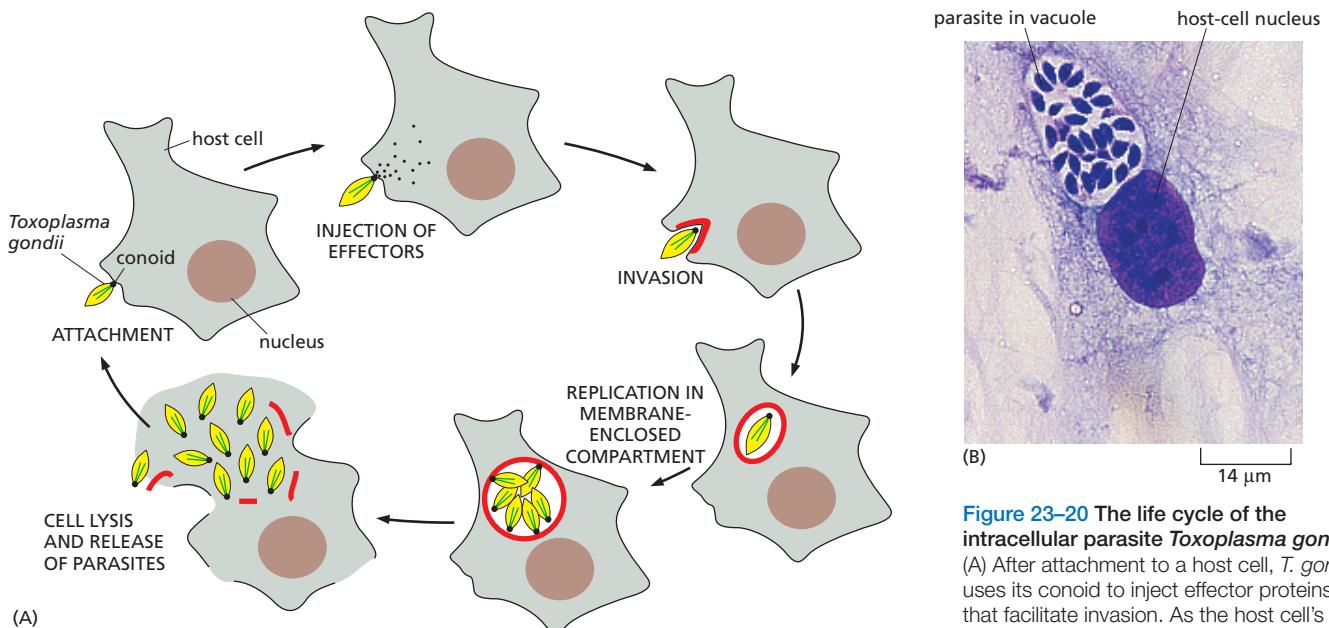


Figure 23–20 The life cycle of the intracellular parasite *Toxoplasma gondii*. (A) After attachment to a host cell, *T. gondii* uses its conoid to inject effector proteins that facilitate invasion. As the host cell's plasma membrane invaginates to surround the parasite, it somehow removes the normal host-cell membrane proteins, 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 parasites to infect other host cells (Movie 23.6). (B) Light micrograph of *T. gondii* replicating within a membrane-enclosed compartment (a vacuole) in a cultured cell. (B, courtesy of Manuel Camps and John Boothroyd.)

The protozoan *Trypanosoma cruzi*, which causes Chagas disease, in Mexico and Central and South America, uses two alternative invasion strategies. In a *lysosome-dependent pathway*, the parasite attaches to host cell-surface receptors, inducing a local increase in Ca^{2+} in the host cell's cytosol. The Ca^{2+} signal recruits lysosomes to the site of parasite attachment, and the lysosomes fuse with the host cell's plasma membrane, allowing the parasites rapid access to the lysosomal compartment (Figure 23–21). In a *lysosome-independent pathway*, the parasite penetrates the host-cell plasma membrane by inducing the membrane to invaginate, without the recruitment of lysosomes.

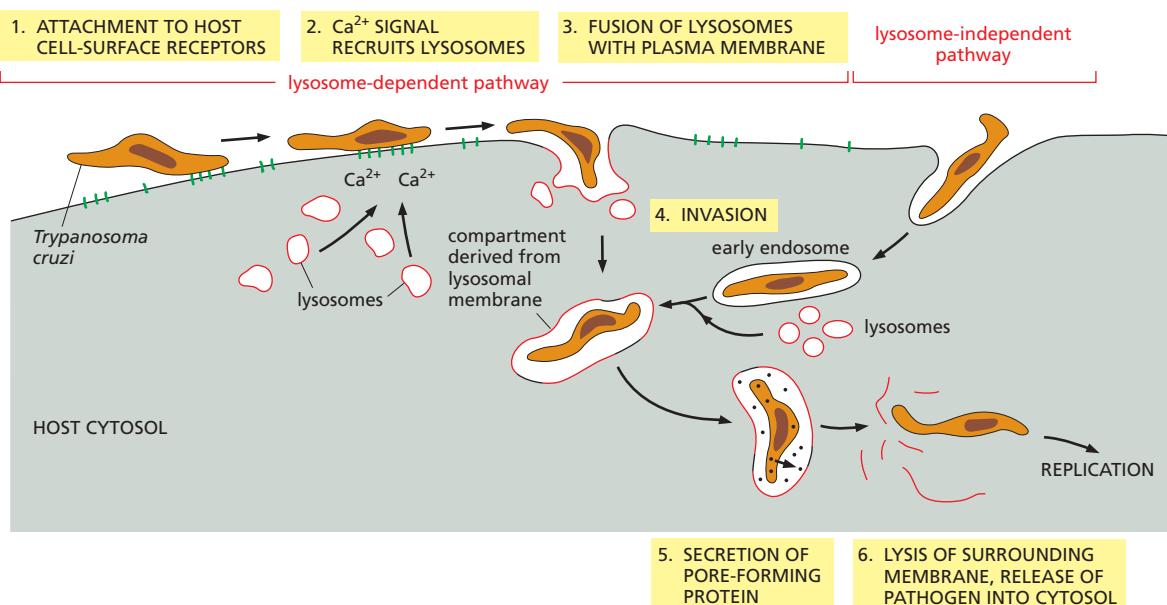


Figure 23–21 The two alternative strategies that *Trypanosoma cruzi* uses to invade host cells. In the lysosome-dependent pathway (left), *T. cruzi* recruits host-cell lysosomes to its site of attachment to the host cell. The lysosomes fuse with the invaginating plasma membrane to create an intracellular 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, thereby allowing the parasite to escape into the host-cell cytosol and proliferate. In the lysosome-independent pathway (right), the parasite induces the host plasma membrane to invaginate and pinch off without recruiting lysosomes; then lysosomes fuse with the endosome prior to the parasite's escape into the cytosol.

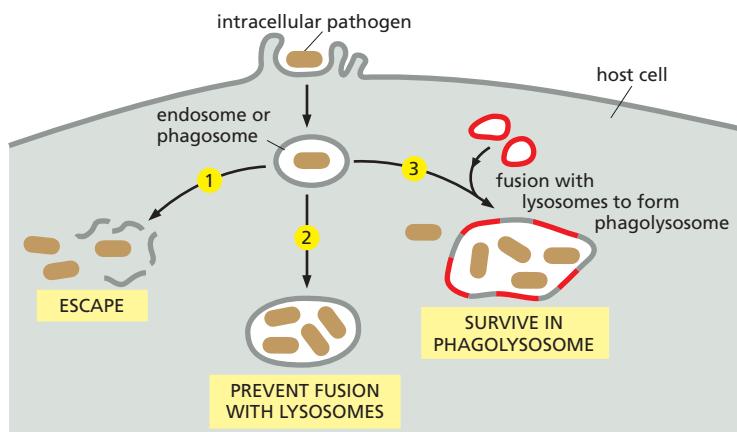


Figure 23–22 Choices that an intracellular pathogen faces. After entry into a host cell, generally through phagocytosis into a membrane-enclosed compartment, intracellular pathogens can use one of three strategies to survive and replicate. Pathogens that follow strategy (1) include all viruses, *Trypanosoma cruzi*, *Listeria monocytogenes*, and *Shigella flexneri*. Those that follow strategy (2) include *Mycobacterium tuberculosis* and *Legionella pneumophila*. Those that follow strategy (3) include *Salmonella enterica*, *Coxiella burnetii*, and *Leishmania*.

Some Intracellular Pathogens Escape from the Phagosome into the Cytosol

The intracellular parasites just discussed raise a general problem that faces all intracellular pathogens, including viruses, bacteria, and eukaryotic parasites: they must find a cell compartment in which they can replicate. After their endocytosis by a host cell, they usually find themselves in an endosomal compartment, which normally would fuse with lysosomes to form a *phagolysosome*—a dangerous place for pathogens. To survive, pathogens use a variety of strategies. Some escape from the endosomal compartment before such fusion. Others remain in the endosomal compartments but modify it so that it no longer fuses with lysosomes. Still others have evolved to weather the harsh conditions in the phagolysosome (Figure 23–22).

Trypanosoma cruzi uses the escape route by secreting a pore-forming toxin that lyses the lysosome membrane, releasing the parasite into the host cell's cytosol (see Figure 23–21). The bacterium *Listeria monocytogenes* uses a similar strategy. Following phagocytosis by the zipper mechanism, it secretes a protein called *listeriolysin O*, which disrupts the phagosomal membrane, releasing the bacteria into the cytosol (Figure 23–23).

Many Pathogens Alter Membrane Traffic in the Host Cell to Survive and Reproduce

The survival and reproduction of many intracellular pathogens requires that they modify membrane (vesicular) traffic in the host cell. They may, for example, prevent the normal fusing of endosomes with lysosomes, or adapt themselves to

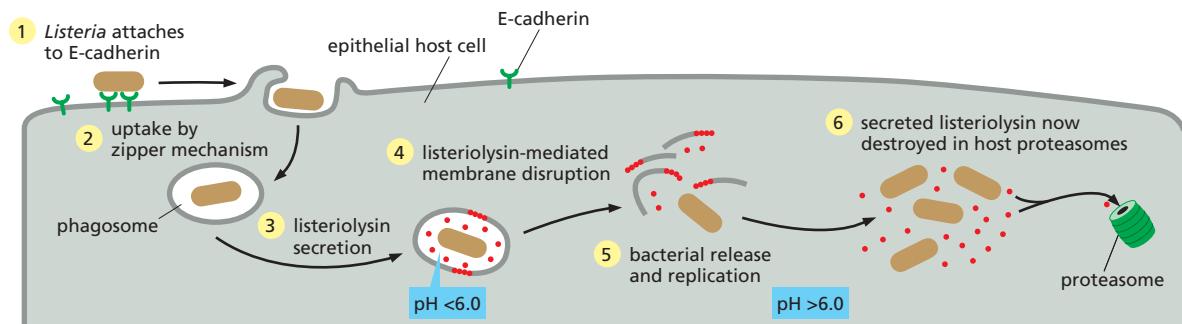


Figure 23–23 Escape of *Listeria monocytogenes* by selective destruction of the phagosomal membrane. The bacterium attaches to E-cadherin on the surface of host epithelial cells and induces its own uptake by the zipper mechanism (see Figure 23–19A). Within the phagosome, the bacterium secretes the protein *listeriolysin O*, which is activated at pH < 6 and forms oligomers in the phagosomal 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 pH in the cytosol is > 6, however, the *listeriolysin O* there is inactive and is also rapidly degraded by proteasomes. Thus, the host cell's plasma membrane remains intact.

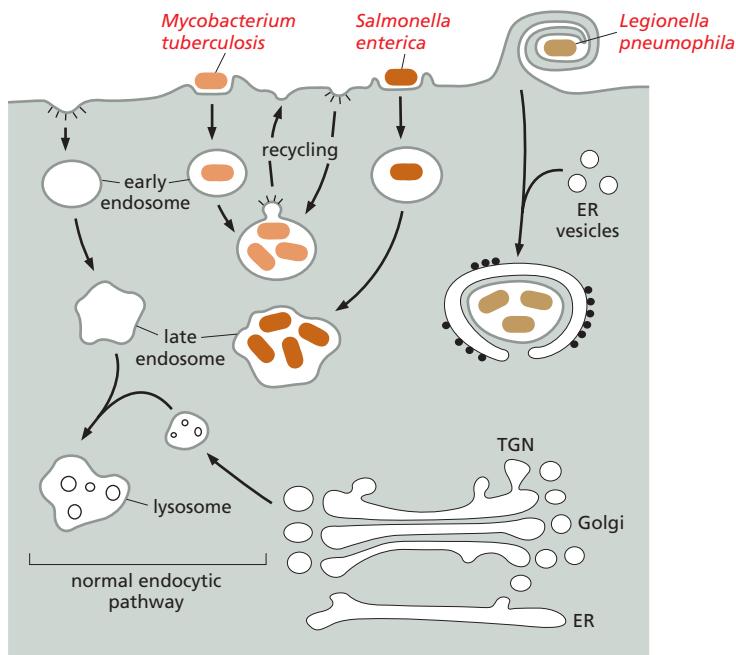


Figure 23–24 Modifications of membrane traffic in host cells by bacterial pathogens. Intracellular bacterial pathogens, including *Mycobacterium tuberculosis*, *Salmonella enterica*, and *Legionella pneumophila*, all replicate in membrane-enclosed compartments, but the compartments differ. *M. tuberculosis* 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 rough endoplasmic reticulum (ER) membrane and communicates with the ER via transport vesicles. TGN, trans Golgi network.

resist the lysosome's antimicrobial armaments. Intracellular pathogens must also provide a pathway for importing nutrients from the host cytosol into their compartment of choice.

Different pathogens have distinct strategies for altering membrane traffic in the host cell (Figure 23–24). *M. tuberculosis* prevents the early endosome that contains the bacteria from maturing, so the endosome never acidifies or acquires the other characteristics of a late endosome or lysosome. This strategy requires the activity of its type VII secretion system, as well as mycobacterial lipid products that mimic host lipids and influence vesicular traffic. Phagosomes containing *Salmonella enterica*, in contrast, acidify and acquire markers of late endosomes and lysosomes, but the bacteria slow the process of phagosomal maturation. They do so by injecting effector proteins through a second type III secretion system. These effectors activate host kinesin motor proteins to pull membrane tubules outward from the phagosome along cytoplasmic microtubules, forming a specialized compartment called the *Salmonella*-containing vacuole (Figure 23–25).

Other bacteria seem to find shelter in intracellular compartments that are distinct from those of the usual endocytic system. One example is *Legionella pneumophila*, which was first recognized as a human pathogen in 1976, when it was found to be the cause of a type of pneumonia known as **Legionnaire's disease**. *L. pneumophila* is normally a parasite of freshwater amoebae, but it is commonly

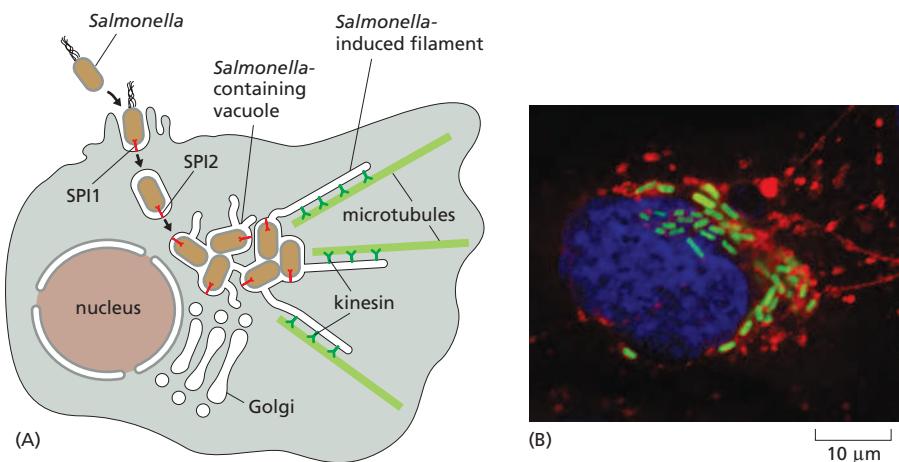


Figure 23–25 *Salmonella enterica* residing in a modified phagosomal compartment called the *Salmonella*-containing vacuole. These bacteria invade the host cell using an SPI1 type III secretion system to inject effector proteins that induce the trigger mechanism of microbe entry illustrated in Figure 23–19B. (A) Following its engulfment into a phagosome, the bacterium inactivates its SPI1 type III secretion system and activates its SPI2 type III secretion system to inject different effector proteins, which remodel the phagosome into the specialized *Salmonella*-containing vacuole. One of the injected effector proteins activates host kinesin motor proteins to pull membrane tubules outward toward the plus ends of the microtubules (see Figure 16–42). (B) Fluorescence micrograph showing *S. enterica* in a *Salmonella*-containing vacuole. The bacteria are stained green, the microtubules red, and the nucleus blue. (B, courtesy of Stephane Meresse.)

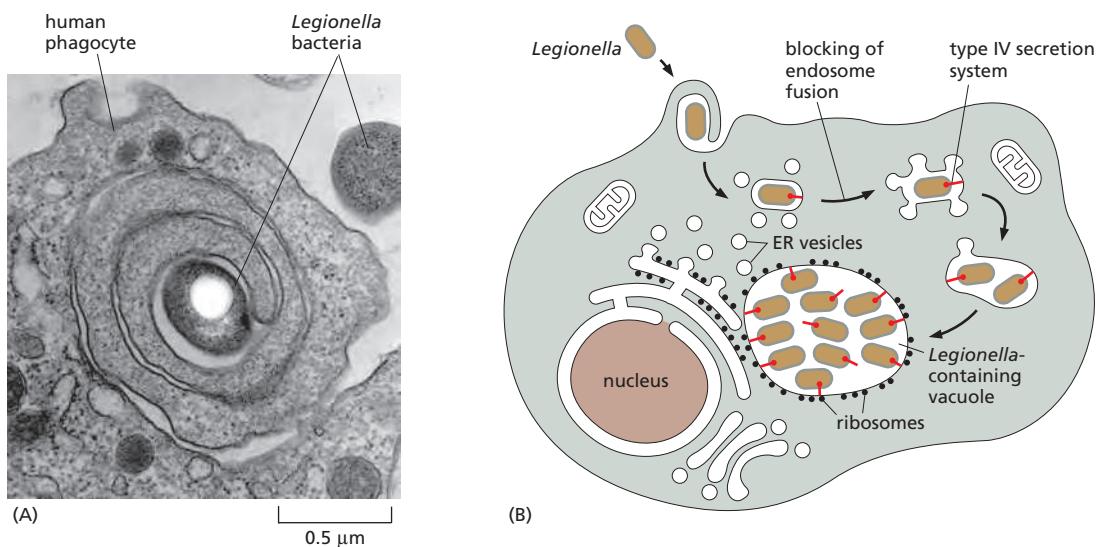


Figure 23–26 *Legionella pneumophila* residing in a compartment with characteristics similar to those of the rough endoplasmic reticulum (ER). (A) Electron micrograph showing the unusual coiled structure that the *Legionella pneumophila* bacterium induces on the surface of a phagocyte during the invasion process. Some other pathogens, including the bacterium *Borrelia burgdorferi*, which causes Lyme disease, the eukaryotic pathogen *Leishmania*, and the yeast *Candida albicans*, can also invade cells using this type of coiling phagocytosis. (B) Following invasion, *L. pneumophila* uses its type IV secretion system to secrete effector proteins that block phagosome–endosome fusion and phagosome maturation. It also secretes effector proteins that promote the fusion of the phagosome with ER-derived vesicles, thereby creating a *Legionella*-containing vacuole with characteristics similar to the rough ER. (A, from M.A. Horwitz, *Cell* 36:27–33, 1984. With permission from Elsevier.)

spread to humans by central air-conditioning systems, which harbor infected amoebae and produce microdroplets of water that are easily inhaled. Once in the lung, the bacteria are engulfed by macrophages by an unusual process called coiling phagocytosis (Figure 23–26A). *L. pneumophila* uses a type IV secretion system to inject effector proteins into the phagocyte that modulate the activity of proteins that regulate vesicular traffic, including SNARE proteins and Rab and Arf family small GTPases (discussed in Chapter 13). The effector proteins thereby prevent the phagosome from fusing with endosomes and promote its fusion with vesicles derived from the endoplasmic reticulum, converting the phagosome into a compartment that resembles the rough endoplasmic reticulum (Figure 23–26B).

Viruses can also alter membrane traffic in the host cell. Enveloped viruses make use of host cell membranes to acquire their own envelope membrane. In the simplest cases, virally encoded glycoproteins are inserted into the endoplasmic reticulum membrane and follow the secretory pathway through the Golgi apparatus to the plasma membrane; the viral capsid proteins and genome assemble into nucleocapsids, which acquire their envelope as they bud off from the plasma membrane (see Figure 23–12). This mechanism is used by many enveloped viruses including HIV-1. Other enveloped viruses such as herpesviruses and vaccinia virus acquire their lipid envelopes in more complex ways (Figure 23–27).

Viruses and Bacteria Use the Host-Cell Cytoskeleton for Intracellular Movement

As mentioned earlier, many pathogens escape into the cytosol rather than remaining in a membrane-enclosed compartment. The cytosol of mammalian cells is extremely viscous, as it is crowded with protein complexes, organelles, and cytoskeletal filaments, all of which inhibit the diffusion of particles the size of a bacterium or a viral nucleocapsid. Thus, to reach a particular region of the host cell a pathogen must be actively moved there. As with transport of intracellular organelles, pathogens generally use the host cell's cytoskeleton for their active movement.

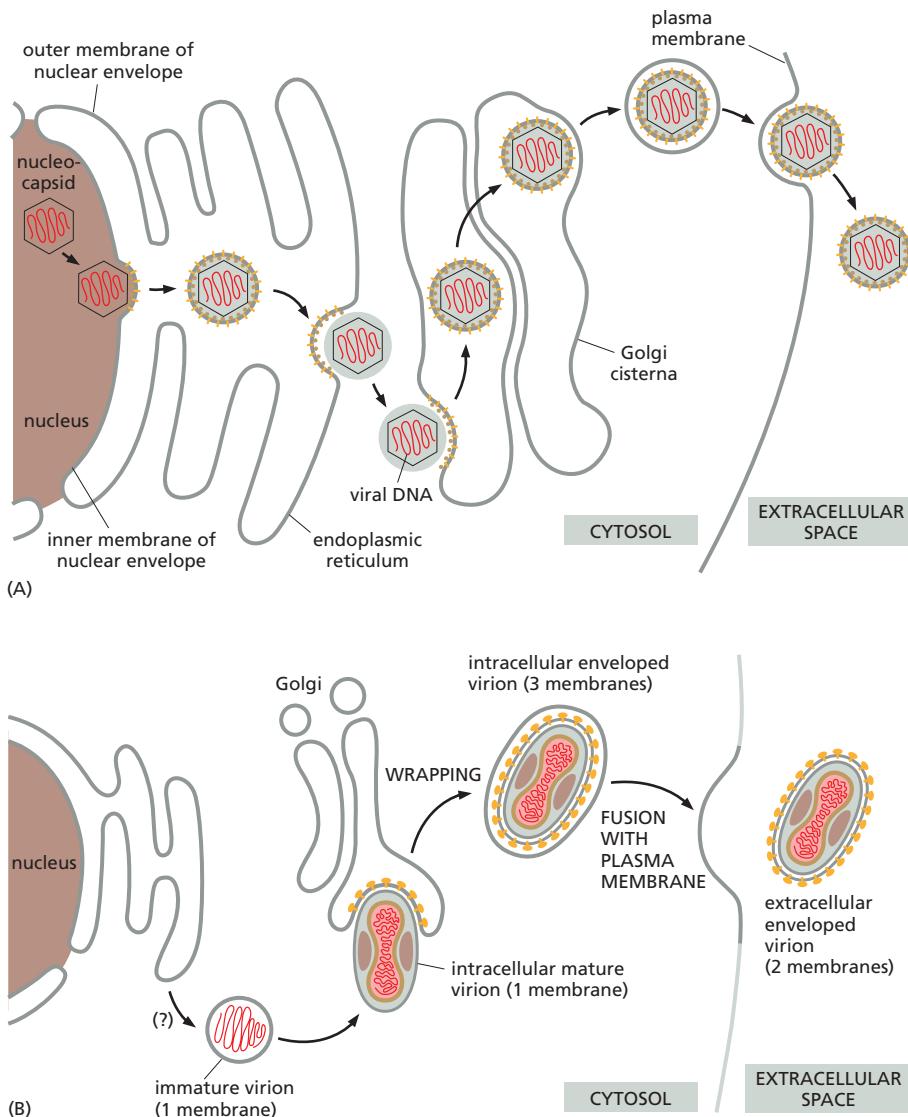


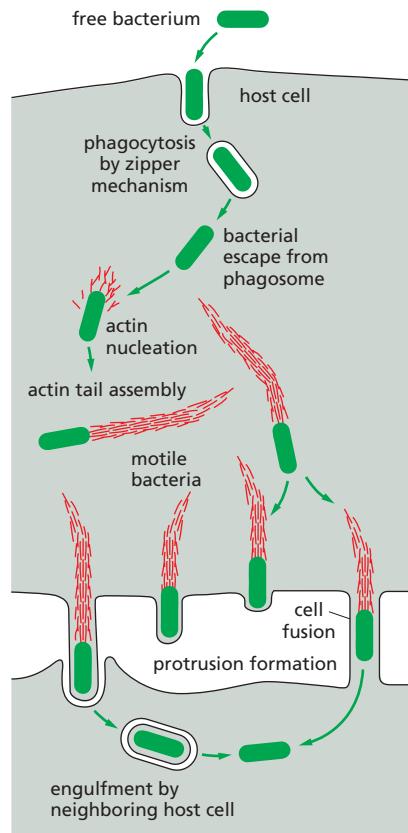
Figure 23–27 Complex strategies for viral envelope acquisition. (A) Herpesvirus nucleocapsids assemble in the nucleus and then bud through the inner nuclear membrane into the space between the inner and outer nuclear membranes, acquiring a lipid bilayer membrane coat. The virus particles then apparently lose this coat when they fuse with the endoplasmic reticulum membrane to escape into the cytosol. Subsequently, the nucleocapsids bud into the Golgi apparatus and bud out again on the other side, thereby acquiring two new membrane coats in the process. The virus then buds from the cell surface 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” in the cytosol, far away from the plasma membrane. The immature virion, with one membrane, is then surrounded by two additional membranes, both acquired from the Golgi apparatus by a poorly understood wrapping mechanism, to form the intracellular enveloped virion. After fusion of the outermost membrane with the host-cell plasma membrane, the extracellular enveloped virion is released from the host cell.

Several bacteria that replicate in the host cell's cytosol have adopted a remarkable mechanism that depends on actin polymerization for movement. These bacteria include the human pathogens *Listeria monocytogenes*, *Shigella flexneri*, *Rickettsia rickettsii* (which causes Rocky Mountain spotted fever), and *Burkholderia pseudomallei* (which causes melioidosis, a disease characterized by severe respiratory symptoms). Baculovirus, an insect virus, also uses this mechanism for intracellular movement. All of these pathogens induce the nucleation and assembly of host-cell actin filaments at one pole of the bacterium or virus. The growing filaments generate force and push the pathogens through the cytosol at rates of up to 1 $\mu\text{m/sec}$ (Figure 23–28). New filaments form at the rear of each pathogen and are left behind like a rocket trail as the microbe advances; the filaments depolymerize within a minute or so as they encounter depolymerizing factors in the cytosol. For *L. monocytogenes* and *S. flexneri*, the moving bacteria collide with the plasma membrane and move outward, inducing the formation of long, thin, host-cell protrusions with the bacteria at their tip. As shown in Figure 23–28, a neighboring cell often engulfs these projections, allowing the bacteria to enter the neighbor's cytoplasm without exposure to the extracellular environment, thereby avoiding antibodies produced by the host's adaptive immune system. For *B. pseudomallei*, movement and collision of the bacteria with the plasma membrane promotes cell–cell fusion, which serves a similar purpose of immune avoidance while allowing continued bacterial replication.

Figure 23–28 The actin-based movement of bacterial pathogens within and between host cells. (A) Following invasion, bacterial pathogens such as *Listeria monocytogenes*, *Shigella flexneri*, *Rickettsia rickettsii*, and *Burkholderia pseudomallei* induce the assembly of actin-rich tails in the host-cell cytoplasm, which drives rapid bacterial movement. For most of these pathogens, the moving bacteria collide with the host-cell plasma membrane to form membrane-covered protrusions, which are engulfed by neighboring cells—spreading the infection from cell to cell. In contrast, for *B. pseudomallei*, collision with the plasma membrane promotes cell–cell fusion, creating a conduit through which bacteria can invade neighboring cells (Movie 23.7).

The molecular mechanisms of pathogen-induced actin assembly differ for the different pathogens, suggesting that they evolved independently (Figure 23–29). *L. monocytogenes* and baculovirus produce proteins that directly bind to and activate the Arp 2/3 complex to initiate the formation of an actin tail and movement (see Figure 16–16). *S. flexneri* produces an unrelated surface protein that binds to and activates N-WASp, which then activates the Arp 2/3 complex. *Rickettsia* species produce a protein that directly polymerizes actin by mimicking the function of host formin proteins (see Figure 16–17).

Many viral pathogens rely primarily on microtubule-dependent motor proteins rather than actin polymerization to move within the host-cell cytosol. Viruses that infect neurons, such as the neurotropic alpha herpesviruses, which include the virus that causes chickenpox, provide important examples. The virus enters sensory neurons at the tips of their axons, and microtubule-based retrograde “backward” axonal transport carries the nucleocapsids down the axon to the nucleus. The transport is mediated by attachment of viral capsid proteins to the motor protein dynein (see Figure 16–58). After replication and assembly in the nucleus, the enveloped virions are then carried by antegrade “forward” axonal transport along microtubules to the axon tips, with the transport being mediated by the attachment of a different viral capsid protein to a kinesin motor protein (see Figure 16–56). A large number of viruses associate with either dynein or kinesin motor proteins to move along microtubules at some stage in their replication. As microtubules serve as oriented tracks for vesicular transport in eukaryotic cells, it is not surprising that many viruses have independently evolved the ability to exploit them for their own transport.



Viruses Can Take Over the Metabolism of the Host Cell

Viruses use basic host cell machinery for most aspects of their reproduction: they depend on host-cell ribosomes to produce their proteins, and most use host-cell DNA and RNA polymerases for their own replication and transcription. Many viruses encode proteins that modify the host transcription or translation apparatus to favor the synthesis of viral RNAs and proteins over those of the host cell, shifting the synthetic capacity of the cell toward the production of new virus particles. Poliovirus, for example, encodes a protease that specifically cleaves the TATA-binding component of TFIID (see Figure 6–17), shutting off transcription of most of the host cell's protein-coding genes. Influenza virus produces a protein that blocks both the splicing and the polyadenylation of host-cell RNA transcripts, preventing their export into the cytosol (see Figure 6–38).

Viruses also alter translation by the host. Translation initiation for most host-cell mRNAs depends on recognition of their 5' cap by translation initiation factors (see Figure 6–70). This initiation process 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 encode endonucleases that cleave off the 5' cap from host-cell mRNAs; some go even further by using 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 their own RNA, using internal ribosome entry sites (IRESs) (see Figure 7–68).

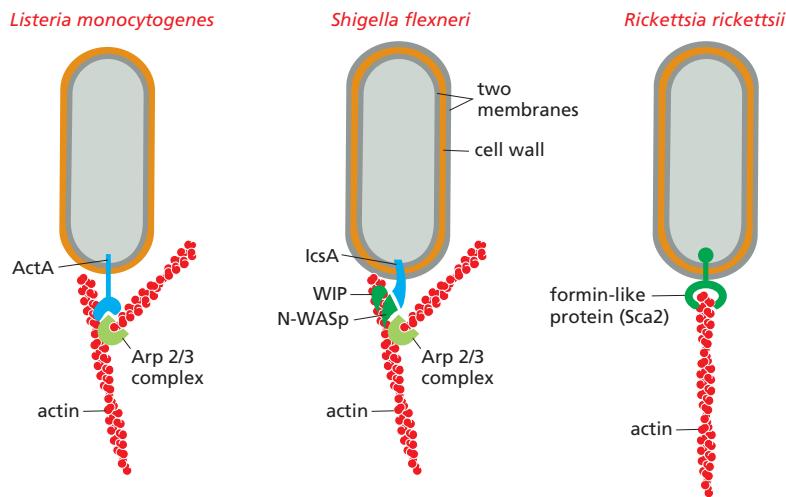


Figure 23-29 Molecular mechanisms for actin nucleation by various bacterial pathogens. *Listeria monocytogenes* and *Shigella flexneri* induce actin nucleation by recruiting and activating the host Arp 2/3 complex (see Figure 16-16), although each uses a different recruitment strategy: *L. monocytogenes* expresses a surface protein, ActA, that directly binds to and activates the Arp 2/3 complex; *S. flexneri* expresses a surface protein, IcsA (unrelated to ActA), that recruits the host protein N-WASp, which in turn recruits the Arp 2/3 complex, along with other host proteins, including WIP (WASp-interacting protein). *Rickettsia rickettsii* uses an entirely different strategy; it expresses a surface protein, Sca2, that directly nucleates actin polymerization by mimicking the activity of host formin proteins.

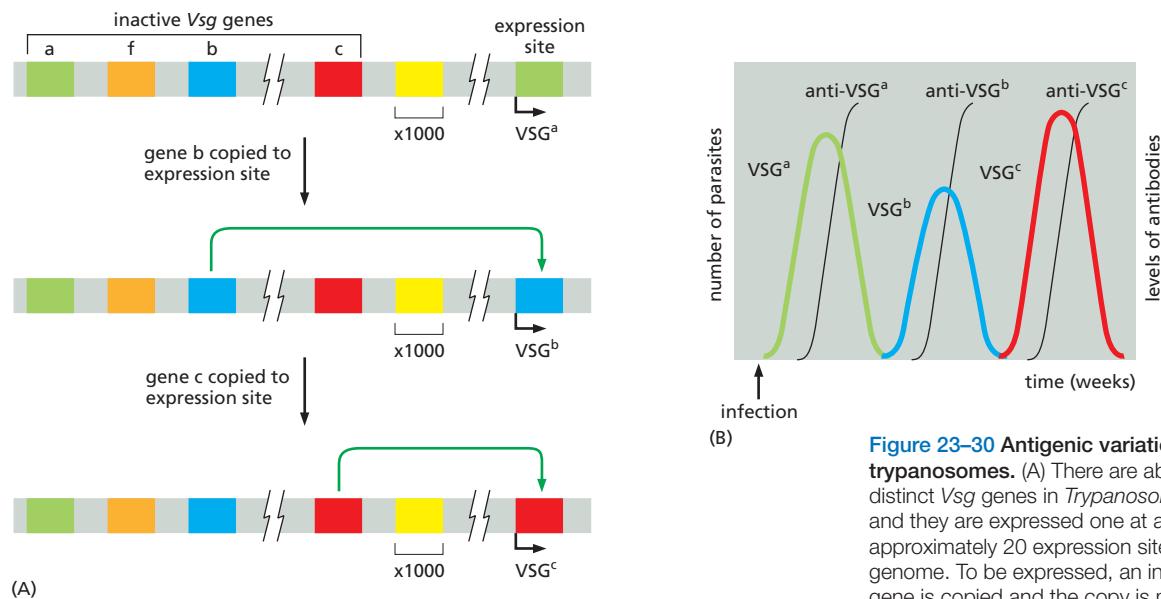
A few DNA viruses use host-cell DNA polymerase to replicate their genome. Unfortunately for these viruses, DNA polymerase is expressed at high levels only during S phase of the cell cycle, and most cells that these viruses infect spend most of their time in G₁ phase. Adenovirus has evolved a mechanism to drive the host cell into S phase, so that the cell produces large amounts of active DNA polymerase, which then replicates the viral genome; to accomplish this, the adenovirus genome also encodes proteins that inactivate both Rb (see Figure 17-61) and p53 (see Figure 17-62), two key suppressors of cell-cycle progression. As might be expected for any mechanism that encourages unregulated DNA replication, these viruses can promote, under some circumstances, the development of cancer. Other DNA viruses, including poxviruses and mimivirus, encode their own DNA and RNA polymerases, as well as some transcription regulators, allowing them to bypass usual host pathways and replicate outside the nucleus.

RNA viruses must always encode their own replication proteins because host cells lack polymerase enzymes that use RNA as a template. For RNA viruses with a single-stranded genome, the replication strategy depends on whether the RNA is a positive [+] strand, which contains translatable information like mRNA, or a complementary negative [-] strand. When the RNA is a positive [+] strand, the incoming viral genome is used to produce the viral RNA polymerase and viral proteins; the viral polymerase is then used to replicate the viral RNA and to generate mRNAs for the production of more viral proteins. For viruses with a negative [-] strand RNA genome (such as influenza and measles virus), an RNA polymerase enzyme is packaged as a structural protein of the incoming viral capsids.

Retroviruses such as HIV-1, which have a positive [+] strand RNA genome, are a special class of RNA virus because they carry with them a viral *reverse transcriptase* enzyme. After entry to the host cell, the reverse transcriptase uses the viral RNA genome as a template to synthesize a double-stranded DNA copy of the viral genome, which enters into the nucleus and integrates into the host cell's chromosomes (see Figure 5-62). It is later transcribed by the cell's DNA-dependent RNA polymerase to produce viral genomes and proteins.

Pathogens Can Evolve Rapidly by Antigenic Variation

The complexity and specificity of the interplay 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 more difficult to prevent or treat. Pathogens have two advantages that enable them to evolve rapidly. First, they replicate very quickly, providing a great deal of material for natural selection to work with. 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 act rapidly on this genetic variation. The host's adaptive immune system and modern microbicidal drugs, both of which destroy pathogens that fail to change, are the main sources of these selective pressures.

An example of an adaptation to the selective pressure imposed by the adaptive immune system 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's surface (discussed in Chapter 24). Many pathogens have evolved mechanisms that deliberately change these antigens during the course of an infection, enabling them to evade antibodies. Some eukaryotic parasites, for example, undergo programmed rearrangements of the genes encoding their surface antigens. A striking example occurs in *Trypanosoma brucei*, a protozoan parasite that causes African sleeping sickness and is spread by tsetse flies. (*T. brucei* is a relative of *T. cruzi*—see Figure 23–21—but it replicates extracellularly rather than intracellularly.) *T. brucei* is covered with a single type of glycoprotein, called *variant-specific glycoprotein* (VSG), which elicits in the host a protective antibody response that rapidly clears most of the parasites. The trypanosome genome, however, contains about 1000 different Vsg genes or pseudogenes, each encoding a VSG with a distinct amino acid sequence. Only one of these genes is expressed at any one time, from one of approximately 20 possible expression sites in the genome. Gene rearrangements that copy different Vsg genes into expression sites repeatedly change the VSG protein displayed on the surface of the pathogen. In this way, a few trypanosomes with an altered VSG escape the initial antibody-mediated clearance, replicate, and cause the disease to recur, leading to a chronic cyclic infection (Figure 23–30).

Bacterial pathogens can also rapidly change their surface antigens. As discussed in Chapter 5, *Salmonella enterica* bacteria switch between expressing either of two versions of the protein flagellin, the structural component of the bacterial flagellum (see Figure 23–3D), in a process called **phase variation** (see Figure 5–65). Species of the genus *Neisseria* are also champions at this. These Gram-negative cocci can cause meningitis and sexually transmitted diseases. They undergo genetic recombination very similar to that just described for eukaryotic pathogens, which enables them to vary the pilin protein they use to attach to host cells. By inserting one of the multiple silent copies of variant *pilin* genes into a single expression locus, they can express many slightly different versions of the protein and repeatedly change the amino acid sequence over time. *Neisseria* bacteria are

Figure 23–30 Antigenic variation in trypanosomes. (A) There are about 1000 distinct Vsg genes in *Trypanosoma brucei*, and they are expressed one at a time from approximately 20 expression sites in the genome. To be expressed, an inactive gene is copied and the copy is moved into an expression site through DNA recombination. Each Vsg gene encodes a different surface protein (antigen). These 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 will 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 continues.

also extremely adept at taking up DNA from their environment by natural transformation and incorporating it into their genomes, further contributing to their extraordinary variability. The end result of this considerable variation is a plethora of different surface compositions with which to bewilder the host adaptive immune system. It is therefore not surprising that it has been difficult to develop an effective vaccine against *Neisseria* infections, although there are now several that protect against *Neisseria meningitidis*, a common cause of fatal meningitis.

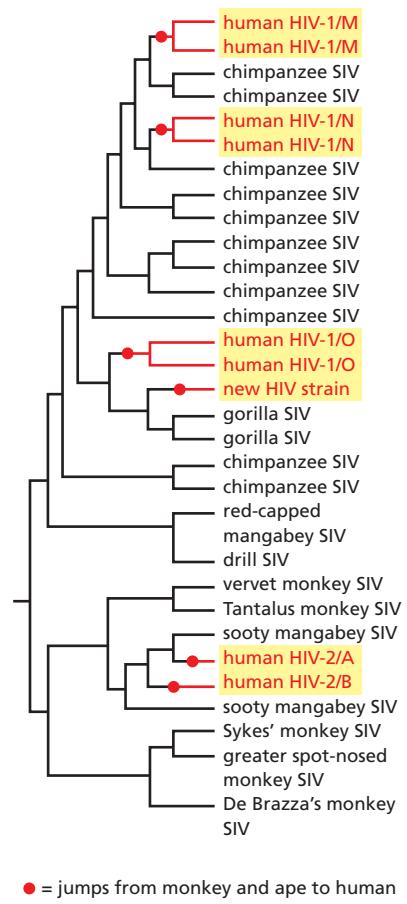
Error-Prone Replication Dominates Viral Evolution

In contrast to the DNA rearrangements in bacteria and parasites, viruses rely on an error-prone replication mechanism for antigenic variation. Retroviral genomes, for example, acquire on average one point mutation every replication cycle, because the viral reverse transcriptase (see Figure 5–62) needed to produce DNA from the viral RNA genome lacks the proofreading activity of DNA polymerases. A typical, untreated HIV infection may eventually produce HIV genomes with every possible point mutation. By a process of mutation and selection within each host, most viruses 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 23–17). 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. 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 (see Figure 7–62), can differ by as much as 30% from one HIV isolate to another. This extraordinary genomic plasticity greatly complicates attempts to develop vaccines against HIV. It has also led to the rapid emergence of new HIV strains. Nucleotide sequence comparisons between various strains of HIV and the very similar simian immunodeficiency virus (SIV) isolated from a variety of monkey species suggest that the most virulent type of HIV, HIV-1, may have jumped from primates to humans multiple independent times, starting as long ago as 1908 (Figure 23–31).

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 RNA strands of the two strains can reassort to form a new type of influenza virus. In normal years, influenza is a mild disease in healthy adults, although it can be life-threatening in the very young and very old. Different influenza strains infect fowl such as ducks and chickens, but only a subset of these strains can infect humans, and transmission from fowl to humans is rare. In 1918, however, a particularly virulent variant of avian influenza crossed the species barrier to infect humans, triggering the catastrophic pandemic of 1918 called the Spanish flu, which killed 20–50 million people worldwide. Subsequent influenza pandemics have been triggered by genome reassortment, in which a new RNA segment from an avian form of the virus replaced one or more of the viral RNA segments from the human form (Figure 23–32). In 2009, a new H1N1 swine virus emerged that derived genes from pig, avian, and human influenza viruses. Such recombination events allowed 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 a new recombinant strain of virus, and the infection rate drops to a steady-state level. 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.

Drug-Resistant Pathogens Are a Growing Problem

The development of drugs that cure rather than prevent infections has had a major impact on human health. **Antibiotics**, which are either bactericidal (they kill bacteria) or bacteriostatic (they inhibit bacterial growth without killing), are the most successful class of such drugs. Penicillin was one of the first antibiotics



● = jumps from monkey and ape to human

Figure 23–31 Diversification of HIV-1, HIV-2, and related strains of SIV. HIV comprises different viral families, all descended from SIV (simian immunodeficiency virus). On three separate occasions, SIV was passed from a chimpanzee to a human, resulting in three HIV-1 groups: major (M), outlier (O), and non-M non-O (N). The HIV-1 M group is the most common and is primarily responsible for the global AIDS epidemic. On two separate occasions, SIV was passed from a sooty mangabey monkey to a human, resulting in the two HIV-2 groups. In 2009, a new strain of HIV was discovered that appears to have resulted from SIV passage from a gorilla to a human.

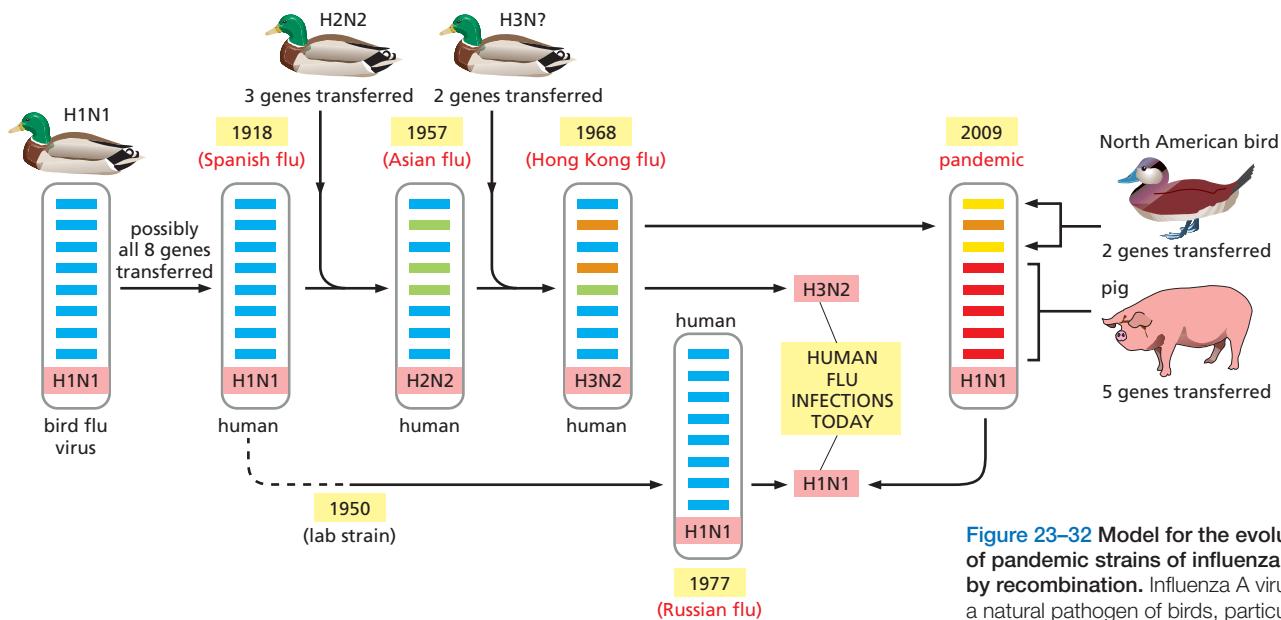


Figure 23–32 Model for the evolution of pandemic strains of influenza virus by recombination. Influenza A virus is a natural pathogen of birds, particularly waterfowl, and it is always present in wild bird populations. In 1918, a particularly virulent form of the virus crossed the species barrier from birds to humans and caused a devastating worldwide epidemic. This strain was designated H1N1, referring to the specific forms of its main antigens, hemagglutinin (H) and neuraminidase (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 three genes 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 two genes were 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. In 2009, a new H1N1 swine virus emerged that had derived five genes from pig influenza viruses, two from avian influenza viruses, and one from a human influenza virus. As indicated, most human influenza today is caused by H1N1 and H3N2 strains.

used to treat infections in humans, just in time to prevent tens of thousands of deaths from infected battlefield wounds in World War II. Because bacteria (see Figure 1–17) are not closely related evolutionarily to the eukaryotes they infect, much of their basic machinery for DNA replication and transcription, RNA translation, and metabolism differs from that of their host. These differences enable us to develop antibacterial drugs that exhibit *selective toxicity*, in that they specifically inhibit these processes in bacteria without disrupting them in the host. Most of the antibiotics that we use to treat bacterial infections are small molecules that inhibit macromolecular synthesis in bacteria by targeting bacterial enzymes that either are distinct from their eukaryotic counterparts or are involved in pathways such as cell wall biosynthesis that are absent in animals (Figure 23–33 and see Table 6–4).

However, bacteria continuously evolve and strains resistant to antibiotics rapidly develop, often within a few years of the introduction of a new drug. 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 for any one of them.

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 modifies or destroys the drug; or (3) it can prevent the drug's access to the drug target by, for example, actively pumping the drug out of the pathogen (Figure 23–34).

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 by horizontal gene transfer. They may even spread between pathogens of different species. The highly effective but expensive antibiotic *vancomycin*, for example, is used as a treatment of last resort for many severe, hospital-acquired, Gram-positive bacterial infections that are resistant to most other known antibiotics. Vancomycin prevents one step in bacterial cell wall synthesis—the cross-linking of peptidoglycan chains in the bacterial cell wall (see Figure 23–3B). Resistance can arise if the bacterium synthesizes a cell wall using different subunits that do not bind vancomycin. The most effective form of vancomycin resistance depends on the acquisition of a transposon (see Figure 5–60) containing seven genes, the products of which work together to sense the

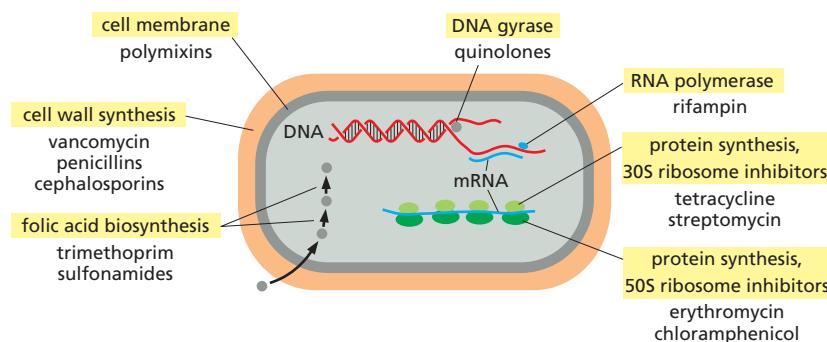


Figure 23–33 Antibiotic targets. Although there are many antibiotics in clinical use, they have a narrow range of targets, which are highlighted in yellow. A few representative antibiotics in each class are listed. Nearly all antibiotics used to treat human infections fall into one of these categories. The vast majority inhibit either bacterial protein synthesis or bacterial cell wall synthesis.

presence of vancomycin, shut down the normal pathway for bacterial cell wall synthesis, and produce a different type of cell wall.

Drug-resistance genes acquired by horizontal transfer frequently come from environmental microbial reservoirs. Nearly all antibiotics used to treat bacterial infections today are based on 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 Gram-positive bacteria of the genus *Streptomyces*, which reside in the soil. It is believed that microorganisms produce antimicrobial compounds, many of which have probably existed on Earth for hundreds of millions of years, as weapons in their competition with other microorganisms in the environment. Surveys of bacteria taken from soil samples that have never been exposed to antibiotic drugs used in modern medicine 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 antibiotic treatments, they can apparently draw upon this immense 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 take antibiotics for symptoms that are typically caused by viruses (flu-like illnesses, colds, sore throats, and earaches) and these drugs have no effects. Persistent and chronic misuse of antibiotics can eventually result in antibiotic-resistant normal flora, which can then transfer the resistance to pathogens. Antibiotics are also misused in agriculture, where they are commonly employed as food additives to promote the growth and 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.

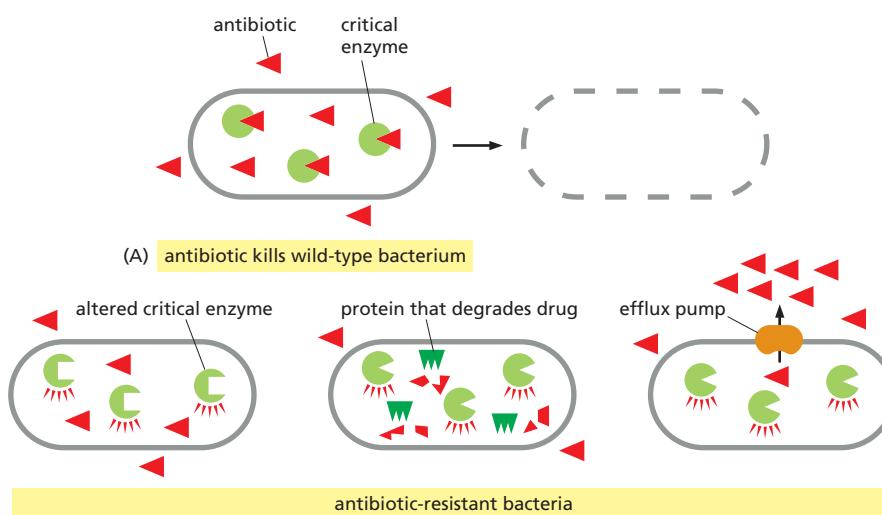


Figure 23–34 Three general mechanisms of antibiotic resistance. (A) A nonresistant wild-type bacterial cell bathed in a drug (red triangles) that binds to and inhibits an essential enzyme (light green) will be killed due to enzyme inhibition. (B) A bacterium that has altered the drug's target enzyme so that the drug no longer binds to the enzyme will survive and proliferate. In many cases, a single point mutation in the gene encoding the target protein can generate resistance. (C) A bacterium that expresses an enzyme (dark green) that either degrades or covalently modifies the drug will survive and proliferate. Some resistant bacteria, for example, make β -lactamase enzymes, which cleave penicillin and similar molecules. (D) A bacterium that expresses or up-regulates 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) will survive and proliferate. Some efflux pumps, such as the TetR efflux pump, are specific for a single drug (in this case, tetracycline), whereas others, called multidrug resistance (MDR) efflux pumps, are capable of exporting a wide variety of structurally dissimilar drugs. Upregulation of an MDR pump can render a bacterium resistant to a very large number of different antibiotics in a single step.

Because the acquisition of drug resistance is almost inevitable, it is crucial that we continue to develop innovative treatments for infectious diseases. We must also take additional measures to delay the onset of drug resistance.

Summary

All pathogens share the ability to interact with host cells in diverse ways that promote the replication and spread of the pathogen. 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 body surfaces in direct contact with the environment. Intracellular pathogens, including all viruses and many bacteria and protozoa, invade host cells 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 for the invasion. 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 survival and replication, frequently altering host membrane traffic and exploiting the host-cell cytoskeleton for intracellular movement. 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.

PROBLEMS

Which statements are true? Explain why or why not.

23–1 Our adult bodies harbor about 10 times more microbial cells than human cells.

23–2 The microbiomes from healthy humans are all very similar.

23–3 Pathogens must enter host cells to cause disease.

23–4 Viruses replicate their genomes in the nucleus of the host cell.

23–5 You should not take antibiotics for diseases caused by viruses.

Discuss the following problems.

23–6 In order to survive and multiply, a successful pathogen must accomplish five tasks. Name them.

23–7 *Clostridium difficile* infection is the leading cause of hospital-associated gastrointestinal illness. It is typically treated with a course of antibiotics, but the infection recurs in about 20% of cases. *C. difficile* infections are difficult to eradicate because the bacteria exist in two forms: a replicating, toxin-producing form and a spore form that is resistant to antibiotics. Fecal microbiota transplantation—the transfer of normal gut microbiota from a healthy individual—can resolve >90% of recurrent infections, a much better cure rate than further antibiotic treatment alone. Why do you suppose microbiota transplantation is so effective?

23–8 What are the three general mechanisms for horizontal gene transfer?

WHAT WE DON'T KNOW

- What are the genetic and molecular features that differ between pathogens and members of the normal human microbiota? How can our immune system distinguish between the two?
- To what extent are common host-cell biological pathways and molecules hijacked by diverse microbes?
- Can host-cell defense molecules be mobilized by drugs to fight infection?

23–9 The Gram-negative bacterium *Yersinia pestis*, the causative agent of the plague, is extremely virulent. Upon infection, *Y. pestis* injects a set of effector proteins into macrophages that suppresses their phagocytic behavior and also interferes with their innate immune responses. One of the effector proteins, YopJ, acetylates serines and threonines on various MAP kinases, including the MAP kinase kinase kinase TAK1, which controls a key signaling step in the innate immune response pathway. To determine how YopJ interferes with TAK1, you transfect human cells with active YopJ (*YopJ*^{WT}) or inactive YopJ (*YopJ*^{CA}) and with FLAG-tagged active TAK1 (*TAK1*^{WT}) or inactive TAK1 (*TAK1*^{K63W}), and assay for total TAK1 and for phosphorylated TAK1, using antibodies against the FLAG tag or against phosphorylated TAK1 (Figure Q23–1). How does YopJ block the TAK1 signaling pathway? How do you suppose the serine/threonine acetylase activity of YopJ might interfere with TAK1 activation?

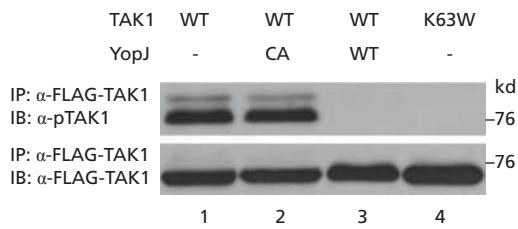


Figure Q23–1 Effects of YopJ on TAK1 phosphorylation (Problem 23–9). TAK1 was immunoprecipitated (IP) using antibodies against the FLAG tag (α -FLAG-TAK1). Total TAK1 in the immunoprecipitation was assayed by immunoblot (IB) using the same antibody. Phosphorylated TAK1 was assayed by IB using antibodies specific for phospho-TAK1 (α -pTAK1). A scale of protein molecular mass is shown at right in kilodaltons. (From N. Paquette et al., Proc. Natl Acad. Sci. USA 109:12710–12715, 2012. With permission from National Academy of Sciences.)

23–10 The intracellular bacterial pathogen *Salmonella typhimurium*, which causes gastroenteritis, injects effector proteins to promote its invasion into nonphagocytic host cells by the trigger mechanism. *S. typhimurium* first stimulates membrane ruffling to promote invasion, and then suppresses membrane ruffling once invasion is complete. This behavior is mediated in part by injection of two effector proteins: SopE, which promotes membrane ruffling and invasion, and SptP, which blocks the effects of SopE. Both effector proteins target the monomeric GTPase, Rac, which in its active form promotes membrane ruffling. How do you suppose SopE and SptP affect Rac activity? How do you suppose the effects of SopE and SptP are staggered in time if they are injected simultaneously?

23–11 John Snow is widely regarded as the father of modern epidemiology. Most famously, he investigated an outbreak of cholera in London in 1854 that killed more than 600 victims before it was finished. Snow recorded where the victims lived, and plotted the data on a map, along with the locations of the water pumps that served as the source of water for the public (Figure Q23–2). He concluded that the disease was most likely spread in the water, although he could find nothing suspicious-looking in it. His conclusion ran counter to the then-current belief that cholera was from “miasmas” in bad air. Very few believed his theory during the next 50 years, with the “bad air” theory persisting until at least 1901. What do you suppose Snow saw in the data that led him to his conclusion? Why do you think most scientists remained skeptical for so long?

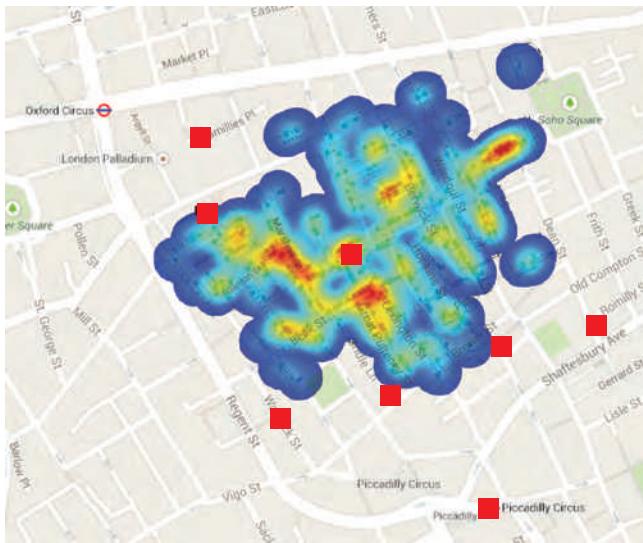


Figure Q23–2 A map of where the victims of the 1854 cholera outbreak lived, superimposed on a modern Google map of the area (Problem 23–11). The locations of the victims' houses are indicated in a gradient of colors from blue (indicating few cases) to orange (indicating many cases). Public water pumps are shown as red squares.

23–12 Influenza epidemics account for 250,000 to 500,000 deaths globally each year. These epidemics are markedly seasonal, occurring in temperate climates in the northern and southern hemispheres during their respec-

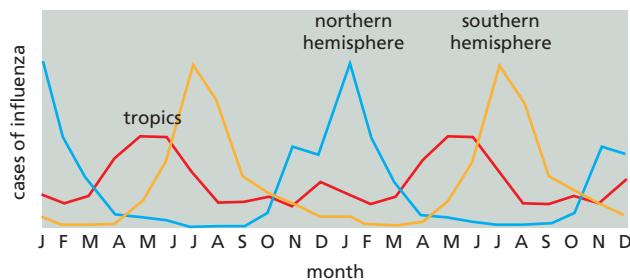


Figure Q23–3 Seasonal patterns of influenza epidemics (Problem 23–12). Cases of influenza at different times of the year are shown for the northern hemisphere (blue), the southern hemisphere (orange), and the tropics (red).

tive winters. By contrast, in the tropics, there is significant influenza activity year round, with a peak in the rainy season (Figure Q23–3). Can you suggest some possible explanations for the patterns of influenza epidemics in temperate zones and the tropics?

23–13 Several negative-strand viruses carry their genome as a set of discrete RNA segments. Examples include influenza virus (eight segments), Rift Valley fever virus (three segments), Hantavirus (three segments), and Lassa virus (two segments), to name a few. Why does segmentation of the genome provide a strong evolutionary advantage for these viruses?

23–14 Avian influenza viruses readily infect birds, but are transmitted to humans very rarely. Similarly, human influenza viruses spread readily to other humans, but have never been detected in birds. The key to this specificity lies in the viral capsid protein, hemagglutinin, which binds to sialic acid residues on cell-surface glycoproteins, triggering virus entry into the cell (Movie 23.8). Hemagglutinin on human viruses recognizes sialic acid in a 2-6 linkage with galactose, whereas avian hemagglutinin recognizes sialic acid in a 2-3 linkage with galactose. Humans make carbohydrate chains that have only the 2-6 linkage between sialic acid and galactose; birds make only the 2-3 linkage; but pigs make carbohydrate chains with both linkages. How does this situation make pigs ideal hosts for generating new strains of human influenza viruses?

23–15 The majority of antibiotics used in the clinic are made as natural products by bacteria. Why do you suppose bacteria make the very agents we use to kill them?

23–16 In the early days of penicillin research, it was discovered that bacteria in the air could destroy the penicillin, a big problem for large-scale production of the drug. How do you suppose this occurs?

23–17 When the Oxford team of Ernst Chain and Norman Heatley had laboriously collected their first two grams of penicillin (probably no more than 2% pure!), Chain injected two normal mice with 1 g each of this preparation, and waited to see what would happen. The mice survived with no apparent ill effects. Their boss, Howard Florey, was furious at what he saw as a waste of good antibiotic. Why was this experiment important?

REFERENCES

General

- Cossart P, Boquet P, Normark S & Rappuoli R (eds) (2005) *Cellular Microbiology*, 2nd ed. Washington, DC: ASM Press.
- Engleberg NC, DiRita V & Dermody T (2012) Schaechter's Mechanisms of Microbial Disease, 5th ed. Philadelphia, PA: Lippincott, Williams & Wilkins.
- Norkin LA (2010) *Virology: Molecular Biology and Pathogenesis*. Washington, DC: ASM Press.
- Wilson BA, Salyers AA, Whitt DD & Winkler ME (2011) *Bacterial Pathogenesis: A Molecular Approach*, 3rd ed. Washington, DC: ASM Press.
- Introduction to Pathogens and the Human Microbiota**
- Aly AS, Vaughan AM & Kappe SH (2009) Malaria parasite development in the mosquito and infection of the mammalian host. *Annu. Rev. Microbiol.* 63, 195–221.
- Baltimore D (1971) Expression of animal virus genomes. *Bacteriol. Rev.* 35, 235–241.
- Clemente JC, Ursell LK, Parfrey LW & Knight R (2012) The impact of the gut microbiota on human health: an integrative view. *Cell* 148, 1258–1270.
- Crick FH & Watson JD (1956) Structure of small viruses. *Nature* 177, 473–475.
- Fauci A & Morens DM (2012) The perpetual challenge of infectious diseases. *N. Engl. J. Med.* 366, 454–461.
- Frost LS, Leplae R, Summers AO & Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* 3, 722–732.
- Galán 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.
- Nelson EJ, Harris JB, Morris JG Jr et al. (2009) Cholera transmission: the host, pathogen and bacteriophage dynamic. *Nat. Rev. Microbiol.* 7, 693–702.
- Pflughoeft KJ & Versalovic J (2012) Human microbiome in health and disease. *Annu. Rev. Pathol.* 7, 99–122.
- Polk DB & Peek RM Jr (2010) *Helicobacter pylori*: gastric cancer and beyond. *Nat. Rev. Cancer* 10, 403–414.
- Poulin R & Morand S (2000) The diversity of parasites. *Q. Rev. Biol.* 75, 277–293.
- Rappleye CA & Goldman WE (2006) Defining virulence genes in the dimorphic fungi. *Annu. Rev. Microbiol.* 60, 281–303.
- Thomas CM & Nielsen KM (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* 3, 711–721.
- Votteler J & Sundquist WI (2013) Virus budding and the ESCRT pathway. *Cell Host Microbe* 14, 232–241.
- Young JAT & Collier RJ (2007) Anthrax toxin: receptor binding, internalization, pore formation, and translocation. *Annu. Rev. Biochem.* 76, 243–265.
- Cell Biology of Infection**
- Alix E, Mukherjee S & Roy CR (2011) Subversion of membrane transport pathways by vacuolar pathogens. *J. Cell Biol.* 195, 943–952.
- Beiting DP & Roos DS (2011) A systems biological view of intracellular pathogens. *Immunol. Rev.* 240, 117–128.
- Brandenburg B & Zhuang X (2007) Virus trafficking – learning from single-virus tracking. *Nat. Rev. Microbiol.* 5, 197–208.
- Cossart P & Sansonetti PJ (2004) Bacterial invasion: the paradigms of enteroinvasive pathogens. *Science* 304, 242–248.
- Daugherty MD & Malik HS (2012) Rules of engagement: molecular insights from host-virus arms races. *Annu. Rev. Genet.* 46, 677–700.
- Davies J & Davies D (2010) Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* 74, 417–433.
- Dimitrov DS (2004) Virus entry: molecular mechanisms and biomedical applications. *Nat. Rev. Microbiol.* 2, 109–122.
- Duffy S, Shackelton LA & Holmes EC (2008) Rates of evolutionary change in viruses: patterns and determinants. *Nat. Rev. Genet.* 9, 267–276.
- Forsberg KJ, Reyes A, Wang B et al. (2012) The shared antibiotic resistome of soil bacteria and human pathogens. *Science* 337, 1107–1111.
- 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 DE, Siliciano RF & Jacobs WR Jr (2012) Outwitting evolution: fighting drug-resistant TB, malaria, and HIV. *Cell* 148, 1271–1283.
- Haglund CM & Welch MD (2011) Pathogens and polymers: microbe-host interactions illuminate the cytoskeleton. *J. Cell Biol.* 195, 7–17.
- Ham H, Sreelatha A & Orth K (2011) Manipulation of host membranes by bacterial effectors. *Nat. Rev. Microbiol.* 9, 635–646.
- Hayward RD, Leong JM, Koronakis V & Campellone KG (2006) Exploiting pathogenic *Escherichia coli* to model transmembrane receptor signalling. *Nat. Rev. Microbiol.* 4, 358–370.
- 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.
- Lusso P (2006) HIV and the chemokine system: 10 years later. *EMBO J.* 25, 447–456.
- Medina RA & García-Sastre A (2011) Influenza A viruses: new research developments. *Nat. Rev. Microbiol.* 9, 590–603.
- Mengaud J, Ohayon H, Gounon P et al. (1996) E-cadherin is the receptor for internalin, a surface protein required for entry of *L. monocytogenes* into epithelial cells. *Cell* 84, 923–932.
- Mercer J, Schelhaas M & Helenius A (2010) Virus entry by endocytosis. *Annu. Rev. Biochem.* 79, 803–833.
- Miller S & Krijnse-Locker J (2008) Modification of intracellular membrane structures for virus replication. *Nat. Rev. Microbiol.* 6, 363–374.
- Mullins JI & Jensen MA (2006) Evolutionary dynamics of HIV-1 and the control of AIDS. *Curr. Top. Microbiol. Immunol.* 299, 171–192.
- 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-Cerdá J & Cossart P (2006) Bacterial adhesion and entry into host cells. *Cell* 124, 715–727.
- Ray K, Marteyn B, Sansonetti PJ & Tang CM (2009) Life on the inside: the intracellular lifestyle of cytosolic bacteria. *Nat. Rev. Microbiol.* 7, 333–340.
- Sibley LD (2011) Invasion and intracellular survival by protozoan parasites. *Immunol. Rev.* 240, 72–91.
- 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.
- Vink C, Rudenko G & Seifert HS (2012) Microbial antigenic variation mediated by homologous DNA recombination. *FEMS Microbiol. Rev.* 36, 917–948.
- Walsh D & Mohr I (2011) Viral subversion of the host protein synthesis machinery. *Nat. Rev. Microbiol.* 9, 860–875.
- Welch MD & Way M (2013) Arp2/3-mediated actin-based motility: a tail of pathogen abuse. *Cell Host Microbe* 14, 242–255.

The Innate and Adaptive Immune Systems

CHAPTER 24

As we discussed in Chapter 23, all living organisms serve as hosts for other species, usually in relationships that are benign or even mutually helpful. But all organisms, and all cells in a multicellular organism, need to defend themselves against infection by harmful invaders, collectively called **pathogens**, which can be microbes (bacteria, viruses, or fungi), or larger parasites. Even bacteria defend themselves against viruses, using intracellular proteins called *restriction factors*, which block viral propagation. Invertebrates use a variety of defense strategies, including protective barriers, toxic molecules, restriction factors, and phagocytic cells that ingest and destroy invading pathogens. Vertebrates, too, depend on such *innate immune responses*, but they can also harness more sophisticated and specific mechanisms, called *adaptive immune responses*. The innate responses occur first, calling the adaptive immune responses into play if required, in which case, both types of responses work together to eliminate the pathogen (Figure 24–1).

Whereas innate immune responses are general defense reactions that can involve almost any cell type in an organism, the adaptive immune responses are highly specific to the particular pathogen that induced them and depend on a class of white blood cells (leukocytes) called *lymphocytes*. There are two major classes of lymphocytes that mount adaptive immune responses—B lymphocytes (*B cells*), which secrete *antibodies* that bind specifically to the pathogen, and T lymphocytes (*T cells*), which can either directly kill cells infected with the pathogen or produce secreted or cell-surface signal proteins that stimulate other host cells to help eliminate the pathogen (Figure 24–2). Unlike innate immune responses, which are generally short-lasting, the adaptive responses provide long-lasting protection: a person who recovers from measles or is vaccinated against it, 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.

Both the innate and adaptive immune systems have evolved sensing mechanisms that enable them to recognize harmful invaders (pathogens) and distinguish them from both the host's own cells and molecules and harmless or beneficial foreign organisms and their molecules. The innate system relies on sensor proteins that recognize particular types or patterns of molecules that are common to pathogens but are absent or sequestered in the host. The adaptive system, by contrast, uses unique genetic mechanisms to produce a virtually limitless diversity of related proteins—receptors on T and B cells and secreted antibodies—that, between them, can bind almost any foreign molecule. This remarkable strategy enables the adaptive immune system to react specifically against any pathogen, even if the animal never encountered it before. But, it also requires that the system learn not to react against self molecules or harmless foreign ones; if these learning mechanisms fail, harmful autoimmune or allergic responses result.

In this chapter, we focus on vertebrate immune responses and the features that distinguish them from other kinds of cell responses. We begin with innate immune defenses and then discuss the highly specialized properties of the adaptive immune system.

IN THIS CHAPTER

THE INNATE IMMUNE SYSTEM

OVERVIEW OF THE ADAPTIVE IMMUNE SYSTEM

B CELLS AND IMMUNOGLOBULINS

T CELLS AND MHC PROTEINS

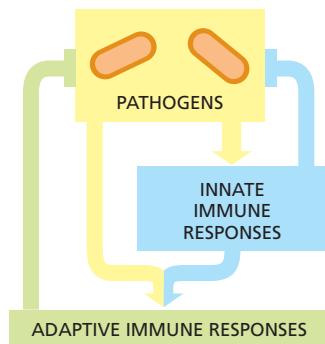


Figure 24–1 Innate and adaptive immune responses. Innate immune responses are activated directly by pathogens and defend all multicellular organisms against infection. In vertebrates, pathogens, together with the innate immune responses they activate, also stimulate adaptive immune responses, which then work together with innate immune responses to help fight the infection.

THE INNATE IMMUNE SYSTEM

Adaptive immune responses are slow to develop when a vertebrate first encounters a new pathogen. This is because the specific B cells and T cells that can respond to a particular pathogen are initially few in number and must be stimulated to proliferate and differentiate before they can mount effective adaptive immune responses, which can take days. By contrast, a single bacterium that divides every hour can generate almost twenty million progeny in a single day, producing a full-blown infection. Vertebrates, therefore, rely on their **innate immune system** to defend them against infection during the first critical hours and days of exposure to a new pathogen. Plants and invertebrates lack adaptive immune systems and therefore rely entirely on innate immunity for protection against pathogens.

In this section, we consider some of the strategies the innate immune system uses to recognize pathogens and to provide a first line of defense against them.

Epithelial Surfaces Serve as Barriers to Infection

In vertebrates, the first encounters with infectious organisms are typically at the epithelial surfaces that form the skin and line the respiratory, digestive, urinary, and reproductive tracts. These epithelia provide both physical and chemical barriers to invasion by pathogens: tight junctions between epithelial cells bar entry between the cells, and a variety of substances secreted by the cells discourage the attachment and entry of pathogens. The keratinized epithelial cells of the skin, for example, form a thick physical barrier, and the sebaceous glands in the skin secrete fatty acids and lactic acid, which inhibit bacterial growth. In addition, epithelial cells in all tissues, including those in plants and invertebrates, secrete antimicrobial molecules called **defensins**. Defensins are positively charged, amphipathic peptides that bind to and disrupt the membranes of many pathogens, including enveloped viruses, bacteria, fungi, and parasites.

The epithelial cells that line internal organs such as the respiratory and digestive tracts also secrete slimy mucus, which sticks to the epithelial surface and makes it difficult for pathogens to adhere. The beating of cilia on the surface of the epithelial cells lining the respiratory tract and the peristaltic action of the intestine also discourage the adherence of pathogens. Moreover, as we discuss in Chapter 23, healthy skin and gut are populated by enormous numbers of harmless (and often helpful) *commensal* microbes, collectively called the *normal flora*, which compete for nutrients with pathogens; some also produce antimicrobial peptides that actively inhibit pathogen proliferation.

Pattern Recognition Receptors (PRRs) Recognize Conserved Features of Pathogens

Pathogens do occasionally breach the epithelial barricades, in which case underlying, nonepithelial cells of the innate immune system provide the next line of defense. These cells sense the presence of pathogens largely through the use of receptor proteins that recognize microbe-associated molecules that either are not present or are sequestered in the host organism. Because these microbial molecules often occur in repeating patterns, they are called **pathogen-associated molecular patterns (PAMPs)**, even though they are not unique to microbes that can cause disease. PAMPs are present in various microbial molecules, including nucleic acids, lipids, polysaccharides, and proteins.

The special receptor proteins that recognize PAMPs are called **pattern recognition receptors (PRRs)**. Some PRRs are transmembrane proteins on the surface of many types of host cells, where they recognize extracellular pathogens; on professional phagocytic cells (phagocytes) such as *macrophages* and *neutrophils* (discussed in Chapter 22), they can mediate the uptake of the pathogens into phagosomes, which then fuse with lysosomes, where the pathogens are destroyed. Other PRRs are located intracellularly, where they can detect intracellular pathogens such as viruses; these PRRs are either free in the cytosol or associated with

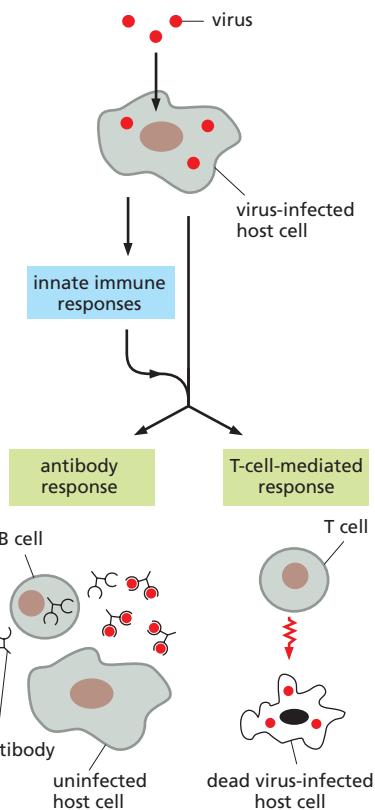


Figure 24–2 The two main classes of adaptive immune responses.

Lymphocytes carry out both classes of adaptive responses—shown here as responses to a viral infection. In one class, B cells secrete antibodies that specifically bind to and neutralize extracellular viruses, by preventing the viruses from infecting host cells. In the other, T cells mediate the response; in this example, they kill the virus-infected host cells. In both cases, innate immune responses help activate the adaptive immune responses through pathways that are not shown.

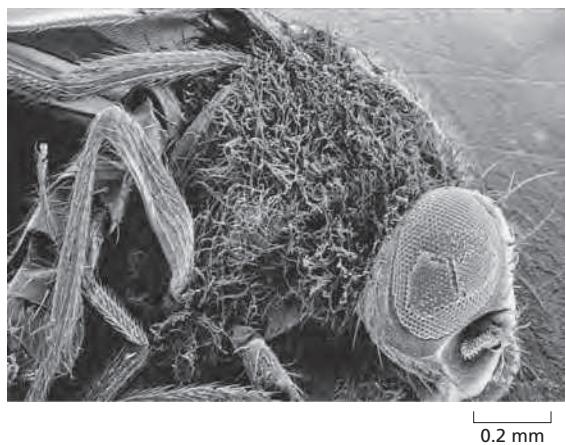


Figure 24–3 A scanning electron micrograph of a mutant fruit fly that died from a fungal infection. The fly is covered with fungal hyphae, as it lacked a Toll receptor, which helps protect *Drosophila* from fungal infections. (From B. Lemaitre et al., *Cell* 86:973–983, 1996.)

the membranes of the endolysosomal system (discussed in Chapter 13). Still other PRRs are secreted and bind to the surface of extracellular pathogens, marking them for destruction by either phagocytes or blood proteins that are part of the *complement system* (discussed later).

There Are Multiple Classes of PRRs

The first PRR identified was the *Toll receptor* in *Drosophila*, which was well-known for its role in fly development (see Figure 21–17). It was later discovered to be also required for the production of antimicrobial peptides that protect the fly against fungal infections (Figure 24–3). Toll is a transmembrane glycoprotein with a large extracellular domain that contains a series of leucine-rich repeats. Soon it was discovered that both plants and animals have a variety of **Toll-like receptors (TLRs)** that function as PRRs in innate immune responses against various pathogens. Mammals make at least 10 different TLRs, each recognizing distinct ligands: TLR3, for example, recognizes double-stranded viral RNA in the endosomal lumen (Figure 24–4); TLR4 recognizes lipopolysaccharide (LPS) on the outer

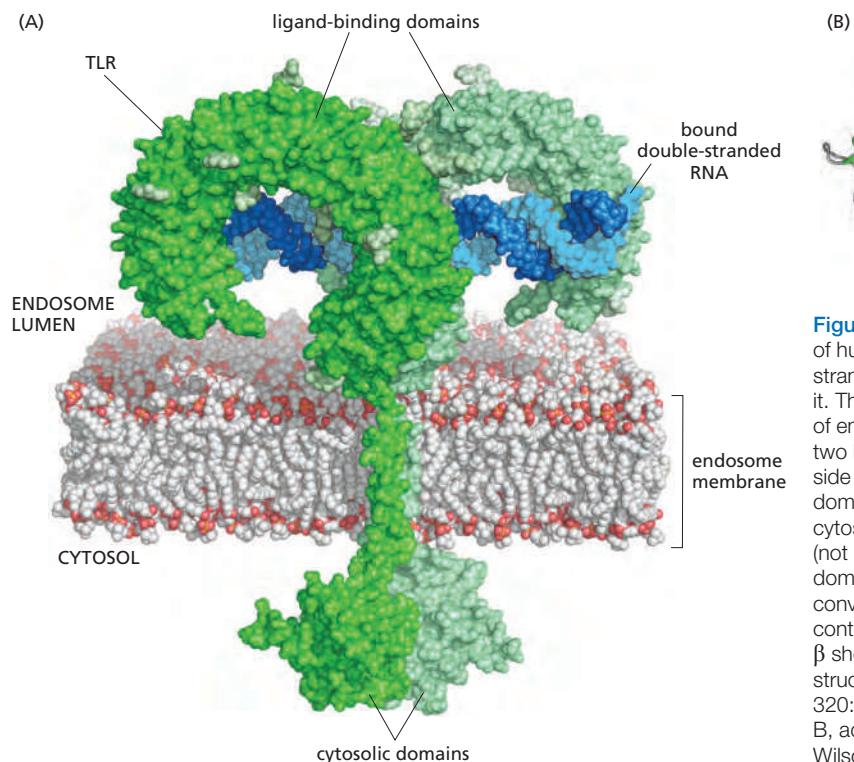


Figure 24–4 A Toll-like receptor. The structure of human TLR3 is shown (green), with a double-stranded RNA molecule (dsRNA, blue) bound to it. The receptor is a homodimer in the membrane of endosomes. The binding of dsRNA to the two horseshoe-shaped domains on the luminal side of the endosome brings the two cytosolic domains together, allowing adaptor proteins in the cytosol to assemble into a large signaling complex (not shown). (B) The crystal structure of a luminal domain of the receptor, which contains 23 conventional leucine-rich repeats, each of which contributes a β strand to the continuous β sheet (red) that lines the concave surface of the structure. (A, adapted from L. Liu et al., *Science* 320:379–381, 2008. With permission from AAAS; B, adapted from J. Choe, M.S. Kelker and I.A. Wilson, *Science* 309:581–585, 2006; PDB: 1ZIW.)

membrane of Gram-negative bacteria; TLR5 recognizes the protein that forms the bacterial flagellum; and TLR9 recognizes short, unmethylated sequences of bacterial, viral, or protozoan DNA, called CpG motifs, which are uncommon in vertebrate DNA.

In addition to TLRs, vertebrates use several other families of PRRs to detect pathogens. One is the large family of **NOD-like receptors (NLRs)**. Like TLRs, NLRs have leucine-rich repeat motifs, but they are exclusively cytoplasmic and recognize a distinct set of bacterial molecules. Individuals who are homozygous for a particular mutant allele of the NLR gene *NOD2* have a greatly increased risk of developing Crohn's disease, a chronic inflammatory disease of the small intestine, possibly triggered by a bacterial infection. Another class of PRRs consists of **RIG-like receptors (RLRs)**, which are members of the RNA helicase family of proteins. They are also exclusively cytoplasmic and detect viral pathogens. A fourth class of PRRs consists of **C-type lectin receptors (CLRs)**, which are trans-membrane cell-surface proteins that recognize carbohydrates (which is why they are called lectins) on various microbes. **Table 24-1** summarizes some PRRs and their ligands and locations in cells. Collectively, these and other PRRs act as an alarm system to alert the innate and adaptive immune systems that an infection is brewing (**Movie 24.1**).

When a cell-surface or intracellular PRR binds a PAMP, it stimulates the cell to secrete a variety of cytokines and other extracellular signal molecules. Some of these inhibit viral replication, but most induce a local inflammatory response that helps eliminate the pathogen, as we now discuss.

Activated PRRs Trigger an Inflammatory Response at Sites of Infection

When a pathogen invades a tissue, it activates PRRs on or in various cells of the innate immune system, resulting in an **inflammatory response** at the site of infection. The inflammatory response depends on changes in local blood vessels and is characterized clinically by local pain, redness, heat, and swelling. 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. At the same time, the endothelial cells lining the local blood vessels are stimulated to express cell adhesion proteins, which promote the attachment and escape of white blood cells or *leukocytes* (see Figure 19–29B), adding to the local swelling; initially neutrophils escape, followed later by lymphocytes and monocytes (the blood-borne precursors of macrophages).

TABLE 24-1 Some Pattern Recognition Receptors (PRRs)

Receptor	Location	Ligand	Origin of ligand
<i>Toll-like receptors (TLRs)</i>			
TLR3	Endolysosomal system	Double-stranded RNA	Viruses
TLR4	Plasma membrane	Bacterial lipopolysaccharide (LPS); viral coat proteins	Bacteria; viruses
TLR5	Plasma membrane	Flagellin	Bacteria
TLR9	Endolysosomal system	Unmethylated CpG DNA	Bacteria, viruses, protozoa
<i>NOD-like receptors (NLRs)</i>			
NOD2	Cytoplasm	Degradation products of peptidoglycans	Bacteria
<i>Retinoic acid-inducible gene 1-like receptors (RLRs)</i>			
RLR1	Cytoplasm	Double-stranded RNA	Viruses
<i>C-type lectin receptors (CLRs)</i>			
Dectin1	Plasma membrane	β-Glucan	Fungi

The activation of PRRs results in the production of a large variety of extracellular signal molecules that mediate the inflammatory response at the site of an infection. These include both lipid signal molecules, such as prostaglandins, and protein (or peptide) signal molecules called **cytokines**. Some of the most important **pro-inflammatory cytokines** are *tumor necrosis factor- α* (*TNF α*), *interferon- γ* (*IFN γ*), a variety of *chemokines* (which recruit leukocytes), and various *interleukins* (*ILs*) that we discuss later, including IL1, IL6, IL12, and IL17. In addition, a secreted PRR (mannose-binding lectin) activates the complement system when the PRR binds to a pathogen; fragments of complement proteins released during complement activation stimulate an inflammatory response (discussed shortly; see Figure 24-7).

When activated by PAMPs, most cell-surface and intracellular PRRs stimulate the production of multiple pro-inflammatory cytokines by activating intracellular signaling pathways that switch on transcription regulators, including NF κ B, to induce the transcription of the relevant cytokine genes (see Figure 15-62). Some PRRs, however, can also stimulate pro-inflammatory cytokine production by a different mechanism: when activated, several cytoplasmic NLRs assemble with adaptor proteins and specific protease precursors of the caspase family (discussed in Chapter 18) to form **inflammasomes**, in which pro-inflammatory cytokines such as IL1 are cleaved from their inactive precursor proteins by activated caspases. These cytokines are then released from the cell by a poorly understood, unconventional secretion pathway. Inflammasomes closely resemble apoptosomes in their assembly and structure, but, in apoptosomes, procaspases are activated to initiate a proteolytic caspase cascade that leads to apoptosis (see Figure 18-7).

NLR-dependent inflammasome assembly can also be triggered in the absence of infection if cells are damaged or stressed. Such cells produce “danger signals,” such as altered or misplaced self molecules, which can activate the relevant NLRs: the arthritis caused by uric acid crystals formed in the joints of individuals with gout, who have abnormally high uric acid levels in their blood, is a painful example.

Phagocytic Cells Seek, Engulf, and Destroy Pathogens

In all animals, the recognition of a microbial invader is usually quickly followed by its engulfment by a phagocytic cell. Macrophages are long-lived phagocytes that reside in most vertebrate tissues; they are among the first cells to encounter invading microbes, whose PAMPs activate the macrophages to secrete pro-inflammatory signal molecules. Neutrophils are short-lived phagocytes that are abundant in blood but are not present in healthy tissues; they are rapidly recruited to sites of infection by various attractive molecules, including formylmethionine-containing peptides (which are released by microbes but are not made by mammalian cells), chemokines secreted by activated macrophages, and peptide fragments produced from cleaved, activated complement proteins. The recruited neutrophils contribute their own pro-inflammatory cytokines.

In addition to their PRRs, macrophages and neutrophils display a variety of cell-surface receptors that recognize fragments of complement proteins or antibodies bound to the surface of a pathogen. The binding of such a pathogen to these receptors leads to its phagocytosis (Figure 24-5) and an attack on the ingested pathogen once inside a phagolysosome. The phagocytes possess an impressive armory of weapons to kill the invader, including enzymes such as lysozyme and acid hydrolases that can degrade the pathogen’s cell wall. The cells assemble *NADPH oxidase complexes* on the phagolysosomal membrane, where the complexes catalyze the production of highly toxic oxygen-derived compounds, including superoxide (O_2^-), 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. Whereas macrophages generally survive this killing frenzy and live to kill again, neutrophils do not. Dead and dying neutrophils are a major component of the pus that forms in acute bacterially infected wounds; their half-life in the human bloodstream is only a few hours.



Figure 24-5 Antibody-activated phagocytosis. Electron micrograph of a neutrophil phagocytosing an antibody-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*. (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.)

If a pathogen is too large to be successfully phagocytosed (if it is a large parasite such as a worm, for example), a group of macrophages, neutrophils, or eosinophils (another type of leukocyte) will gather around the invader. They secrete defensins and other damaging agents and release the toxic products of the respiratory burst. This barrage is often sufficient to destroy the pathogen (Figure 24–6).

Complement Activation Targets Pathogens for Phagocytosis or Lysis

The blood and other extracellular fluids contain numerous proteins with antimicrobial activity, some of which are produced in response to an infection, while others are produced constitutively. The most important of these are components of the **complement system**, which consists of about thirty interacting soluble proteins that are mainly made continuously by the liver and are inactive until an infection or another trigger activates them. They were originally identified by their ability to amplify and “complement” the action of antibodies made by B cells, but some are also secreted PRRs, which directly recognize PAMPs on microbes.

The *early complement components* consist of three sets of proteins, 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 cleave and activate C3, which is the pivotal complement component (Figure 24–7); 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 protein cleavages liberate a biologically active small fragment that can attract neutrophils, plus a membrane-binding larger fragment. The binding of the large fragment to a cell membrane, usually the surface of a pathogen, helps stimulate the next reaction in the sequence. In this way, complement activation is largely kept confined to the cell surface where it began. In particular, the large fragment of C3, called C3b, binds covalently to the surface of the pathogen. Here, it recruits protein fragments produced by cleavage of other early complement components to form proteolytic complexes that catalyze the subsequent steps in the complement cascade. The early events in complement activation have diverse functions: C3b-binding receptors on phagocytic cells enhance the ability of these cells to phagocytose the pathogen, and similar receptors on B cells enhance the

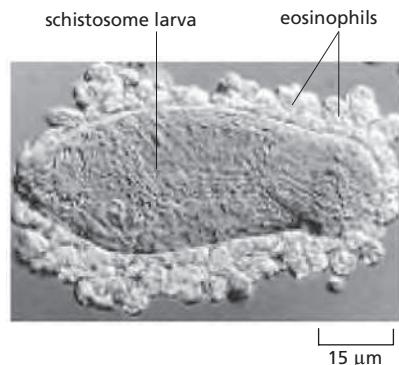


Figure 24–6 Eosinophils attacking a parasite. Phagocytes cannot ingest large parasites such as the schistosome larva shown here. When the larva is coated with antibody or complement components, however, eosinophils (and other leukocytes) can recognize it and collectively kill it by secreting a large variety of toxic molecules. (Courtesy of Anthony Butterworth.)

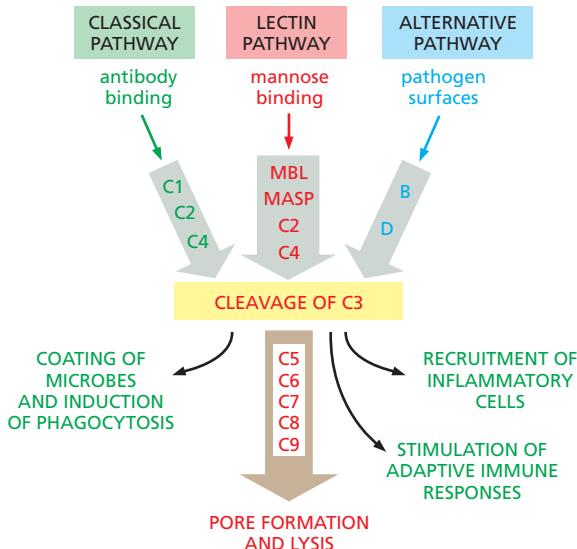


Figure 24–7 The principal stages in complement activation by the classical, lectin, and alternative pathways. In all three pathways, the reactions of complement activation usually take place on the surface of an invading microbe, such as a bacterium, and lead to the cleavage of C3 and the various consequences shown. As indicated, the complement proteins C1 to C9, mannose-binding lectin (MBL), MBL-associated serine protease (MASP), and factors B and D are the central components of the complement system. The early components are shown within gray arrows, while the late components are shown within a brown arrow. The functions of the protein fragments produced during complement activation are indicated by the black arrows. The various complement proteins that regulate the system are omitted.

ability of these cells to make antibodies against various microbial molecules on C3b-coated pathogens; the smaller fragment of C3 (called C3a), as well as small fragments of C4 and C5, act independently as diffusible signals to promote an inflammatory response by recruiting leukocytes to the site of infection.

As indicated in Figure 24–7, antibodies bound to the surface of a pathogen activate the *classical pathway*. *Mannose-binding lectin*, mentioned earlier, is a secreted PRR that initiates the *lectin pathway* of complement activation when it recognizes bacterial or fungal glycolipids and glycoproteins bearing terminal mannose and fucose sugars in a particular spatial conformation. These initial binding events in the classical and lectin pathways cause the recruitment and activation of the early complement components. Finally, molecules on the surface of pathogens will often directly activate the *alternative pathway*.

Host cells produce various plasma membrane molecules that prevent complement reactions 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 (see Figure 10–16). Because pathogens generally lack sialic acid, they are singled out for complement-mediated destruction, while host cells are spared. Some pathogens, including the bacterium *Neisseria gonorrhoeae* that causes the sexually transmitted disease gonorrhea, coat themselves with a layer of sialic acid to effectively hide from the complement 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*. These protein complexes assemble in the pathogen membrane near the site of C3 activation, forming aqueous pores through the membrane (Figure 24–8). For this reason, and because they perturb the structure of the lipid bilayer in their vicinity, they make the membrane leaky and can, in some cases, cause the microbe to lyse.

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, ensuring 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 complement components once the components have been activated by proteolytic cleavage. Second, many of the activated components in the cascade are unstable; unless they bind immediately to either the next component in the complement cascade or to a nearby membrane, they rapidly inactivate.

Virus-Infected Cells Take Drastic Measures to Prevent Viral Replication

Because host-cell ribosomes make a virus's proteins and host-cell lipids form the membranes of enveloped viruses, PAMPs are generally not present on the surface of viruses. Therefore, the only general way that a host cell PRR can recognize the presence of a virus is to detect unusual elements of the viral genome, such as the

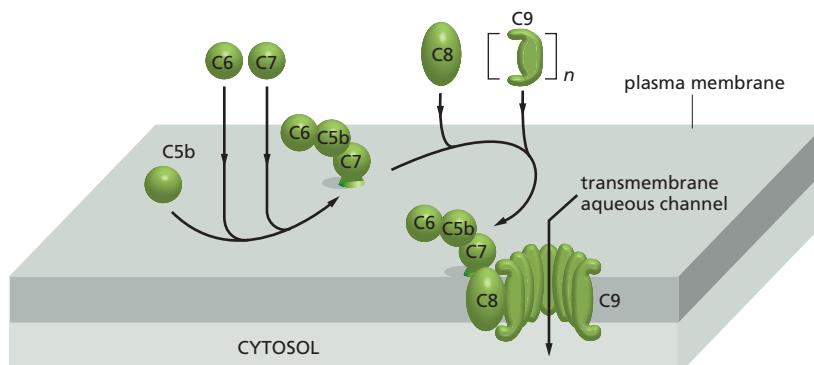


Figure 24–8 Assembly of the late complement components to form a membrane attack complex. The cleavage of the early complement components (shown within gray arrows in Figure 24–7) results in the formation of C3b-containing proteolytic complexes called C5 convertases (not shown). These then cleave the first of the late components, C5, to produce C5a and C5b. As illustrated, C5b rapidly assembles with C6 and C7 to form C567, which then binds firmly via C7 to the membrane. One molecule of C8 binds to the complex 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 membrane. 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 ring of C9 molecules forms a large transmembrane channel in the membrane.

double-stranded RNA (dsRNA) that is an intermediate in the life cycle of many viruses and is recognized by several PRRs including the Toll-like receptor TLR3; in addition, DNA virus genomes frequently contain significant amounts of the CpG motifs discussed earlier, which can be recognized by TLR9 (see Table 24-1, p. 1300).

Mammalian cells are particularly adept at recognizing the presence of dsRNA, which activates intracellular PRRs that induce the host cell to produce and secrete two antiviral cytokines—**interferon-α (IFNα)** and **interferon-β (IFNβ)**. These interferons are referred to as *type I interferons* to distinguish them from IFNγ, which is a type II interferon and has different functions, as we discuss later. Type I interferons act in both an autocrine fashion on the infected cells that produced it and a paracrine fashion on uninfected neighbors. They bind to a common cell-surface receptor, which activates the JAK-STAT intracellular signaling pathway (see Figure 15-56) to stimulate specific gene transcription and thereby the production of more than 300 proteins, including many cytokines, reflecting the complexity of the cell's acute response to a viral infection.

The production of type I interferons appears to be a general response of mammalian cells to a viral infection, and viral components other than dsRNA can trigger it. The type I interferons help block viral replication in multiple ways. They activate a latent ribonuclease that nonspecifically degrades single-stranded RNA. They also indirectly activate a protein kinase that phosphorylates and inactivates the protein synthesis initiation factor eIF2 (discussed in Chapter 6), thereby shutting down most protein synthesis in the infected host cell. Apparently, by destroying most of its own 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 an even more extreme step to prevent the virus from replicating: it kills itself by undergoing apoptosis, often with the help of immune killer cells, as we discuss next.

Natural Killer Cells Induce Virus-Infected Cells to Kill Themselves

Type I interferons also have less direct ways of blocking viral replication. One of these is to enhance the activity of **natural killer cells (NK cells)**, which are leukocytes related to T and B cells but are part of the innate immune system and are recruited early to sites of inflammation. Like *cytotoxic T cells* of the adaptive immune system (discussed later), NK cells destroy virus-infected cells by inducing the infected cells to kill themselves by undergoing apoptosis (discussed in Chapter 18). We consider how killer cells induce apoptosis later, when we discuss how cytotoxic T cells do it (see Figure 24-43). Although they kill in the same way, the means by which cytotoxic T cells and NK cells distinguish the surface of virus-infected cells from that of uninfected cells are different ([Movie 24.2](#)).

Both cytotoxic T cells and NK cells recognize the same special class of cell-surface proteins on a host cell to help determine if the cell is virus-infected, but they use distinct receptors to do so. The special cell-surface proteins recognized are called **class I MHC proteins**, because they are encoded by genes in the *major histocompatibility complex*; almost all nucleated cells in vertebrates express these genes, and we discuss them in detail later. Cytotoxic T cells use both **T cell receptors (TCRs)** and **co-receptors** to recognize peptide fragments of viral proteins bound to class I MHC proteins on the surface of virus-infected host cells and then induce the infected cells to kill themselves. By contrast, NK cells have cell-surface **inhibitory receptors** that monitor the level of class I MHC proteins on the surface of other host cells: the high levels of these MHC proteins normally present on healthy cells engage these receptors and thereby inhibit the killing activity of the NK cells. The NK cells thus focus primarily on host cells expressing abnormally low levels of class I MHC proteins and induce them to kill themselves; these are mainly virus-infected cells and some cancer cells ([Figure 24-9](#)). NK cell killing activity is stimulated when various **activating receptors** on the NK cell surface recognize specific proteins that are greatly increased on the surface of virus-infected cells and some cancer cells.

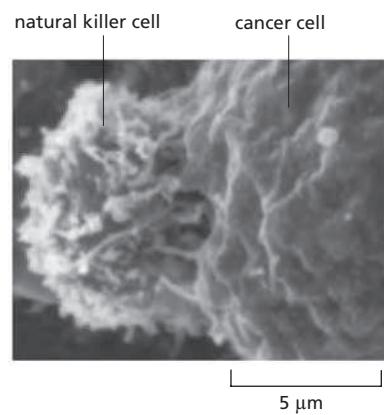


Figure 24-9 A natural killer (NK) cell attacking a cancer cell. This scanning electron micrograph was taken shortly after the NK cell attached to the cancer cell, but before it induced the cell to die by apoptosis. (Courtesy of J.C. Hiserodt, in Mechanisms of Cytotoxicity by Natural Killer Cells [R.B. Herberman and D. Callewaert, eds.]. New York: Academic Press, 1995.)

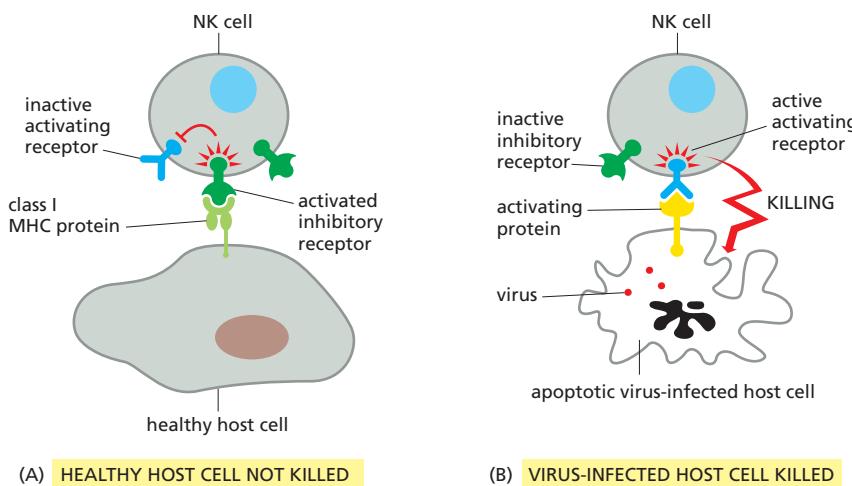


Figure 24–10 How an NK cell recognizes its target. An NK cell preferentially attacks infected host cells and cancer cells because these cells have on their surface both activating proteins and, in some cases, abnormally low levels of class I MHC proteins. (A) The high levels of class I MHC proteins found on normal host cells activate inhibitory receptors on the NK cell that suppress the killing activity of the NK cell. (B) In contrast, the activating proteins on infected cells and cancer cells bind to activating receptors on the NK cell and stimulate the killing activity of the cell.

The reason that class I MHC protein levels are often low on virus-infected cells is that many viruses have developed a variety of mechanisms to inhibit the expression of these proteins on the surface of the host cells they infect, in order to avoid detection by cytotoxic T cells: some viruses encode proteins that block class I MHC gene transcription; others block the intracellular assembly of peptide-MHC complexes; still others block the transport of these complexes to the cell surface. By evading recognition by cytotoxic T cells in these ways, however, a virus incurs the wrath of NK cells, which recognize the infected cells as being different—both because the infected cells express little class I MHC protein and because they express large amounts of other surface proteins that are recognized by the activating receptors on the NK cells (Figure 24–10).

Dendritic Cells Provide the Link Between the Innate and Adaptive Immune Systems

Dendritic cells are crucially important components of the innate immune system. They are a heterogeneous class of cells that are widely distributed in the tissues and organs of vertebrates. They express a large variety of PRRs, which enable dendritic cells to recognize and phagocytose invading pathogens and their products and to become activated in the process. The activated dendritic cells cleave the proteins of the pathogen into peptide fragments, which bind to newly synthesized MHC proteins, which then carry the fragments to the dendritic cell surface. The activated cells then migrate to a nearby lymphoid organ such as a lymph node (also called a lymph gland), where they present the peptide-MHC complexes to T cells of the adaptive immune system, activating the T cells to join in the battle against the specific pathogen (Figure 24–11).

In addition to the complexes of MHC proteins and microbial peptides displayed on their cell surface, activated dendritic cells also display cell-surface *co-stimulatory proteins* that help activate T cells (see Figure 24–11). As we discuss later, 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 mounts slower but more powerful and highly specific responses to attack an invader, as we now discuss.

Summary

All multicellular organisms possess innate immune defenses against invading pathogens; these defenses include physical and chemical barriers and various defensive cell responses. In vertebrates, these innate defense responses can also

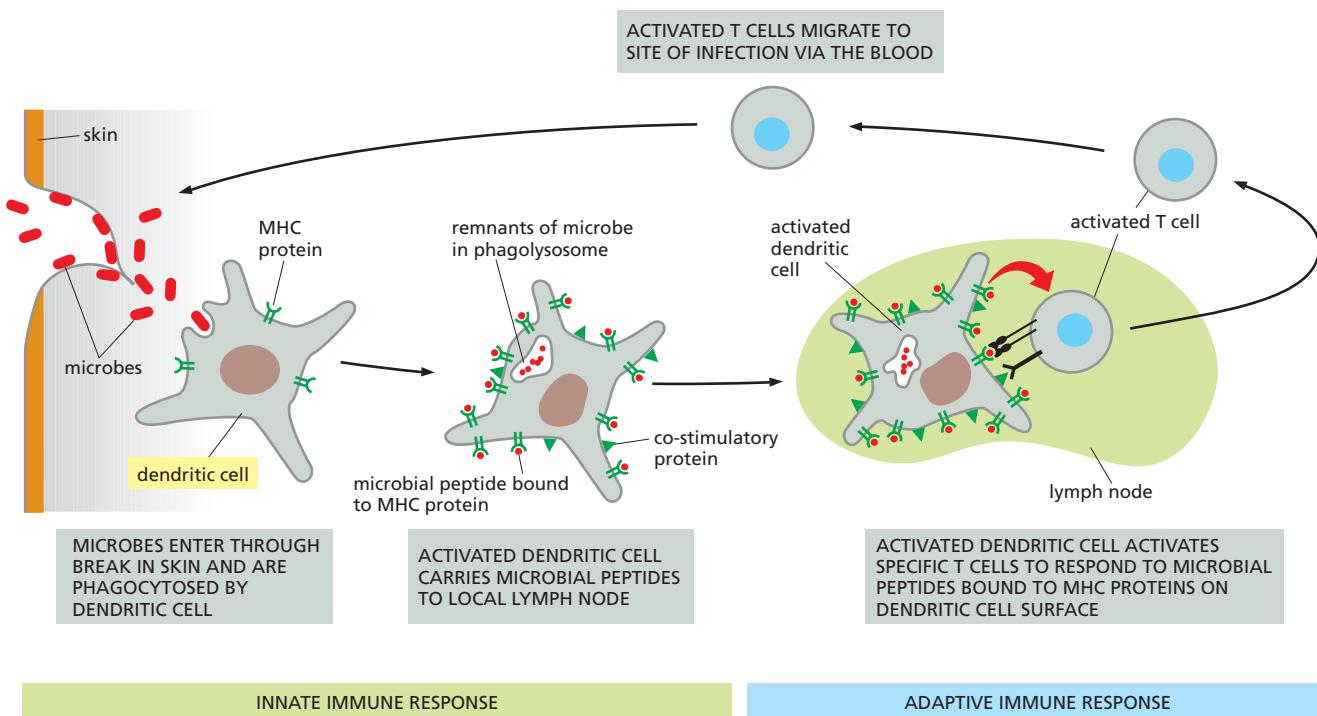


Figure 24–11 Dendritic cells as functional links between the innate and adaptive immune systems. Dendritic cells pick up invading microbes or their products at the site of an infection. The microbial PAMPs activate the dendritic cells to express co-stimulatory proteins and increased amounts of MHC proteins on their surface and to migrate via lymphatic vessels to a nearby lymph node. In the lymph node, the activated dendritic cells activate T cells that express appropriate receptors for the co-stimulatory proteins and the microbial peptides bound to MHC proteins on the dendritic cell surface. The activated T cells proliferate, and some of their progeny migrate to the original site of infection, where they help eliminate the microbes, either by activating local macrophages or by killing infected host cells (not shown). In addition, some of the activated T cells help stimulate specific B cells in the lymph node to secrete antibodies against the microbe (not shown).

A crucial feature of dendritic cell activation is that the pathogen provides an individual dendritic cell with both the peptides for presentation to T cells and the PAMP signals that activate the dendritic cell to express co-stimulatory proteins. In this way, the individual dendritic cell has all it needs to activate specific T cells that recognize the peptide–MHC complexes on its surface ([Movie 24.3](#)).

recruit specific and more powerful adaptive immune responses to help fight the infection. Innate immune responses rely on the ability of host cells to recognize characteristic features of microbial molecules called pathogen-associated molecular patterns, or PAMPs, which can be associated with a pathogen's proteins, lipids, sugars, or nucleic acids. PAMPs are mainly recognized by pattern recognition receptors (PRRs), including the Toll-like receptors (TLRs) found on or in both plant and animal cells. In vertebrates, some PRRs are secreted and can activate complement when they bind microbial PAMPs. The complement system, which can also be activated by antimicrobial antibodies bound to pathogens, consists of a group of blood proteins that are activated in sequence to help fight infections, by disrupting the pathogen's membrane, stimulating an inflammatory response, or targeting the microbe for phagocytosis—mainly by macrophages and neutrophils. The phagocytes use a combination of degradative enzymes, antimicrobial peptides, and oxygen-derived toxic molecules to kill invading pathogens; in addition, they secrete various signal molecules that help trigger an inflammatory response.

Cells infected by a virus produce and secrete type I interferons (IFNa and IFN β), which induce a complex set of host-cell responses that inhibit viral replication. The interferons also enhance the killing activity of natural killer (NK) cells. An NK cell kills infected host cells because they express large amounts of surface proteins that activate the NK cell; the killing is especially efficient when infected cells express reduced amounts of class I MHC proteins, which, when present in normal amounts on a host cell surface inhibit the killing activity of NK cells.

Dendritic cells of the innate immune system functionally link innate immune responses to adaptive immune responses. The cells become activated when their PRRs pick up microbes and their products at sites of infection and phagocytose them. The activated cells cleave the microbial proteins into peptide fragments, which bind to newly made MHC proteins, which transport the fragments to the cell surface. The activated dendritic cells then carry the peptide–MHC complexes to a lymph organ, where they activate appropriate T cells to make specific adaptive immune responses against the microbes.

OVERVIEW OF THE ADAPTIVE IMMUNE SYSTEM

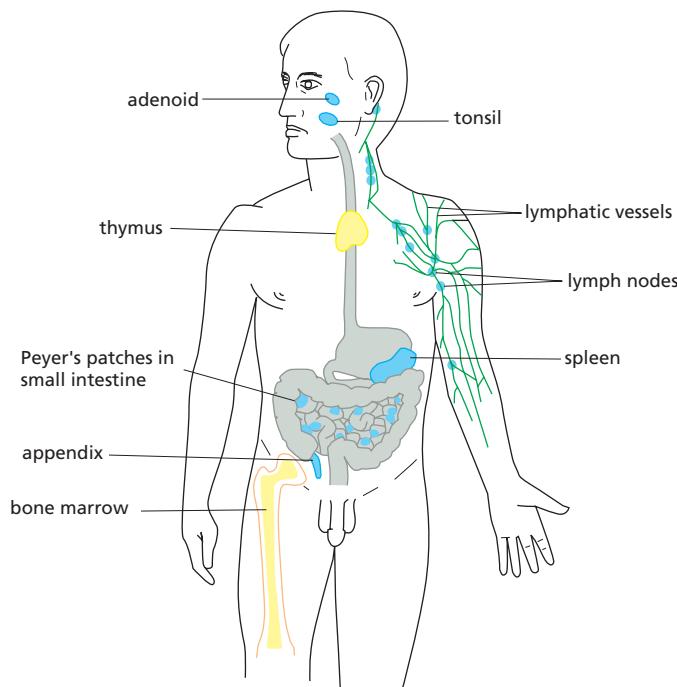
A dramatic “big bang” in immune defense mechanisms occurred when jawed vertebrates evolved and acquired an **adaptive immune system**. This sophisticated defense system depends on B and T lymphocytes (B and T cells), which, during their development, rearrange particular DNA sequences in various combinations so that, together, the cells can produce an almost limitless variety of B and T cell receptors and antibodies. Collectively, these proteins can bind to essentially any molecule, including small chemicals, carbohydrates, lipids, and proteins; individually, they can distinguish between molecules 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 small molecule. By this strategy, the adaptive immune system can recognize and respond specifically to any pathogen, including new mutant forms. However, because the genetic rearrangement process produces both receptors that can bind to self molecules as well as receptors that can bind to foreign molecules, vertebrates have had to evolve special mechanisms to ensure that B and T cells do not react against the host’s own molecules and cells—a process called *immunological self-tolerance*.

Moreover, many harmless foreign substances enter the body, for example, as food or inhaled material, and it would be pointless and potentially dangerous to mount adaptive immune responses against them. Such inappropriate responses are normally avoided because innate immune responses are required to call adaptive immune responses into play and do so only when the innate cells’ PRRs recognize microbial PAMPs, as we discussed earlier. One can trick the adaptive immune system into responding to a harmless foreign molecule, such as a foreign protein, by co-injecting a molecule (often of microbial origin) called an **adjuvant**, which activates PRRs. This trick is called **immunization** and is the basis of vaccination. Any substance capable of stimulating B or T cells to make a specific adaptive immune response against it is referred to as an **antigen** (*antibody generator*).

There are two broad classes of adaptive immune responses—*antibody responses* and *T-cell-mediated immune responses*, and most pathogens induce both classes of responses. In **antibody responses**, B cells are activated to secrete antibodies, which are proteins that circulate in the bloodstream and permeate the other body fluids, where they can bind specifically to the foreign antigen that stimulated their production (see Figure 24–2). Binding of antibody neutralizes extracellular viruses and microbial toxins (such as tetanus toxin or cholera toxin) by blocking their ability to bind to receptors on host cells. Antibody binding also marks invading pathogens for destruction, both by making it easier for phagocytes of the innate immune system to ingest and destroy them and by activating the complement system.

In **T-cell-mediated immune responses**, T cells recognize foreign antigens that are bound to MHC proteins on the surface of host cells such as dendritic cells, which are specialized for presenting antigen to T cells and are therefore referred to as *professional antigen-presenting cells (APCs)*. Because MHC proteins carry fragments of pathogen proteins from inside a host cell to the cell surface, T cells can detect pathogens hiding inside a host cell and either kill the infected cell (see Figure 24–2) or stimulate phagocytes or B cells to help eliminate the pathogens.

In this section, we discuss the origins and general properties of B and T cells. In later sections, we consider the specific properties and functions of these cells.

**Figure 24–12** Human lymphoid organs.

Lymphocytes develop from lymphoid progenitor cells 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).

B Cells Develop in the Bone Marrow, T Cells in the Thymus

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. They occur in large numbers in the blood and lymph (the colorless fluid in the lymphatic vessels, which 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, and spleen (Figure 24–12), and many are also found in other organs, including skin, lung, and gut.

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 adult mammals, develop in the *bone marrow*. Both types of cells develop from lymphoid progenitor cells that are produced from multipotent *hematopoietic stem cells*, which are found mainly in the bone marrow (Figure 24–13). The hematopoietic stem cells give rise to more than just lymphocytes: as discussed in Chapter 22, they produce all of the cells of the

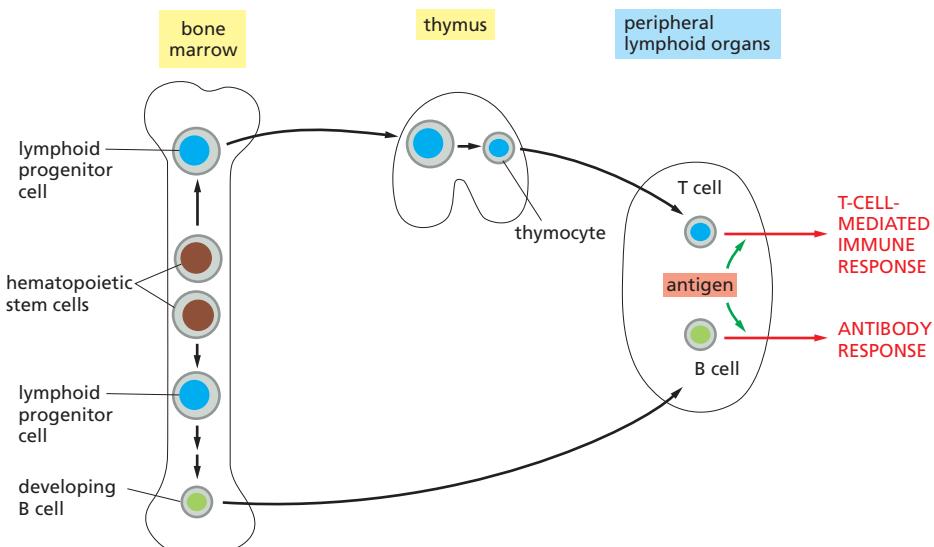


Figure 24–13 The development of B and T cells. The central lymphoid organs, where lymphocytes develop from lymphoid progenitor cells, are labeled in yellow boxes. The lymphoid progenitor cells develop from multipotent hematopoietic stem cells in the bone marrow. Some lymphoid progenitor cells develop 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). Foreign antigens activate B cells and T cells mainly in peripheral lymphoid organs, such as lymph nodes or the spleen.

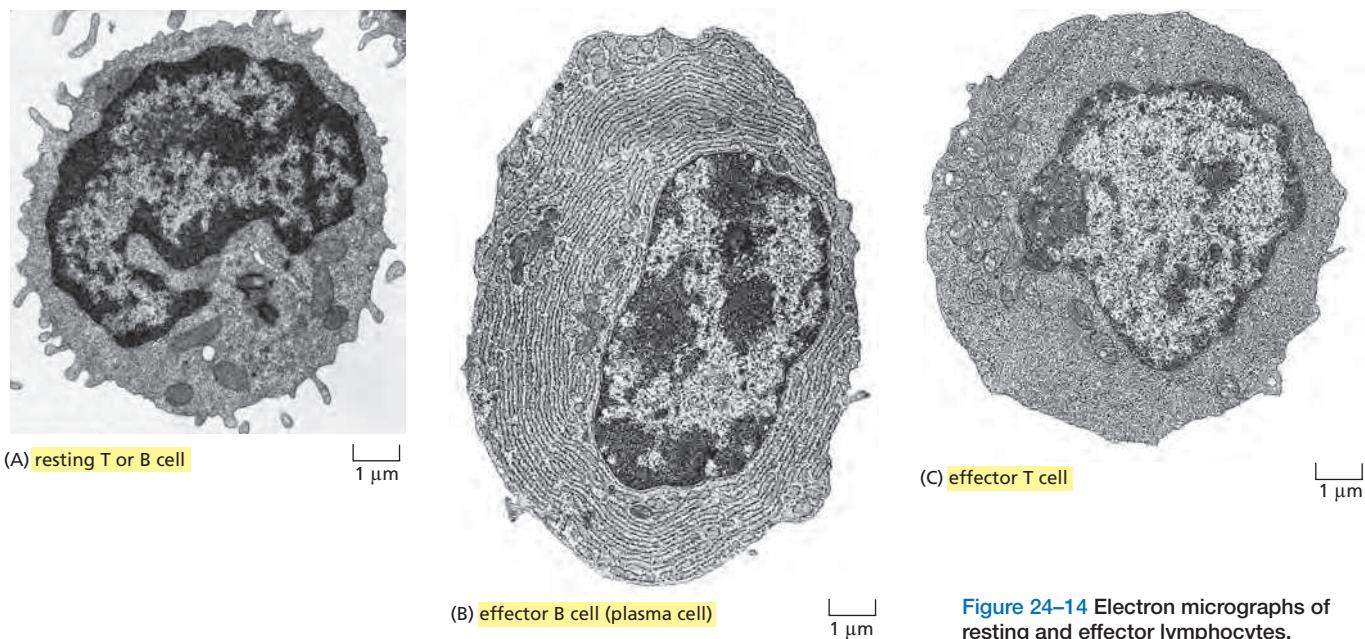


Figure 24-14 Electron micrographs of resting and effector lymphocytes.

(A) This resting lymphocyte could be either a B cell or a T cell, as these cells are difficult to distinguish morphologically until antigen activates them 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 that are secreted in large amounts. (C) An effector T cell, which has relatively little rough ER but is filled with free ribosomes; it secretes cytokines, but in relatively small amounts. 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.)

hematopoietic system, including erythrocytes, leukocytes, and platelets (see Figure 22-32).

Because they are sites where lymphocytes develop from lymphoid progenitor cells, the thymus and bone marrow are referred to as **central (primary) lymphoid organs** (see Figure 24-12). As we discuss later, most B and T cells 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. It is in these peripheral lymphoid organs that foreign antigens activate B and T cells (see Figure 24-13).

B and T cells become morphologically distinguishable from each other only after antigen has activated them: resting B and T cells look very similar, even in an electron microscope (Figure 24-14A). 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 24-14B). In contrast, effector T cells (Figure 24-14C) contain very little endoplasmic reticulum and secrete a variety of cytokines rather than antibodies. Whereas B-cell-derived antibodies are widely distributed by the bloodstream, T-cell-derived cytokines mainly act locally on neighboring cells, although some are carried via the blood and act on distant host cells.

Immunological Memory Depends On Both Clonal Expansion and Lymphocyte Differentiation

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, collectively, can make more than 10^{12} different antibody molecules that react specifically with the antigen that induced their production. How can B cells and T cells respond specifically to such an enormous diversity of foreign antigens? The answer for both B and T cells is the same. 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 receptors that specifically bind the antigen. When a lymphocyte encounters its antigen in a peripheral lymphoid organ, the binding of the antigen to the receptors (with help from co-stimulatory signals, discussed later)

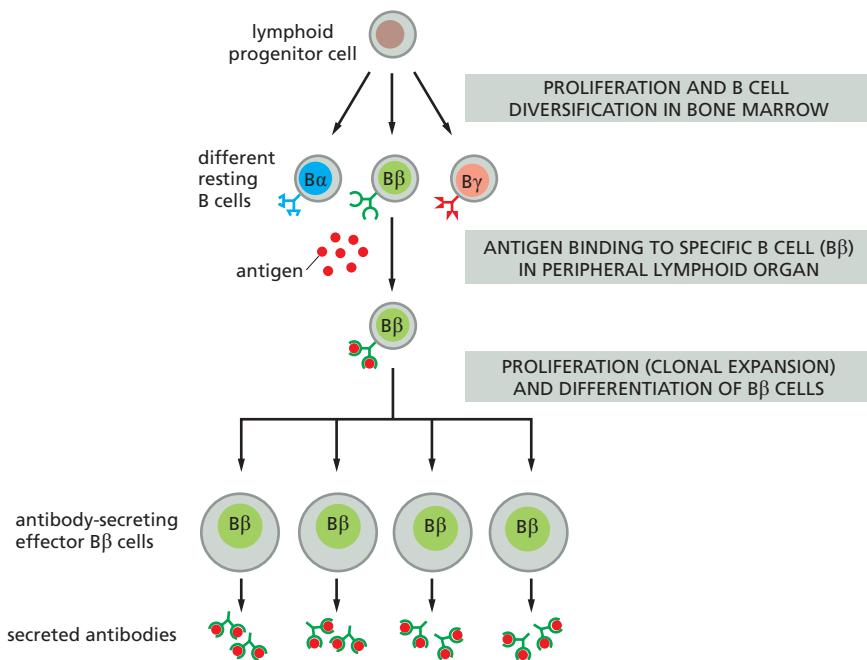


Figure 24–15 Clonal selection. An antigen activates only those lymphocytes that are already committed to respond to it. The committed cell expresses cell-surface receptors that specifically recognize the antigen. The human adaptive immune system consists of many millions of different T and B lymphocyte clones, with cells within a clone expressing the same unique antigen receptor. Before its 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, each expressing a different antigen receptor that binds either a different part of the antigen or the same part with a different binding affinity. Although only B cells are shown here, T cells are selected in a similar way. Note that the antigen receptors on the B cells labeled β in this diagram have the same antigen-binding site as the antibodies secreted by the effector $B\beta$ cells. As we discuss later, B cells require co-stimulatory signals from T cells to become activated by antigen to proliferate and differentiate into antibody-secreting cells (not shown).

activates the lymphocyte; this causes the lymphocyte to proliferate, thereby producing many more cells with the same receptor—a process called *clonal expansion*. The encounter with antigen also causes some of 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 24–15). This arrangement, called **clonal selection**, provides an explanation for **immunological memory**, whereby we develop lifelong immunity to many common infectious diseases after our initial exposure to the pathogen—either through natural infection or vaccination.

It is easy to demonstrate such immunological memory in experimental animals. If an animal is immunized once with antigen A, an immune response (antibody, T-cell-mediated, or both) can be detected after several days; the response rises rapidly and 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, because there are now many more preexisting B or T cells (or both) with specificity for antigen A, 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 therefore reflects antigen-specific immunological memory for antigen A (Figure 24–16).

Immunological memory depends on both lymphocyte proliferation and differentiation. 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 specific foreign 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, whereas effector T cells either kill infected cells or influence the response of other immune cells—by secreting cytokines, for example). Some of the antigen-stimulated naïve cells multiply and differentiate into **memory cells**, which are more easily and more quickly induced to become effector cells by a later encounter with the same antigen: like naïve cells, when memory cells encounter their antigen, they give rise to either effector cells or more memory cells (Figure 24–17).

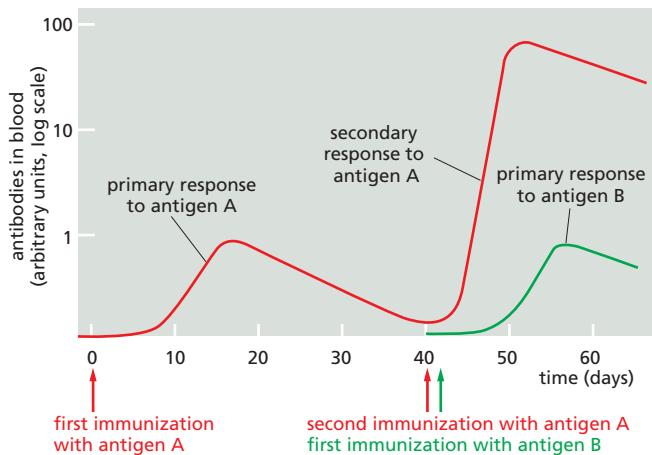


Figure 24–16 Immunological memory: primary and secondary antibody responses. The secondary response induced by a second exposure to antigen A is faster and greater than the primary response and is specific for A, indicating that the adaptive immune system has specifically remembered its previous encounter with antigen A. The same type of immunological memory is observed in T-cell-mediated responses (not shown). 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.

Thus, during the primary response, clonal expansion and differentiation of antigen-stimulated naïve cells creates many memory cells, which are able to respond to the same antigen more sensitively, rapidly, and effectively. 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. Although most effector B and T cells die after an immune response is over, some survive as effector cells and help provide long-term protection against the pathogen. A small proportion of the plasma cells produced in a primary B cell response, for example, can survive for many months or years in the bone marrow, where they continue to secrete their specific antibodies into the bloodstream.

Lymphocytes Continuously Recirculate 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, microbes or their products must travel from these entry points to a peripheral lymphoid organ, such as a lymph node or the spleen, which are the sites where lymphocytes are activated (see Figure 24–11). The route and destination depend on the site of entry. Lymphatic vessels 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 (see Figure 24–12). As discussed earlier (see Figure 24–11), in many cases, activated 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, as we discuss later.

But only a tiny fraction of naïve B and T cells can recognize a particular microbial antigen in a peripheral lymphoid organ, a reasonable estimate being between 1/10,000 and 1/1,000,000 of each class of lymphocyte, depending on the antigen. How do these rare cells find an antigen-presenting cell displaying their specific antigen? The answer is that the lymphocytes continuously recirculate 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*

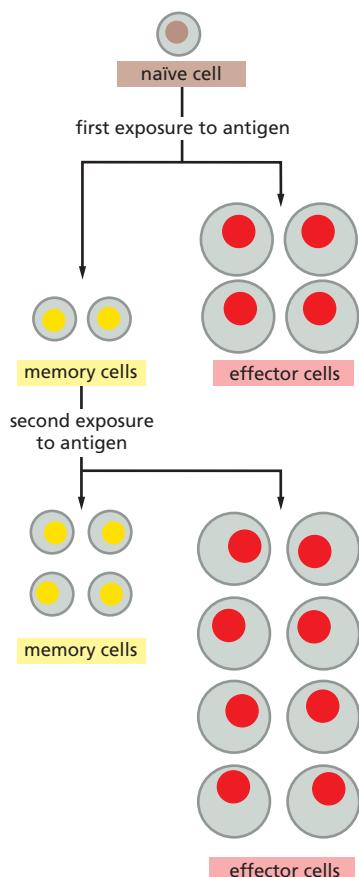


Figure 24–17 A model for the cellular basis of immunological memory.

When stimulated by their specific antigen and co-stimulatory signals, naïve lymphocytes 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. Some memory T cells also develop from a minority of effector T cells (not shown). It is not known how the decision to become an effector cell versus a memory cell is made.

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 24–12). 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 24–18).

The continuous recirculation of a lymphocyte between the blood and lymph ends only if its specific antigen activates it in a peripheral lymphoid organ. In that case, 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 blood to the site of infection (see Figure 24–11), whereas others stay in the lymphoid organ and help activate (or suppress) other immune 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. Lymphocytes that enter a lymph node via the blood, for example, adhere weakly to specialized endothelial cells lining the postcapillary venules via *homing receptors* that belong to the *selectin* family of cell-surface lectins that bind to specific sugar groups on the endothelial cell surface (see Figure 19–28). The lymphocytes roll slowly along the surface of the endothelial cells until another, much stronger adhesion system, dependent on an integrin protein, is called into play by chemokines secreted by the endothelial cells. Now, the lymphocytes stop rolling, and they crawl out of the blood vessel into the lymph node by using yet another cell adhesion protein called CD31 (Figure 24–19). Although B and T cells initially enter the same region of a lymph node, different chemokines guide them to separate regions of the node—B cells to *lymphoid follicles* and T cells to the *paracortex* (Figure 24–20).

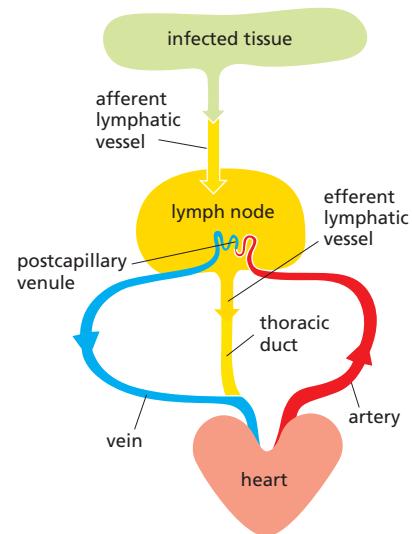


Figure 24–18 The path followed by lymphocytes as they continuously recirculate between the lymph and blood. The circulation through a lymph node (yellow) is shown here. Microbial antigens are usually carried into the lymph node by activated dendritic cells (not shown), which enter the node via afferent lymphatic vessels draining an infected tissue (green). B and T cells, by contrast, enter via the blood, migrating out of the bloodstream into the lymph node through postcapillary venules. Unless they encounter their antigen, the B and T 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, completing the circulation cycle for T and B cells. A typical circulation cycle for these lymphocytes takes about 12–24 hours.

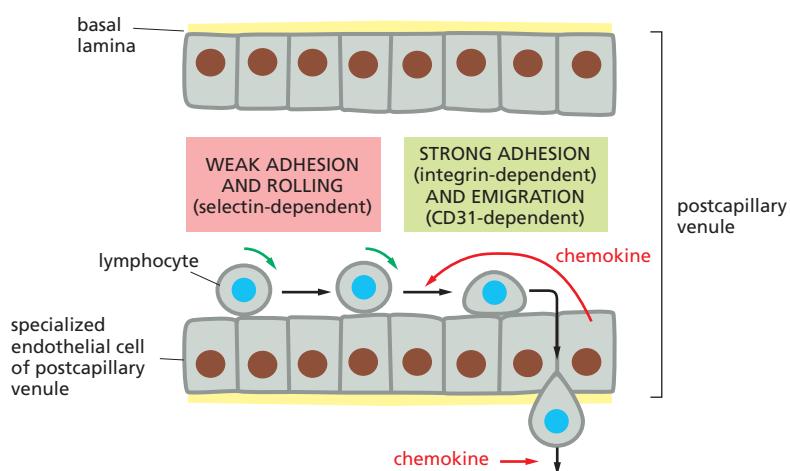


Figure 24–19 Migration of a lymphocyte out of the bloodstream into a lymph node.

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 (discussed in Chapter 19) on the lymphocyte surface. The adhesion is sufficiently weak to enable the lymphocyte, pushed by the flow of blood, to roll along the surface of the endothelial cells. Stimulated by chemokines secreted by specialized endothelial cells in the node (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 immunoglobulin-like cell adhesion protein (CD31) to bind to the junctions between adjacent endothelial cells and migrate out of the venule. The subsequent migration of the lymphocyte in the lymph node depends on chemokines produced within the node (straight red arrow). The migration of other types of leukocytes out of the bloodstream into sites of infection occurs in a similar way (see Figure 19–28 and Movie 19.2).

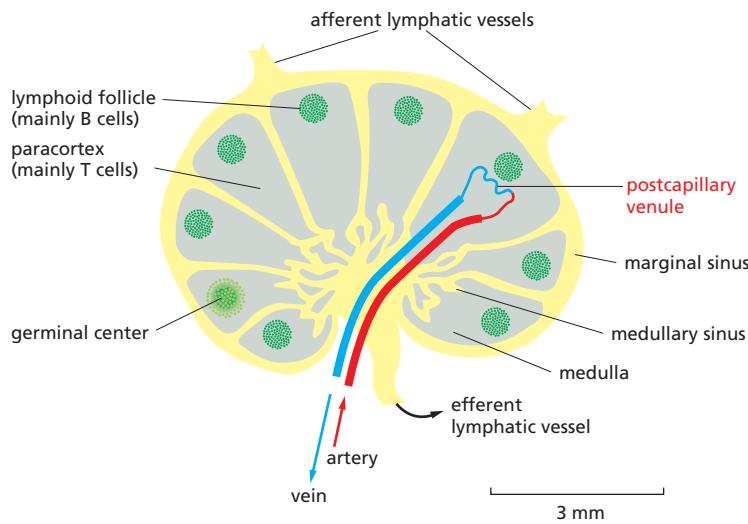


Figure 24–20 A simplified drawing of a human lymph node. B cells are primarily clustered in structures called lymphoid follicles, whereas T cells are found mainly in the paracortex. Chemokines attract both types of lymphocytes into the lymph node from the blood via postcapillary venules (see Figure 24–19). B and T cells then migrate to their respective areas, attracted by different chemokines. If they do not encounter their specific antigen, both B cells and T 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 24–18). During an infection, proliferation of pathogen-specific B cells produces a germinal center in some lymphoid follicles.

Unless they encounter their antigen, both B and T 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 B cell and T cell areas, where the rare antigen-specific B and T cells can interact, leading to their proliferation and differentiation 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—effector plasma B cells to the bone marrow and effector T cells to sites of infection.

Immunological Self-Tolerance Ensures That B and T Cells Do Not Attack Normal Host Cells and Molecules

As discussed earlier, cells of the innate immune system use PRRs to distinguish microbial molecules from self molecules made by the host. The adaptive immune system has the far more difficult recognition task of responding specifically to an almost unlimited number of foreign molecules while not responding to the large number of self molecules. How does it accomplish this feat? It helps that self molecules normally do not induce the innate immune reactions required to activate adaptive immune responses. But even when an infection or tissue injury triggers innate reactions, the vast majority of self molecules normally still fail to induce an adaptive immune response. Why?

One important reason is that the adaptive immune system “learns” not to respond to self molecules. Normal mice, for example, cannot mount an immune response against one of their own protein components of the complement system called C5 (see Figure 24–7). However, mutant mice that lack the gene encoding C5 but are otherwise genetically identical to normal mice of the same strain can make a strong immune response to this blood protein when immunized with it. The **immunological self-tolerance** exhibited by normal mice persists only for as long as the self molecule remains 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, as new B and T cells develop in the absence of C5. Thus, the adaptive immune system is genetically capable of responding to self molecules, but it learns not to do so.

Self-tolerance depends on a number of distinct mechanisms, including the following (**Figure 24–21**):

1. In *receptor editing*, developing B cells that recognize self molecules change their antigen receptors so that the cells no longer do so.
2. In *clonal deletion*, potentially self-reactive B and T cells die by apoptosis when they encounter their particular self molecule.

3. In *clonal inactivation* (also called clonal anergy), self-reactive B and T cells become functionally inactivated when they encounter their self molecule.
4. In *clonal suppression*, self-reactive *regulatory T cells* (discussed later) suppress the activity of other types of potentially self-reactive lymphocytes.

Some of these mechanisms—especially the first two, receptor editing in B cells and clonal deletion of B and T cells—operate in central lymphoid organs when newly formed self-reactive B and T cells first encounter their self molecules, and they are largely responsible for the process called *central tolerance*. Clonal inactivation and clonal suppression, by contrast, operate mainly when mature B and T cells encounter their self molecules in peripheral lymphoid organs, and they are largely responsible for the process called *peripheral tolerance*. Clonal deletion, however, can also operate peripherally, and clonal inactivation can also operate centrally.

Why does the binding of a self molecule lead to tolerance rather than activation? The answer is still not completely known. As we discuss later, the activation of a B or T cell by its antigen in a peripheral lymphoid organ requires more than just antigen binding: it requires co-stimulatory signals, which are provided by a *helper T cell* (discussed later) in the case of a B cell and by an activated dendritic cell in the case of a naïve T cell. The production of such signals is usually triggered by exposure to a pathogen, but a self-reactive lymphocyte normally encounters its self antigen in the absence of such signals. Under these conditions, the lymphocyte will not only fail to be activated, it will often be rendered tolerant—being either killed or inactivated, or actively suppressed by a regulatory T cell (see Figure 24–21). In peripheral lymphoid organs, both T cell tolerance and activation usually occur on the surface of a dendritic cell.

For reasons that are usually unknown, self-tolerance mechanisms sometimes fail, causing T or B cells (or both) to react against the animal's own molecules.

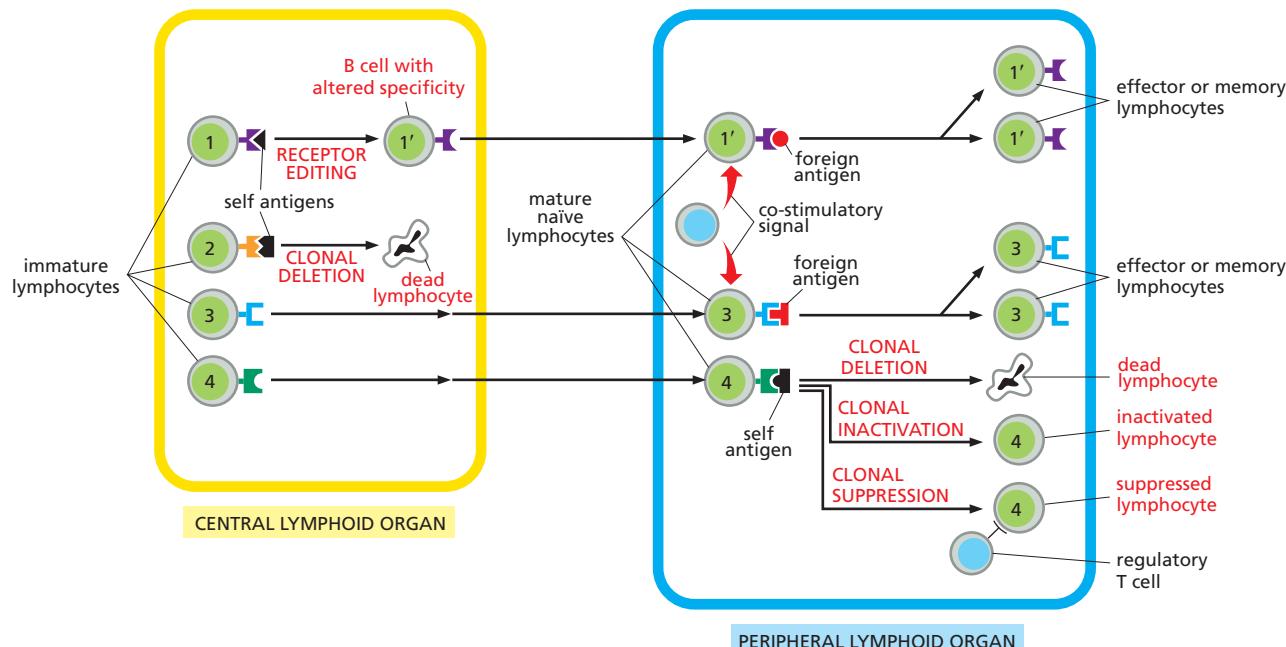


Figure 24–21 Mechanisms of immunological self-tolerance. When a self-reactive immature B cell binds its self molecule in the central lymphoid organ where the cell is produced, it may alter its antigen receptor so that it is no longer self-reactive (cell 1); this process is called receptor editing. Alternatively, when either an immature B or T cell binds its self molecule in a central lymphoid organ, it 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 B or T cell escapes tolerance in the central lymphoid organ and binds its self molecule in a peripheral lymphoid organ (cell 4), or in another peripheral tissue, it will generally not be activated, because the binding usually occurs in the absence of sufficient co-stimulatory signals; instead, the cell may die by apoptosis (often after a period of proliferation), be inactivated, or be suppressed by a regulatory T cell. These forms of tolerance (shown on the right) are called *peripheral tolerance*. As discussed later, the cells providing the co-stimulatory signals are T lymphocytes for B cells and usually dendritic cells for T cells (not shown). For T cells at least, both activation and tolerance in a peripheral lymphoid organ usually occurs on the surface of a dendritic cell, although the dendritic cells are different in the two cases.

Myasthenia gravis is an example of such an **autoimmune disease**. Most of the affected individuals make antibodies against the acetylcholine receptors on their own skeletal muscle cells; these receptors are required for the muscle to contract normally in response to nerve stimulation, which releases acetylcholine (see Figure 11–39). The antibodies interfere with the normal functioning of the receptors so that the patients become weak and may die because they cannot breathe. Similarly, in *juvenile (type 1) diabetes*, adaptive immune reactions against insulin-secreting β cells in the pancreas kill these cells, leading to severe insulin deficiency.

Summary

Innate immune responses triggered by pathogens at sites of infection help activate adaptive immune responses in peripheral lymphoid organs. The adaptive immune system is composed of many millions of B and T cell clones, with the cells in each clone sharing a unique cell-surface receptor that enables them to bind a particular pathogen antigen. The binding of antigen to these receptors, with the help of co-stimulatory signals, stimulates the lymphocyte to proliferate and differentiate into an effector cell that can help eliminate the pathogen. Effector B cells secrete antibodies, which can act over long distances to help eliminate extracellular pathogens and their toxins. Effector T cells, by contrast, produce cell-surface and secreted co-stimulatory molecules, which mainly act locally to help other immune cells eliminate the pathogen; in addition, some T cells can induce infected host cells to kill themselves.

During a primary adaptive immune response to an antigen, lymphocytes that recognize the antigen proliferate so that there are more of them to respond the next time, during a secondary response to the same antigen; moreover, during a primary response, some lymphocytes differentiate into memory cells, which can respond faster and more efficiently the next time the same pathogen invades. These two mechanisms are largely responsible for immunological memory. 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 (in the case of B cells) or are eliminated or inactivated; they can also be suppressed by regulatory T cells. These mechanisms collectively are responsible for immunological self-tolerance, which ensures that the adaptive immune system normally avoids attacking the molecules and cells of the host.

B CELLS AND IMMUNOGLOBULINS

Vertebrates inevitably die of infection if they are unable to make antibodies. **Antibodies** are secreted proteins that defend us against extracellular pathogens in several ways. They bind to viruses and microbial toxins, thereby preventing them from binding to and damaging host cells (see Figure 24–2). When bound to an extracellular pathogen or its products, antibodies also recruit some of the components of the innate immune system, including various types of leukocytes and components of the complement system, which work together to inactivate or eliminate the invaders.

Synthesized exclusively by B cells, antibodies are produced in billions of forms, each with a different amino acid sequence. They belong to the class of proteins called **immunoglobulins** (abbreviated as **Igs**) and are among the most abundant protein components in the blood. In this section, we discuss the structure and function of immunoglobulins and how they are made in so many different forms.

B Cells Make Immunoglobulins (Igs) as Both Cell-Surface Antigen Receptors and Secreted Antibodies

The first Igs 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. They are called **B cell receptors (BCRs)**, and each B cell has approximately 10^5 of

them in its plasma membrane. Each BCR is stably associated with invariant transmembrane proteins that activate intracellular signaling pathways when antigen binds to the BCR; we discuss these invariant proteins later, when we consider how B cells are activated with the assistance of *helper T cells*.

Each B cell clone produces a single species of BCR, with a unique antigen-binding site. When an antigen and a helper T cell activate a naïve or a memory B cell, the B cell proliferates and differentiates into an effector cell, which then produces and secretes large amounts of soluble (rather than membrane-bound) Ig. The secreted Ig is now called an antibody, and it has the same unique antigen-binding site as the BCR (Figure 24–22).

A typical Ig molecule is bivalent, with two identical antigen-binding sites. It consists of four polypeptide chains—two identical *light chains* and two identical *heavy chains*. The N-terminal parts of both light and heavy chains usually cooperate to form the antigen-binding surface, while the more C-terminal parts of the heavy chains form the tail of the Y-shaped protein (Figure 24–23). The tail mediates many of the activities of antibodies, and antibodies with the same antigen-binding sites can have any one of a number of different tail regions, each of which gives the antibody different functional properties, such as the ability to activate complement or to bind to receptor proteins on phagocytic cells that bind a specific type of antibody tail.

Mammals Make Five Classes of IgS

In mammals, there are five major *classes* of IgS, each of which mediates a characteristic biological response following antigen binding to an antibody: 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. Moreover, there are four human IgG subclasses (IgG1, IgG2, IgG3, and IgG4), with γ_1 , γ_2 , γ_3 , and γ_4 heavy chains, respectively. There are also two IgA subclasses in humans. In addition to the various classes and subclasses of heavy chains, 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, but an individual Ig molecule 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 Ig's antigen-binding sites are always identical (see Figure 24–22).

The various heavy chains give a distinctive conformation to the tail region of antibodies, so that each class (and subclass) has characteristic properties of its own. IgM is always the first class of Ig that a developing B cell in the bone marrow makes. It forms the BCRs on the surface of *immature naïve B cells*. After these cells

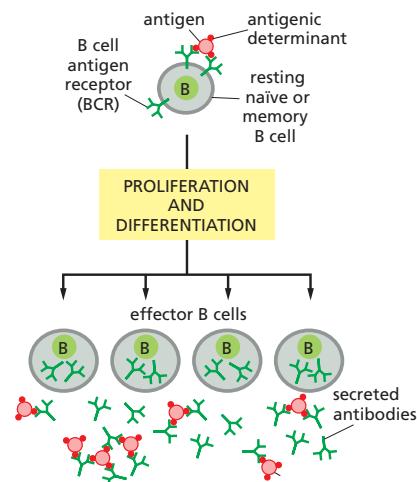


Figure 24–22 The B cell receptors (BCRs) and secreted antibodies made by a B cell clone. The binding of an antigen to BCRs on either a naïve or memory B cell (together with co-stimulatory signals provided by helper T cells—not shown) activates the cell to proliferate and differentiate into effector B cells. The effector cells produce and secrete antibodies with a unique antigen-binding site, which is the same as that of the cell-surface BCRs. Because antibodies have two identical antigen-binding sites, they can cross-link antigens, as shown for an antigen with multiple identical *antigenic determinants*.

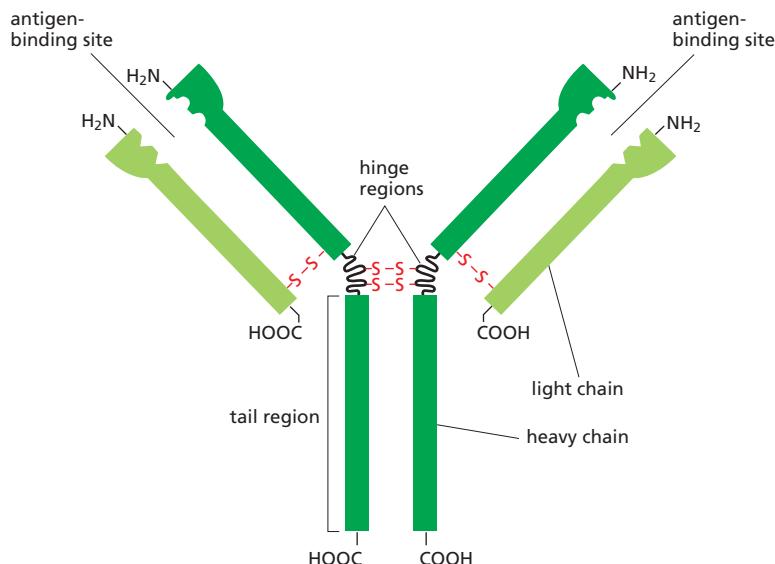
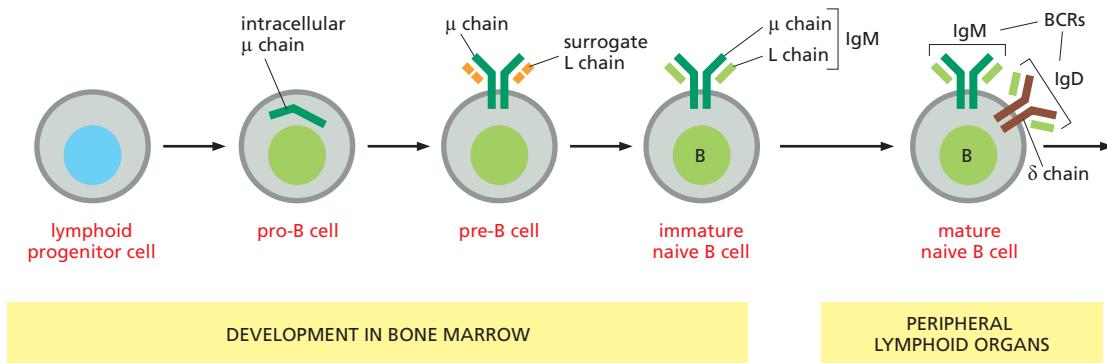


Figure 24–23 A schematic drawing of a bivalent antibody molecule. The two heavy chains each have a hinge region, which, because of its flexibility, improves the efficiency with which the antibody can cross-link antigens (see Figure 24–22). The two heavy chains also form the tail of the antibody, which determines its functional properties. The heavy and light chains are held together by both covalent S-S bonds (red) and noncovalent bonds (not shown).



leave the bone marrow, they start to produce IgD BCRs as well, with the same antigen-binding site as the IgM BCRs. These cells are now called *mature naïve B cells*, as they can now respond to their specific foreign antigen in peripheral lymphoid organs (Figure 24–24). IgM is also the major class of antibody secreted into the blood in the early stages of a primary antibody response on first exposure to an antigen. In its secreted form, IgM is a wheel-like pentamer composed of five four-chain units, giving it a total of 10 antigen-binding sites that allow it to bind strongly to pathogens; in its antigen-bound form, IgM is highly efficient at activating complement, which is important in early antibody responses to pathogens.

The major antibody class in the blood is IgG. These antibodies are four-chain monomers (see Figure 24–23), and they are produced in especially large quantities during secondary antibody responses. The tail region of some subclasses of IgG antibodies that are bound to antigen can activate complement and also bind to specific receptors on macrophages and neutrophils. 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; the activated Fc receptors also signal the phagocyte to secrete pro-inflammatory cytokines (Movie 24.4).

The tail region of IgE antibodies binds to another class of Fc receptors on the surface of *mast cells* in tissues and of *basophils* in the blood. Because antigen-free IgE antibodies bind with high affinity to such Fc receptors, the antibodies act as antigen receptors on these cells. Antigen binding to the bound antibodies activates the Fc receptors and stimulates the cells to secrete a variety of cytokines and biologically active amines, especially *histamine*, which causes blood vessels to dilate and become leaky; this helps leukocytes, 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 leukocytes called *eosinophils*, which also have Fc receptors that bind IgE molecules and can kill extracellular parasitic worms, especially if the worms are coated with IgE antibodies (see Figure 24–6).

IgA is the principal antibody class in secretions, including saliva, tears, milk, and respiratory and intestinal secretions. Yet another class of Fc receptors, located on the relevant epithelial cells, guides the secretion by binding antigen-free IgA dimers and transporting them across the epithelium. The properties of the various classes of antibodies in humans are summarized in Table 24–2.

All classes of Ig 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 Ig molecules (BCRs) 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 Ig molecules made occurs because the activation of B cells by antigen and helper T cells induces a change in the way in which the heavy-chain RNA transcripts are made and processed in the nucleus (see Figure 7–59).

Figure 24–24 Stages of B cell development. All of the stages shown occur before the cells bind their specific antigen. The first cells in the B cell lineage that make Ig are called *pro-B cells*; they make μ heavy chains, which remain in the endoplasmic reticulum until a special type of light chain is made called a surrogate light chain. The surrogate light chains substitute for genuine light chains and assemble with μ chains to form a receptor molecule that inserts into the plasma membrane. The cells are now called *pre-B cells*. Signaling from this pre-B cell receptor allows the cells to make bona fide light chains, which combine with μ chains to form four-chain IgM molecules that serve as cell-surface BCRs on *immature naïve B cells*. After these cells leave the bone marrow, they start to express IgD BCRs as well, which have the same antigen-binding sites as the IgM BCRs; it is this *mature naïve B cell* that reacts with its specific foreign antigen in peripheral lymphoid organs.

TABLE 24-2 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				
Number of four-chain units	5	1	1	1 or 2	1
Percentage of total Ig in blood	10	<1	75	15	<1
Activates classical complement pathway	+	-	+ (some subclasses)	-	-
Crosses from mother to fetus	-	-	+ (some subclasses)	-	-
Binds to macrophages and neutrophils	+ (macrophages only)	-	+ (some subclasses)	+	-
Binds to mast cells and basophils	-	-	+ (some subclasses)	-	+

Ig Light and Heavy Chains Consist of Constant and Variable Regions

Both light and heavy chains have a variable amino acid sequence at their N-terminal ends but a constant sequence at their C-terminal ends. Whereas the **constant region** and **variable region** of a light chain are the same size, the constant region of a heavy chain is about three or four times longer, depending on the class (**Figure 24-25**).

The variable regions of the light and heavy chains come together to form the antigen-binding sites, and the variability of their amino acid sequences provides the structural basis for the diversity of these binding sites. The greatest diversity occurs in three small **hypervariable regions** in the variable regions of both light and heavy chains. Only about 5–10 amino acids in each hypervariable region form the actual antigen-binding site (**Figure 24-26**). As a result, the size of the **antigenic determinant** that an Ig molecule recognizes is generally comparably small: it can consist of fewer than 10 amino acids on the surface of a globular protein, for example (see Figure 24-22).

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 24-27A**, a light chain consists of one variable (V_L) and one constant (C_L) domain, whereas a heavy chain has one variable and three or four constant domains: the variable domains of the light and heavy chains pair to form the antigen-binding region. Each Ig domain has a very similar three-dimensional structure, consisting of a sandwich of two β

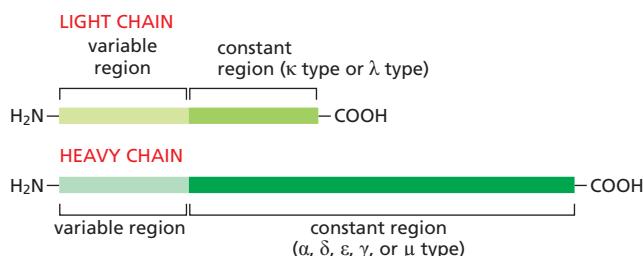


Figure 24-25 Constant and variable regions of Ig chains. The variable regions of the light and heavy chains form the antigen-binding sites, while the constant regions of the heavy chains determine the other biological properties of an Ig protein. The different subclasses of IgG antibodies have different γ -chain constant regions.

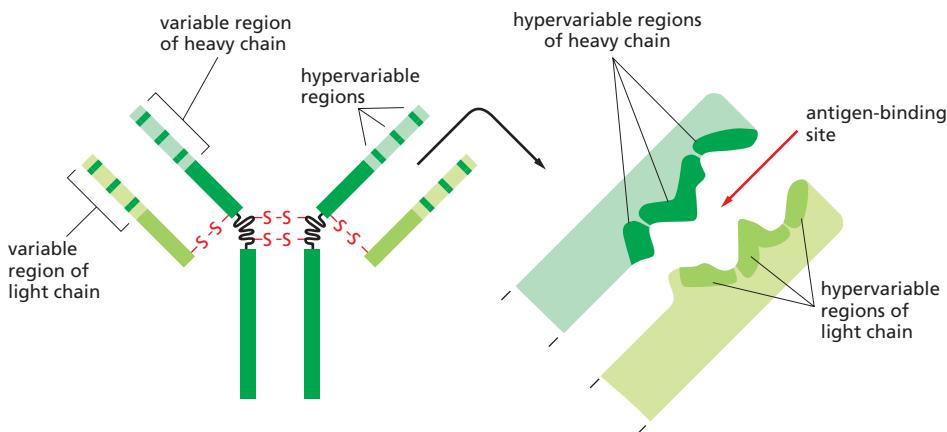


Figure 24–26 Ig hypervariable regions. Highly schematized drawing of how the three hypervariable regions in each light and heavy chain together form each antigen-binding site of an Ig protein.

sheets held together by a disulfide bond; the variable domains are unique in that each has its particular set of hypervariable regions, which are arranged in three *hypervariable loops* that cluster together at the ends of the variable domains to form the antigen-binding site (Figure 24–27B).

Ig Genes Are Assembled From Separate Gene Segments During B Cell Development

Even in the absence of antigen stimulation, a human can probably make more than 10^{12} different Ig molecules—its preimmune, **primary Ig repertoire**. The primary repertoire consists of IgM and IgD proteins and is apparently large enough to ensure that there will be an antigen-binding site to fit almost any potential

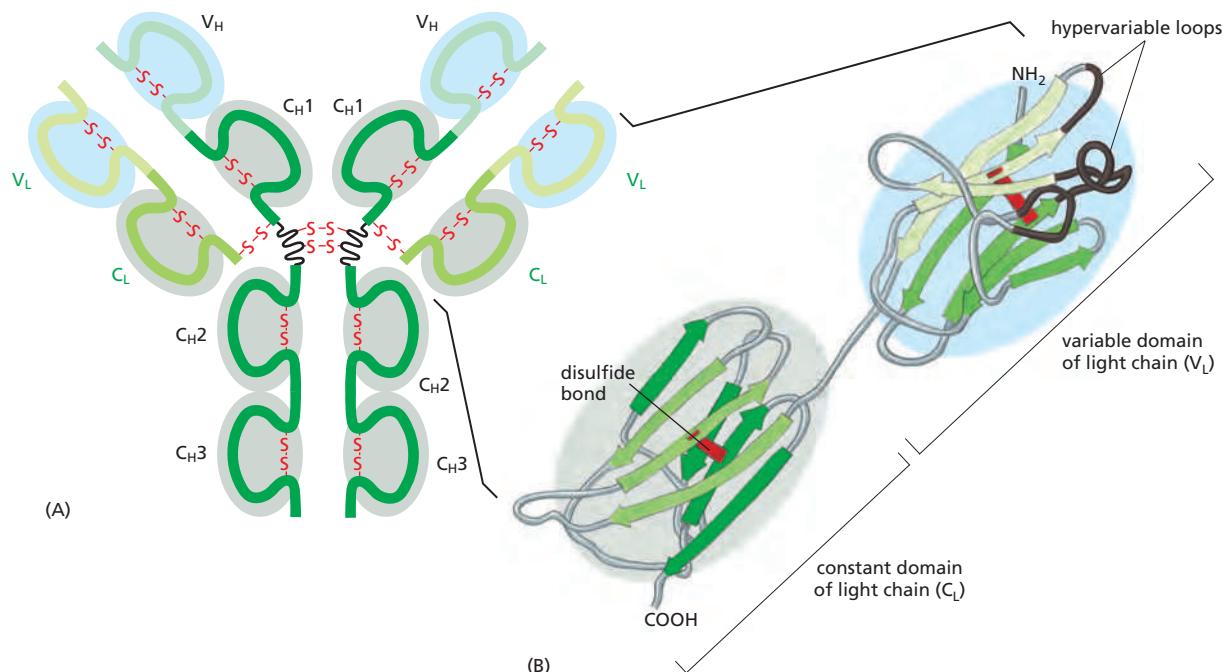


Figure 24–27 Ig domains. (A) The light and heavy chains in an Ig protein are each folded into similar repeating domains. The variable domains (shaded in blue) of the light and heavy chains (V_L and V_H) make up the antigen-binding sites, while the constant domains (shaded in gray) of the heavy chains (mainly C_{H2} and C_{H3}) determine the other biological properties of the protein. The heavy chains of IgM and IgE 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 Ig molecule: V_L binds to V_H , C_L binds to C_{H1} , and so on. (B) X-ray crystallography-based structures of the Ig domains of a light chain (Movie 24.5). Both the variable and constant domains have a similar overall structure, consisting of two β sheets joined by a disulfide bond (red). Note that all the hypervariable regions (black) form loops at the far end of the variable domain, where they come together to form part of the antigen-binding site. All Igs are glycosylated on their C_{H2} domains (not shown); the attached oligosaccharide chains vary from Ig to Ig and can greatly influence the biological properties of the protein, largely by affecting its binding to Fc receptors on immune cells.

antigenic determinant, albeit with low affinity— $K_a \approx 10^5\text{--}10^7$ liters/mole. After stimulation by antigen and helper T cells, B cells can switch from making IgM and IgD to making other classes of Ig—a process called *class switching*. In addition, the binding affinity of these IgS for their antigen progressively increases over time—a process called *affinity maturation*. Thus, antigen stimulation generates a **secondary Ig repertoire**, with a greatly increased affinity (K_a up to 10^{11} liters/mole) and diversity of both Ig classes and antigen-binding sites.

How can each of us make so many different IgS? The problem is not quite as formidable as it might first appear. Recall that the variable regions of the Ig light and heavy chains usually combine to form the antigen-binding site. Thus, if we had 1000 genes encoding light chains and 1000 genes encoding heavy chains, we could, in principle, combine their products in 1000×1000 different ways to make 10^6 different antigen-binding sites. Nonetheless, we have evolved special genetic mechanisms to enable our B cells to generate an almost unlimited number of different light and heavy chains in a remarkably economical way. We do so in two steps. First, before antigen stimulation, developing B cells join together separate *gene segments* in DNA to create the genes that encode the primary repertoire of low-affinity IgM and IgD proteins. Second, after antigen stimulation, the assembled *Ig genes* can undergo two further changes—mutations that can increase the affinity of their antigen-binding site and DNA rearrangements that switch the class of Ig made. Together, these changes produce the secondary repertoire of high-affinity IgG, IgE, and IgA proteins.

We produce our primary Ig repertoire by joining separate **Ig gene segments** together during B cell development. Each type of Ig 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 Ig chain, and one or more gene segments encoding the C region. During the development of a B cell in the bone marrow, a complete coding sequence for each of the two Ig chains to be synthesized is assembled by site-specific genetic recombination (discussed in Chapter 5). Once a V-region coding sequence is assembled next to a C-region sequence, it can then be co-transcribed and the resulting RNA transcript processed to produce an mRNA molecule that codes for the complete Ig polypeptide chain.

Each light-chain V region, for example, is encoded by a DNA sequence assembled from two gene segments—a long **V gene segment** and a short *joining* or **J gene segment** (Figure 24–28). 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 24–29). In addition to bringing together the separate gene segments of the Ig gene, these rearrangements also activate transcription from the gene promoter through changes in the relative positions of the *cis*-regulatory DNA sequences acting on the gene. Thus, a complete Ig chain can be synthesized only after the DNA has been rearranged.

The large number of inherited *V*, *J*, and *D* gene segments available for encoding Ig chains contributes substantially to Ig diversity, and the combinatorial joining of these segments (called *combinatorial diversification*) greatly increases this contribution. Any of the 35 or so functional V segments in our κ light-chain locus, for example, can be joined to any of the 5 J segments (see Figure 24–28), so that this locus can encode at least 175 (35×5) 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 23 or so D segments and to any of the 6 J segments to encode at least 5520 ($40 \times 23 \times 6$) different heavy-chain V regions. By this mechanism alone, called **V(D)J recombination**, a human can produce 295 different V_L regions (175 κ and 120 λ) and 5520 different V_H regions. In principle, these could then be combined to make over 1.5×10^6 (295×5520) different antigen-binding sites.

V(D)J recombination is mediated by an enzyme complex called ***V(D)J recombinase***, which recognizes recombination signal sequences in the DNA that flanks each gene segment to be joined. Although the process ensures that only appropriate gene segments recombine, a variable number of nucleotides are often lost from the ends of the recombining gene segments, and one or more randomly

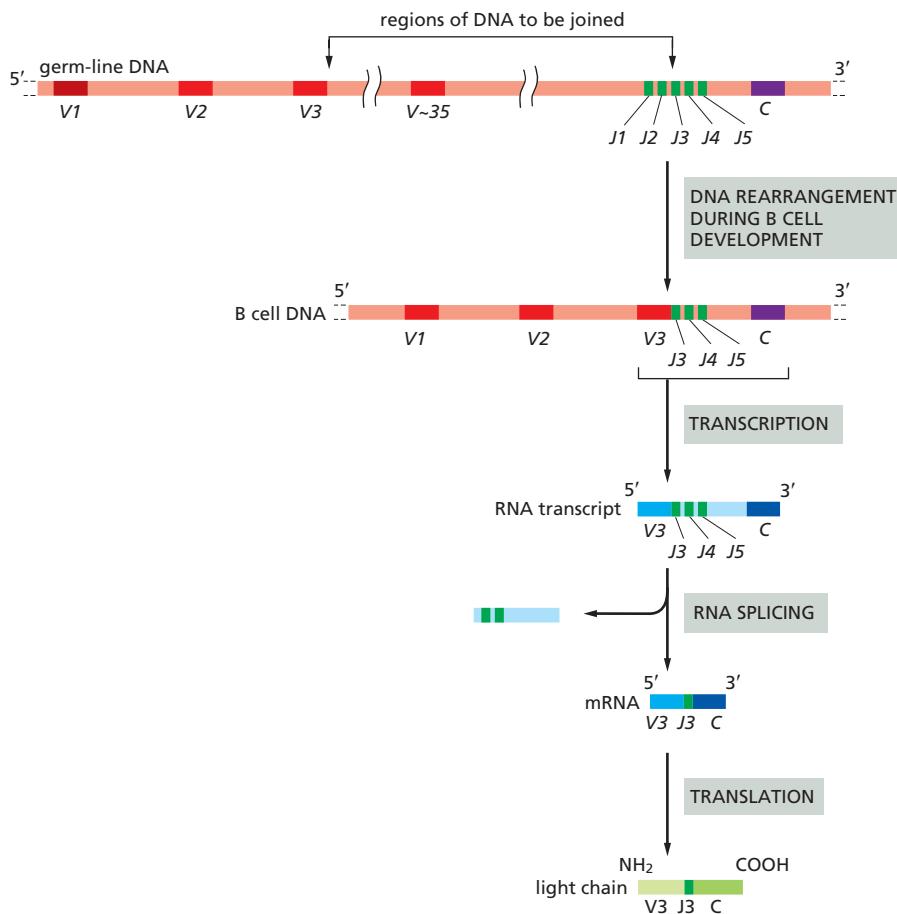


Figure 24–28 The *V-J* joining process involved in making a human *κ* light chain. In the “germ-line” DNA (where the Ig gene segments are not rearranged and are therefore not being expressed), the cluster of five *J* gene segments is separated from the *C*-region coding sequence by a short intron and from the 35 or so functional *V* gene segments by thousands of nucleotide pairs. During the development of a B cell, a randomly chosen *V* gene segment (*V*3 in this case) is moved to lie precisely next to one of the *J* gene segments (*J*3 in this case). The “extra” *J* gene segments (*J*4 and *J*5) and the intron sequence are transcribed (along with the joined *V*3 and *J*3 gene segments and the *C*-region coding sequence) and then removed by RNA splicing to generate mRNA molecules with contiguous *V*3, *J*3, and *C* sequences, as shown. These mRNAs are then translated into *κ* light chains. A *J* gene segment encodes the 15 or so C-terminal amino acids of the *V* region, and a short sequence containing the *V-J* segment junction encodes the third hypervariable region, which is the most variable part of the light-chain *V* region.

chosen nucleotides are also 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 (up to about 10^8 -fold), specifically in the third hypervariable region. This increased diversification comes at a price, however. In many cases, it shifts the reading frame to produce a nonfunctional gene, in which case the developing B cell fails to make a functional Ig molecule and consequently dies in the bone marrow. Once a B cell makes a functional heavy chain and light chain that form an antigen-binding site, it turns off the *V(D)J* recombination process, thereby ensuring that the cell makes Ig of only one antigen-binding specificity.

B cells making BCRs that bind strongly to self antigens in the bone marrow would be dangerous. Such B cells maintain expression of an active *V(D)J* recombinase and are activated by such self-binding to undergo a second round of *V(D)J* recombination in a light-chain locus, thereby changing the specificity of its BCR—the process of **receptor editing** discussed earlier; self-reactive B cells that fail to change their specificity die by apoptosis, in the process of clonal deletion (see Figure 24–21).

Antigen-Driven Somatic Hypermutation Fine-Tunes Antibody Responses

As mentioned earlier, with the passage of time following an infection or vaccination, there is usually a progressive increase in the affinity of the antibodies produced against the pathogen. This phenomenon of **affinity maturation** is due to

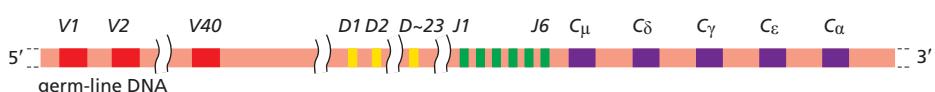


Figure 24–29 The human heavy-chain locus. There are 40 *V* segments, about 23 *D* segments, 6 *J* segments, and an ordered cluster of *C*-region coding sequences, each cluster encoding a different class of heavy chain. The *D* segment (and part of the *J* segment) encodes amino acids in the third hypervariable 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 24–28 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 rearrangements lead to the production of a *VDJC* mRNA that encodes a complete Ig heavy chain. The figure is not drawn to scale: the total length of the heavy-chain locus is over two megabases. Moreover, a number of details are omitted: for example, the exons encoding each *C*-region Ig domain and the hinge region (see Figure 24–27) and the different subclasses of *C_γ*-coding segments are not shown.

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 and form *germinal centers* (see Figure 24–20). 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, the process is called **somatic hypermutation**.

Very few of the altered Iggs generated by hypermutation will have an increased affinity for the antigen. But, because the same Ig genes produce both BCRs and secreted antibodies, the antigen will stimulate preferentially those few B cells that do make BCRs 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 adaptive immune response, providing progressively better protection against the pathogen ([Movie 24.6](#)).

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) during transcription of 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. Somatic hypermutation affects only actively transcribed DNA, because AID works only on single-stranded DNA (which is transiently exposed during transcription) and because proteins involved in the transcription of V-region coding sequences are required to recruit the AID enzyme. AID is also required for activated B cells to switch from IgM and IgD production to the production of the other classes of Ig, as we now discuss.

B Cells Can Switch the Class of Ig They Make

After a developing B cell leaves the bone marrow, before it interacts with antigen, it expresses both IgM and IgD BCRs on its surface, both with the same antigen-binding sites (see Figure 24–24). Stimulation by antigen and helper T cells activates many of these mature naïve B 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 (discussed later) stimulates many of the B cells to switch from making membrane-bound IgM and IgD to making IgG, IgE, or IgA, in the process of **class switching**. Some of these cells become memory cells that express the corresponding class of Ig as BCRs on their surface, while others become effector cells that secrete the Ig molecules as antibodies. The IgG, IgE, and IgA molecules retain their original antigen-binding site and are collectively referred to as *secondary classes* of Igs, because they are produced only after antigen stimulation, dominate secondary antibody responses, and make up the secondary Ig repertoire.

As discussed earlier, the constant region of an Ig heavy chain determines the class of the Ig. 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 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.

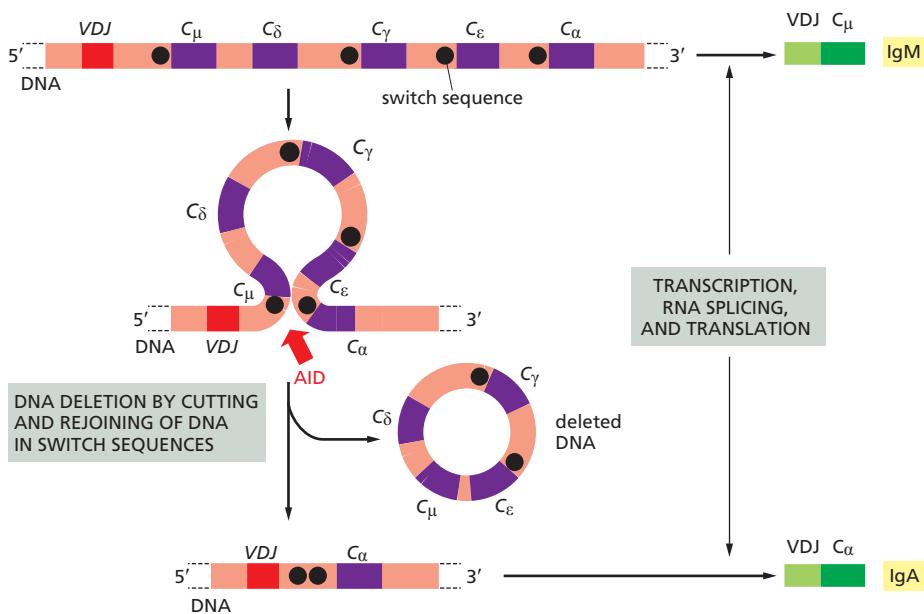


Figure 24–30 An example of the DNA rearrangement that occurs in class switch recombination. A B cell making IgM molecules with a V region encoded by a particular assembled *VDJ* DNA sequence is stimulated to switch to making IgA molecules with the same V region. In the process, it deletes the DNA between 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*, as B cells don't switch from *C_μ* to *C_δ*) can recombine with one another, with the deletion of the intervening DNA, as shown here. As discussed in the text, the recombination process depends on AID, the same enzyme that is involved in somatic hypermutation. When switching from IgM to IgG or IgE, the C-region coding sequences downstream of *C_γ* or *C_δ*, which remain after the DNA deletion, are removed during RNA splicing.

When a B cell switches from making IgM and IgD to one of the secondary classes of Ig, an irreversible change occurs in the DNA—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 flank the different *C_H*-coding segments. (3) It involves cutting and joining the switch sequences, which are noncoding sequences, and leaves the assembled *V_H*-region coding sequence unchanged (Figure 24–30). (4) Most importantly, the molecular mechanism is different. It depends on AID, which is also involved in somatic hypermutation, rather than on the V(D)J recombinase. The cytokines that activate class switching induce the production of transcription regulators that activate transcription from the relevant switch sequences, allowing the recruitment of AID to these sites.

Once bound, AID initiates switch recombination by deaminating some cytosines to uracil in the vicinity of these switch sequences. Excision of these uracils is thought to lead 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 Ig repertoire in humans (and mice) is generated by V(D)J joining mediated by V(D)J recombinase, the secondary antibody repertoire is generated by somatic hypermutation and class switch recombination, both of which are mediated by AID. Figure 24–31 lists the main mechanisms that we have discussed in this chapter that diversify Igs.

Summary

Each B cell clone makes Ig molecules with a unique antigen-binding site. Initially, the Ig molecules are inserted into the plasma membrane and serve as B cell receptors (BCRs) for antigen. Antigen binding to the BCRs, together with co-stimulatory signals from 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 antigen-binding site as the BCRs.

A typical Ig molecule is composed of four polypeptide chains—two identical heavy chains and two identical light chains. Parts of both the heavy and light chains form the two identical antigen-binding sites. There are multiple classes of Ig (IgA, IgD, IgE, IgG, and IgM), each with a distinctive heavy chain, which determines the

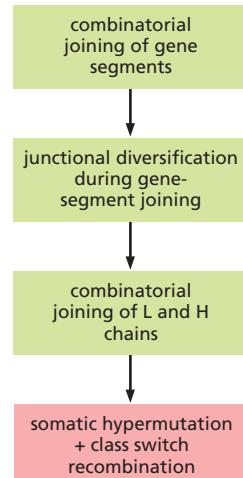


Figure 24–31 The main mechanisms of Ig diversification in mice and humans. Those shaded in green occur during B cell development in the bone marrow, whereas the two mechanisms shaded in red occur when B cells are stimulated by foreign antigen and helper T cells in germinal centers in peripheral lymphoid organs, either late in a primary response or in a secondary response.

biological properties of the Ig class. Each light and heavy chain is composed of a number of Ig domains. The amino acid sequence variation in the variable domains of both light and heavy chains is concentrated in several small hypervariable regions, which form loops at one end of these domains to produce the antigen-binding site.

Igs are encoded by loci on three different chromosomes, each of which is responsible for producing a different polypeptide chain—a κ light chain, a λ light chain, or a heavy chain. Each locus contains separate gene segments that encode different parts of the variable region of the particular Ig 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, before antigen stimulation, separate gene segments are brought together by site-specific recombination that depends on a V(D)J recombinase. 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 newly 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 Ig polypeptide chain.

By randomly combining inherited gene segments that code for the variable 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 domains come together in the final Ig molecule, the heavy and light chains can potentially pair to form Igs with millions of different antigen-binding sites. A loss or gain of nucleotides at the site of gene-segment joining increases this number enormously. The Igs made by such V(D)J recombination before antigen stimulation are IgMs and IgDs with low affinity for binding antigen, and they constitute the primary Ig repertoire.

Igs 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 together produce the high-affinity IgG, IgE, and IgA Igs that constitute the secondary Ig repertoire. The process of class switching allows the same antigen-binding site to be incorporated into antibodies that have different tails and therefore 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 have entered the interior of host cells, where they are invisible to B cells and antibodies. 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, a T cell is activated by foreign antigen to proliferate and differentiate into effector cells only when the antigen is displayed on the surface of an *antigen-presenting cell* (APC), usually a dendritic cell in a peripheral lymphoid organ. One reason T cells require APCs for activation is that the form of antigen they recognize is different from that recognized by the Igs produced by B cells. Whereas Igs can recognize antigenic determinants on the surface of pathogens and soluble folded proteins, for example, T cells can only recognize fragments of protein antigens that have been produced by partial proteolysis inside a host cell. As mentioned earlier, newly formed *MHC proteins* capture these peptide fragments and carry them to the surface of the host cell, where T cells can recognize them.

The second difference is that, once activated, effector T cells act mainly at short range, either within a secondary lymphoid organ or after they have migrated to 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 (if it is an infected host cell, for example) or signal in some way (if it is a B cell or macrophage, for example). We will refer to such host cells as *target cells*. As is the case with APCs, target cells must display an antigen bound to an MHC protein on their surface for a T cell to recognize them.

There are three main classes of T cells—cytotoxic T cells, helper T cells, and regulatory T cells. When activated, they function as effector cells (see Figure 24–17), each with their own distinct activities. 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 immune cells—mainly macrophages, dendritic cells, B cells, and cytotoxic T cells; as we will see, there are a variety of functionally distinct subtypes of helper T cells. Effector *regulatory T cells* suppress the activity of other immune cells.

In this section, we describe these classes and subclasses of T cells and their respective functions. We discuss how they recognize foreign antigens on the surface of APCs or target cells and the crucial part played by MHC proteins in the recognition process. We begin by considering the cell-surface receptors that T cells use to recognize antigen.

T Cell Receptors (TCRs) Are Ig-like Heterodimers

T cell receptors (TCRs), unlike IgGs made by B cells, exist only in membrane-bound form. They are composed of two transmembrane, disulfide-linked polypeptide chains, each of which contains two Ig-like domains—one variable and one constant. On most T cells, the TCRs have one α chain and one β chain (Figure 24–32).

The genetic loci that encode the α and β chains are located on different chromosomes. Like an Ig heavy-chain locus (see Figure 24–29), 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 antigen-binding site diversity of their TCRs as B cells use to generate antigen-binding site diversity of their IgGs, and they use the same V(D)J recombinase; thus, humans or mice deficient in this recombinase cannot make functional B or T cells. The mechanism that does not operate in TCR diversification is antigen-driven somatic hypermutation. Thus, the affinities of TCRs tend to be low ($K_a \approx 10^5$ – 10^7 liters/mole). Various co-receptors and cell-cell adhesion proteins, however, greatly strengthen the binding of a T cell to an APC or target cell.

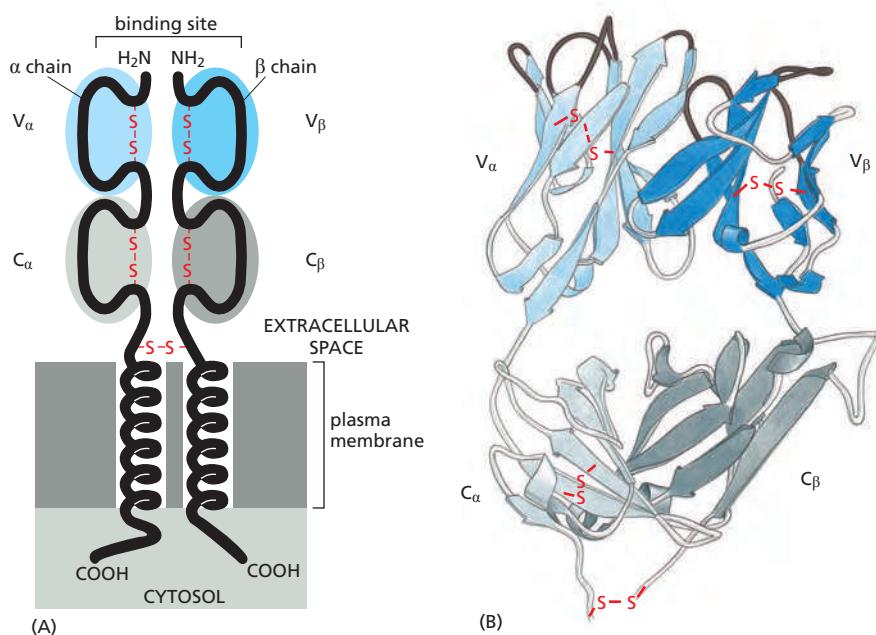


Figure 24–32 A T cell receptor (TCR) heterodimer. (A) Schematic drawing showing that the receptor is composed of an α and a β polypeptide chain. Each chain has a large extracellular part that is folded into two Ig-like domains—one variable (V) and one constant (C). A V_α and a V_β domain (shaded in blue) form the antigen-binding site. Unlike IgGs, which have two binding sites for antigen, TCRs have only one. The $\alpha\beta$ -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 their specific antigen (see Figure 24–45B). A typical T cell has about 30,000 TCRs 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 (black), and it is similar in its overall dimensions and geometry to the antigen-binding site of an Ig molecule. (B, based on K.C. Garcia et al., *Science* 274:209–219, 1996.)

Instead of making α and β chains, a minority of T cells makes a different but related type of TCR heterodimer, composed of γ chains and δ chains. Although these γ/δ T cells normally make up only 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). They have some properties in common with natural killer (NK) cells and with an enlarging category of T-like cells that have features of both innate and adaptive immune cells, which are sometimes collectively referred to as *innate lymphoid cells*. The cells in all these categories tend to be enriched in mucosal tissues, respond early to infection, display little immunological memory, and, compared with B and T cells, have surface receptors of restricted diversity. We will not discuss them further.

As with BCRs, 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. We will discuss these proteins in more detail later, when we consider some of the molecular events involved in T and B cell activation. First, we consider the special ways in which T cells recognize foreign antigen on the surface of an APC or target cell.

Activated Dendritic Cells Activate Naïve T Cells

Generally, naïve T cells, including naïve helper and cytotoxic T cells, proliferate and differentiate into effector cells and memory cells only when they see their specific antigen on the surface of an activated **dendritic cell** in a peripheral lymphoid organ (Figure 24–33). The activated dendritic cell displays the antigen in a complex with MHC proteins on its surface, along with co-stimulatory proteins (see Figure 24–11). The memory T cells that develop, however, can be activated by the same antigen-MHC complex on the surface of other types of APCs (target cells), including macrophages and B cells—as well as by dendritic cells.

Immature dendritic cells are located in most tissues—underlying epithelial layers of the skin and gut, for example—where they are constantly sampling and processing proteins in their environment. They become activated to mature when their pattern recognition receptors (PRRs) encounter pathogen associated molecular patterns (PAMPs) on an invading pathogen or its products. The pathogen or products are ingested, and the microbial proteins are cleaved into peptide fragments, which are loaded onto MHC proteins, as we discuss later. The activated dendritic cells then migrate via the lymph from the site of infection to local lymph nodes or gut-associated lymphoid organs, where they present the foreign antigens, displayed as peptide-MHC complexes on the dendritic cell surface, for recognition by the relevant T cells (see Figure 24–11).

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 memory cell (Figure 24–34): (1) **MHC proteins**, which present foreign peptides to the TCRs; (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 dendritic cell for long enough to become activated, typically several hours. In addition, activated dendritic cells secrete a variety of cytokines that influence the type of effector helper T cell that develops, and different types of dendritic cells promote different outcomes (discussed later).

T Cells Recognize Foreign Peptides Bound to MHC Proteins

MHC proteins capture and display peptide fragments of foreign proteins for presentation to T cells. There are two main classes of MHC proteins, which are structurally and functionally distinct. **Class I MHC proteins** mainly present foreign peptides to cytotoxic T cells, whereas **class II MHC proteins** mainly present foreign peptides to helper and regulatory T cells (Figure 24–35). Some class-I-like MHC proteins present microbial lipid and glycolipid antigens to T cells, but they are not encoded within the MHC region of the genome, and we will not consider them further.

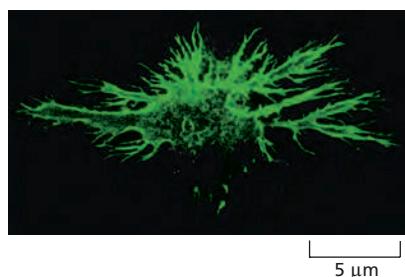


Figure 24–33 Immunofluorescence micrograph of a dendritic cell in culture. These APCs derive their name from their long processes, or “dendrites.” The cell has been labeled with a monoclonal antibody that recognizes a surface antigen on these cells. (Courtesy of David Katz.)

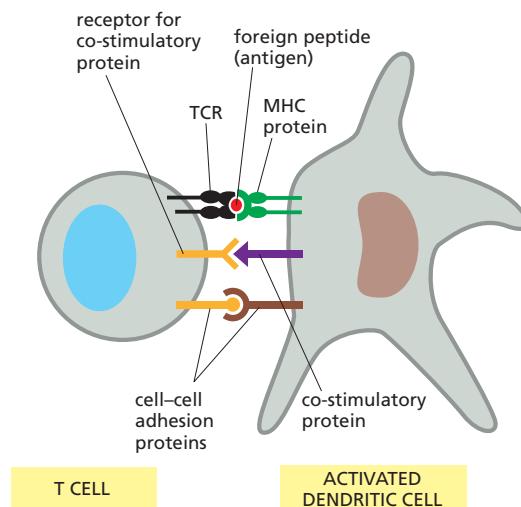


Figure 24–34 The three general types of proteins on the surface of an activated dendritic cell involved in activating a T cell. Although only membrane-bound co-stimulatory molecules are shown, activated dendritic cells also secrete soluble co-stimulatory molecules. The invariant polypeptide chains that are always stably associated with the TCR are not shown; they are illustrated in Figure 24–45B and [Movie 24.7](#).

Both class I and class II MHC proteins are heterodimers, in which two extracellular domains form a *peptide-binding groove*, which always has a variable small peptide bound in it. In class I MHC proteins, the two domains that form the peptide-binding groove are provided by the transmembrane α chain, which is noncovalently associated with a small subunit called β_2 -microglobulin; in class II MHC proteins, a different α chain and a large noncovalently associated β chain each contribute an extracellular domain to form the peptide-binding groove (Figure 24–36). A TCR binds to both the peptide and the ridges of the binding groove. Humans have three major class I proteins, called *HLA-A*, *HLA-B*, and *HLA-C*, and three class II proteins, called *HLA-DR*, *HLA-DP*, and *HLA-DQ* (*HLA* stands for *human-leukocyte-associated*, as these proteins were first demonstrated on human leukocytes). Figure 24–37 shows how the genes that encode these proteins are arranged on human chromosome 6.

There are important differences between the class I and class II MHC proteins with regard to the cell types that express them and the origin of the peptides in their peptide-binding grooves. Almost all of our nucleated cells express class I proteins. Their peptide-binding groove displays one of a diverse collection of peptides (typically 8–10 amino acids in length). In a healthy cell, the peptides originate from the cell's own cytosolic and nuclear proteins that have undergone partial degradation in proteasomes in the processes of normal protein turnover and quality control mechanisms. Some of the peptide fragments produced in this way are actively transported into the lumen of the endoplasmic reticulum (ER), through a specialized transporter in the ER membrane, where they are loaded onto newly synthesized class I MHC α chains; once a peptide binds, the α chain can assemble with its partner chain. The resulting self-peptide-MHC complex is then transported through the Golgi apparatus to the cell surface. Such complexes are not dangerous, however, because the cytotoxic T cells that could recognize

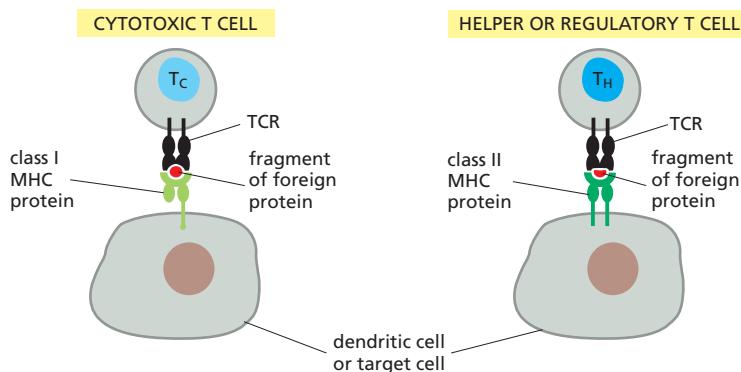


Figure 24–35 Recognition by T cells of foreign peptides bound to MHC proteins. Cytotoxic T cells recognize foreign peptides in association with class I MHC proteins, whereas helper T cells and regulatory T cells recognize foreign peptides in association with class II MHC proteins. In both cases, the T cell recognizes the peptide-MHC complexes on the surface of an APC—either a dendritic cell or a target cell. Some regulatory T cells recognize self peptides in association with class II MHC proteins (not shown).

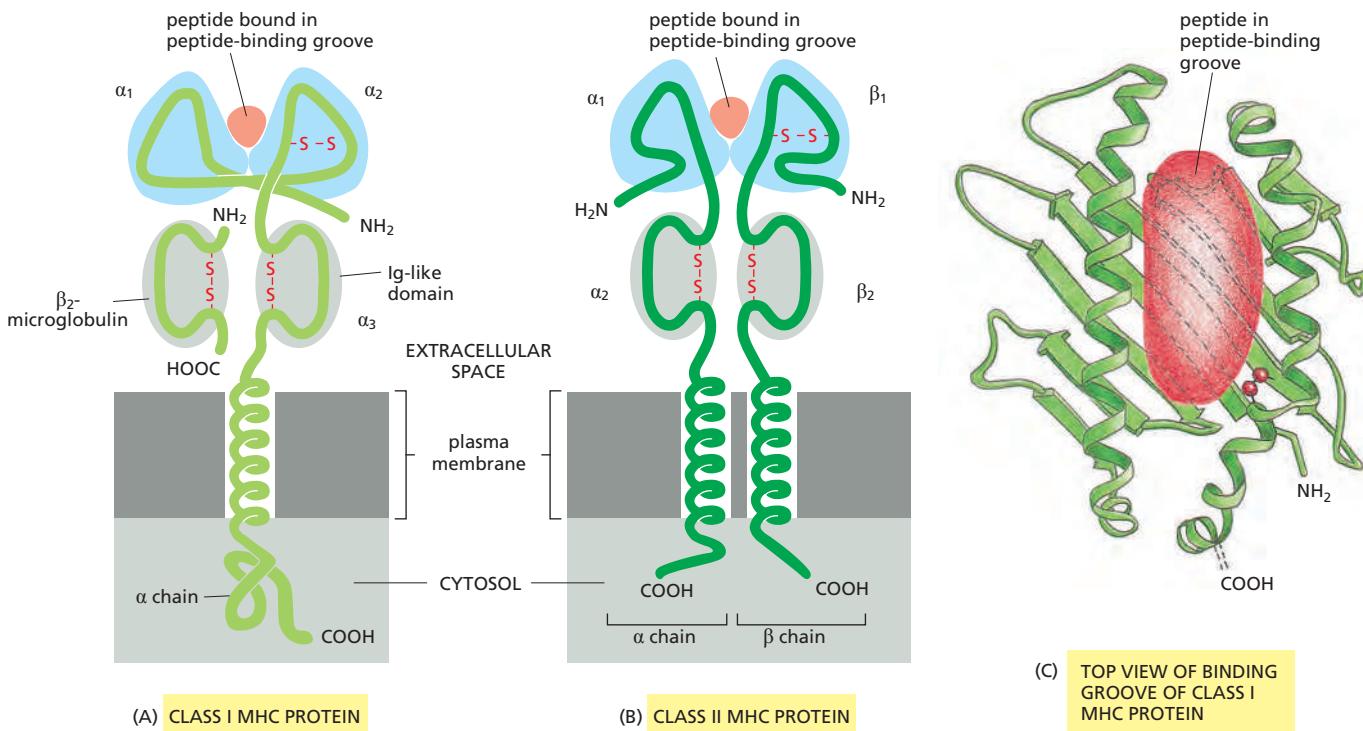
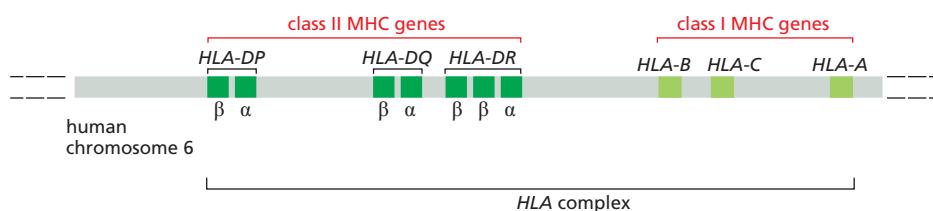


Figure 24-36 Class I and class II MHC proteins. (A) The α chain of the class I molecule has three extracellular domains, α_1 , α_2 , and α_3 , each encoded by a separate exon. The α chain is noncovalently associated with a smaller polypeptide chain, β_2 -microglobulin, which is not encoded within the MHC region of the genome. 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 the α chain and the β chain are encoded within the MHC and 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. (C) The three-dimensional structure of the peptide-binding groove of a human class I MHC protein is viewed from above, with bound peptide shown schematically; a peptide must be bound in the groove for the MHC protein to assemble and be transported to the cell surface. The sides of the groove are formed by two α helices, and the floor is formed by a β pleated sheet. The S-S disulfide bond is shown in red (Movie 24.8 and Movie 24.9). (C, adapted from P.J. Bjorkman et al., *Nature* 329:506–512, 1987. With permission from Macmillan Publishers Ltd.)

them have been either eliminated or inactivated, or suppressed by regulatory T cells in the process of self-tolerance. By contrast, in a cell infected by a pathogen such as a virus, the pathogen proteins will be processed in the same way, and peptides derived from them will be displayed on the infected cell surface bound to class I MHC proteins; there, they are recognized by cytotoxic T cells expressing the appropriate TCRs, thereby targeting the infected cell for destruction (Figure 24-38).

In general, only **antigen-presenting cells (APCs)** express class II MHC proteins. Dendritic cells are referred to as *professional APCs*, as they are specialized for this function and only they can activate naïve T cells. Other immune cells that are targets of effector T cell regulation, including B cells and macrophages, are *nonprofessional APCs*. All APCs load their newly synthesized class II MHC proteins with peptides derived mainly from extracellular proteins that are endocytosed and delivered to endosomes. The newly synthesized class II MHC proteins initially contain an *invariant chain*, which occupies the peptide-binding groove

Figure 24-37 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 MHC 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. Because of the extreme polymorphism of the MHC genes, the chances are very low that the maternal and paternal alleles will be the same. The number of class II MHC proteins that can be made is greater than six because there are two $DR \beta$ genes and because maternally encoded and paternally encoded polypeptide chains can sometimes pair. The entire region shown spans about seven million base pairs and contains other genes that are not shown.



and prevents it from prematurely binding a peptide until the class II MHC protein reaches specialized vesicles, which fuse with endosomes. Here, the invariant chain is removed and peptide fragments (typically 12–20 amino acids long) produced from endocytosed proteins can bind to the groove of the class II MHC proteins, which are then transported to the plasma membrane for display on the surface of the APC. In a healthy host cell, class II MHC protein grooves are loaded with self-peptides derived from normal proteins and will be ignored by T cells because of self-tolerance mechanisms. During an infection, however, pathogen proteins are also endocytosed and processed in the same way, enabling APCs to present pathogen peptides bound to class II MHC proteins to T cells expressing an appropriate TCR (Figure 24–39).

The distinction just discussed between the antigen-processing pathways for loading peptides onto class I and class II MHC proteins 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, specialized subsets of dendritic cells use a process called **cross-presentation**, which begins when these noninfected dendritic cells phagocytose virus-infected host cells or their fragments. The ingested viral proteins are then released by an unknown mechanism from phagolysosomes into the cytosol, where they are degraded in proteasomes; the resulting protein fragments are then transported into the ER lumen, where they load onto assembling class I MHC proteins. Cross-presentation in dendritic cells is not confined to endocytosed pathogens and their products: it also operates to activate cytotoxic T cells against tumor antigens of cancer cells and the MHC proteins of foreign organ grafts.

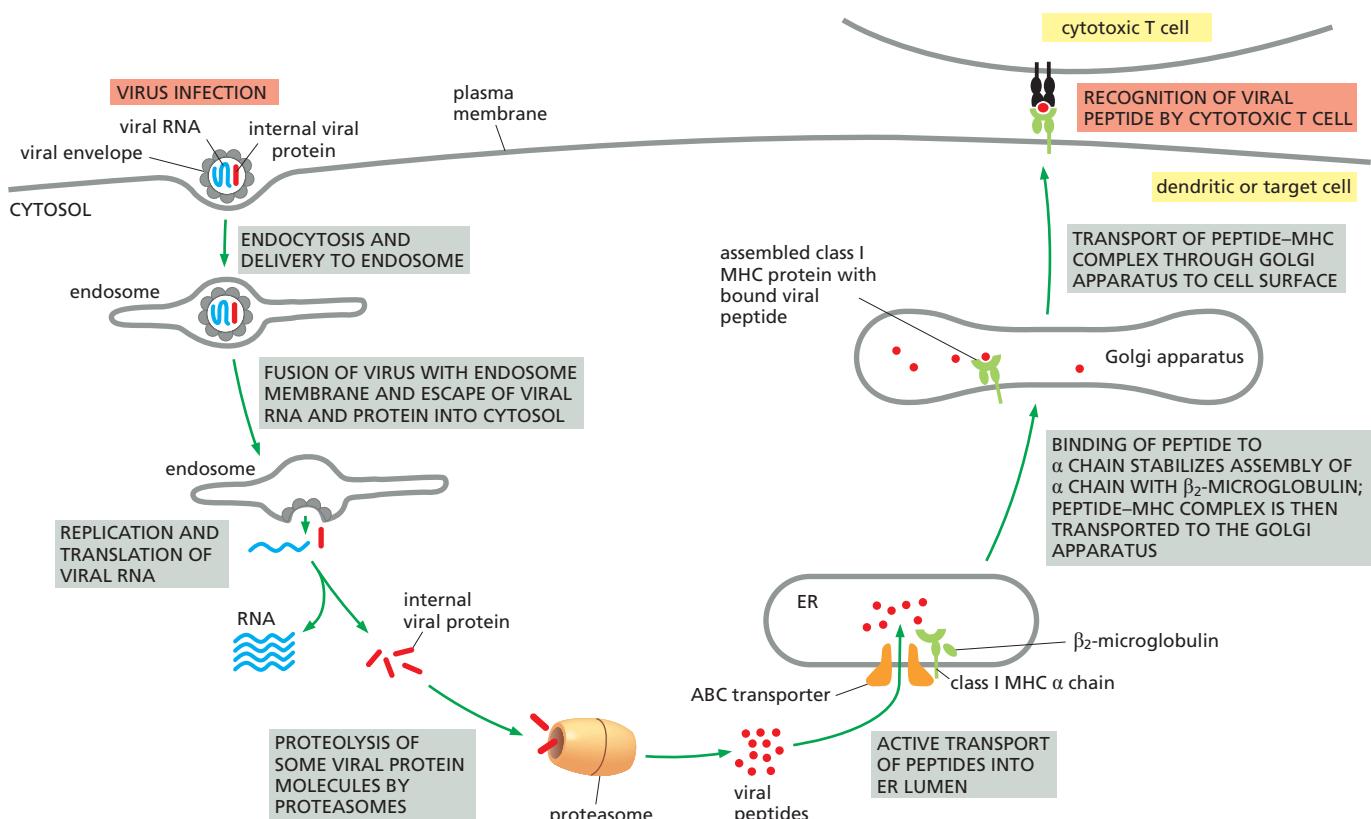


Figure 24–38 The processing of an extracellular foreign protein for presentation to cytotoxic T cells. An effector cytotoxic T cell kills a virus-infected cell when it recognizes fragments of an internal viral protein bound to class I MHC proteins on the surface of the infected cell. Not all viruses 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 in the ER lumen aid the folding and assembly of class I MHC proteins (not shown). The assembly of class I MHC proteins and their transport to the cell surface require the binding of either a self or foreign peptide (Movie 24.10).

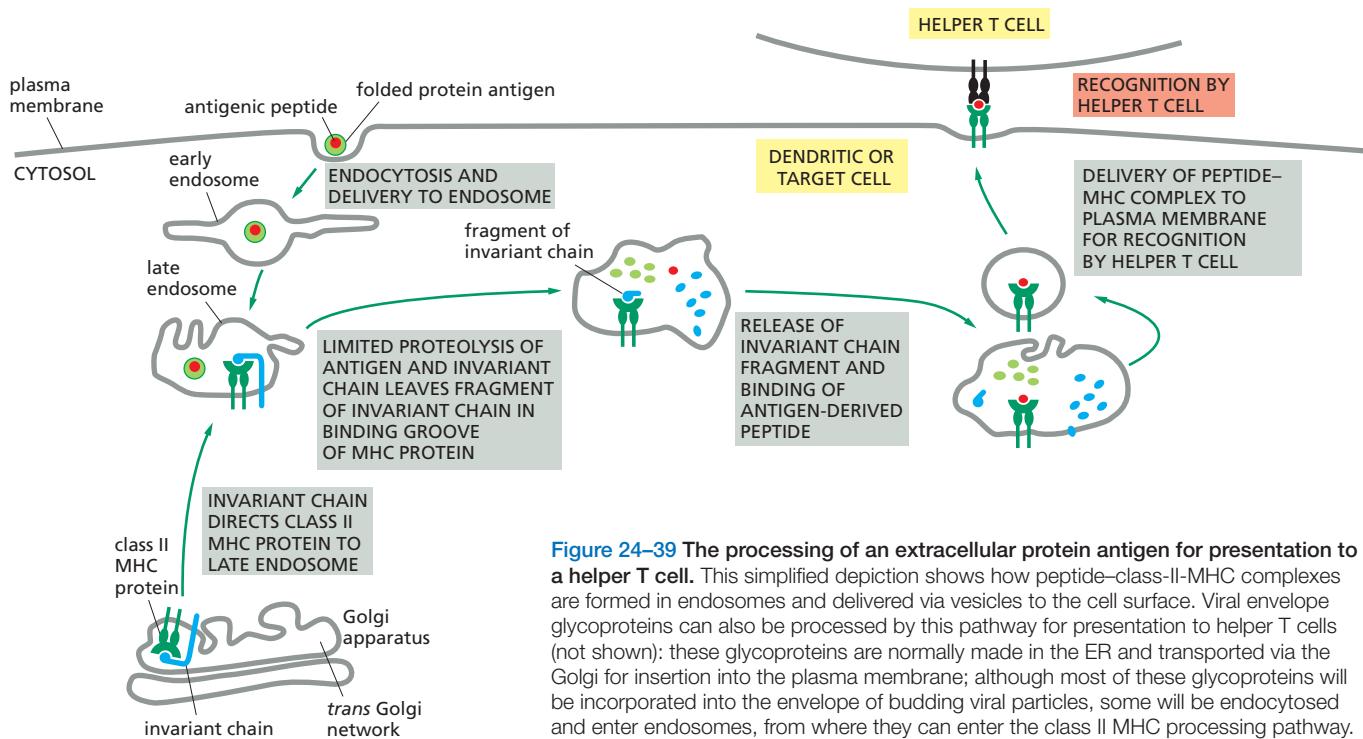


Figure 24–39 The processing of an extracellular protein antigen for presentation to a helper T cell. This simplified depiction shows how peptide–class-II–MHC complexes are formed in endosomes and delivered via vesicles to the cell surface. Viral envelope glycoproteins can also be processed by this pathway for presentation to helper T cells (not shown): these glycoproteins are normally made in the ER and transported via the Golgi for insertion into the plasma membrane; although most of these glycoproteins will be incorporated into the envelope of budding viral particles, some will be endocytosed and enter endosomes, from where they can enter the class II MHC processing pathway.

During an infection, only a small fraction of the many thousands of MHC proteins on the surface of an APC or target cell will have pathogen peptides bound to them. This is sufficient, however: fewer than 50 copies of such a peptide–MHC complex on a dendritic cell, for example, can activate a helper T cell that has a TCR that binds the complex with a high-enough affinity. **Table 24–3** compares the properties of class I and class II MHC proteins.

MHC Proteins Are the Most Polymorphic Human Proteins Known

Although any individual can make only a small number of different class I and class II MHC proteins, together, these proteins must be able to present peptide fragments from almost any foreign protein to T cells. Thus, unlike the antigen-binding site of an Ig protein, the peptide-binding groove of each MHC protein must be

TABLE 24–3 Properties of Human Class I and Class II MHC Proteins

	Class I	Class II
Genetic loci	<i>HLA-A, HLA-B, HLA-C</i>	<i>HLA-DP, HLA-DQ, HLA-DR</i>
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

able to bind a very large number of different peptides. The genes encoding class I and class II MHC proteins (see Figure 24–37) are the most *polymorphic* known in higher vertebrates: in the human population, for example, there are more than 2000 allelic variants of these genes. The corresponding variations in the MHC proteins are concentrated in the floor and walls of the peptide-binding grooves and allow MHC molecules in different individuals to bind different arrays of peptides.

It is thought that infectious diseases have been an important driving force for generating this remarkable MHC polymorphism. In the evolutionary war between pathogens and the adaptive immune system, pathogens will tend to change their proteins through mutation so that the peptides derived from them will not fit in the MHC peptide-binding grooves. When a pathogen succeeds, it can sweep through a population as an epidemic. In such circumstances, the few individuals who produce a new allelic form of MHC protein that can bind peptides derived from the altered pathogen will have a large selective advantage. This type of selection will tend to promote and maintain a large diversity of MHC proteins in the population. In West Africa, for example, individuals with a specific MHC allele (HLA-B53) have a reduced susceptibility to a severe form of malaria that is endemic there; although this allele is rare elsewhere, it is found in 25% of the West African population.

The extensive diversity of human MHC proteins is the main reason that individuals who receive a foreign organ transplant must be treated with strong immunosuppressive drugs to prevent the immunological rejection of the grafted organ. Of all the foreign proteins that the graft expresses, the MHC proteins are by far the most powerful stimulators of the recipient's T cells, which would rapidly destroy the graft if they were not prevented from doing so by such drugs. Foreign MHC proteins are powerful T cell stimulants because T cells respond to them in the same way they respond to self MHC proteins that have foreign peptides bound to them; for this reason, the proportion of a person's T cells that can specifically recognize any foreign MHC protein is relatively high.

CD4 and CD8 Co-receptors on T Cells Bind to Invariant Parts of MHC Proteins

The affinity of TCRs for peptide–MHC complexes on an APC 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 are invariant and do not bind to foreign peptides. Once bound to the surface of a dendritic cell, for example, a T cell increases the strength of the binding by activating an integrin adhesion protein (discussed in Chapter 19), which then binds more strongly to an Ig-like protein on the surface of the dendritic cell. This increased adhesion enables the T cell to remain bound long enough to become activated.

When an accessory receptor has a direct role in activating the T cell by generating its own intracellular signals, it is called a **co-receptor**. 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 MHC 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 24–40).

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. Thus, the recognition of class I MHC proteins by CD8 allows cytotoxic T cells to focus on any type of infected host cell, while the recognition of class II MHC proteins by CD4 allows helper and regulatory T cells to focus on the target immune cells that they help or suppress, respectively. The cytoplasmic tail of the CD4 and CD8 proteins is associated with a member of the Src family of cytoplasmic tyrosine kinases

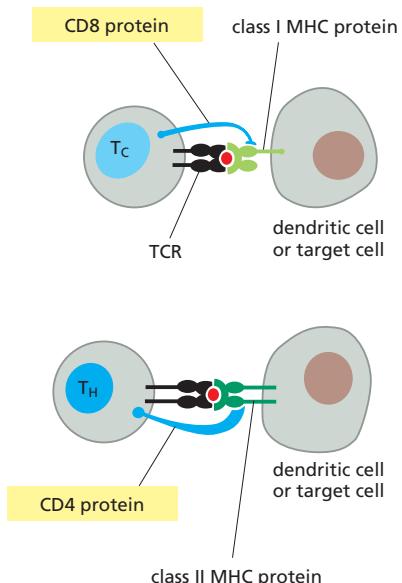


Figure 24–40 CD4 and CD8 co-receptors on the surface of T cells. Cytotoxic T cells (T_c) express CD8, which recognizes class I MHC proteins, whereas helper T cells (T_H) and regulatory T cells (not shown) express CD4, which recognizes class II MHC proteins. Note that the co-receptors bind to the same 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 binding groove.

(discussed in Chapter 15) called *Lck*, which phosphorylates various intracellular proteins on tyrosines and thereby participates in the activation of the T cell (discussed later).

The AIDS virus (HIV) uses CD4 molecules (as well as chemokine receptors) to enter helper T cells (see Figure 23–17). 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 anti-HIV drugs. HIV also uses CD4 and chemokine receptors to enter macrophages, which also have both types of receptors on their surface.

Developing Thymocytes Undergo Negative and Positive Selection

T cell development begins when bone-marrow-derived lymphoid progenitor cells enter the thymus from the bloodstream. There, the cells receive a variety of signals from thymus stromal cells, epithelial cells, macrophages, and dendritic cells, which promote their stepwise development into mature **thymocytes**. At one step, the progenitor cells are induced to express V(D)J recombinase and begin to rearrange their TCR gene segments. Soon thereafter, the cells express both CD4 and CD8 co-receptors, and these so-called *double-positive thymocytes* migrate inward and interact with thymus dendritic cells or epithelial cells expressing self peptides bound to class I and class II MHC proteins. If the TCR on the thymocyte binds with high affinity to these complexes, a strong signal will be transmitted, causing the cell to undergo apoptosis. This process, called **negative selection**, is an example of clonal deletion (see Figure 24–21), and it eliminates thymocytes that could potentially attack normal host cells and tissues and thereby cause an autoimmune disease if the cells were to continue to mature and leave the thymus.

If its TCR is unable to bind at all to a self-peptide-MHC complex in the thymus, the thymocyte will fail to receive the signals it needs to survive and will die of “neglect;” without the ability to recognize self-MHC proteins, a T cell would generally be of no use, as T cells can only see pathogen-derived peptides in the context of self-MHC proteins. Thymocytes that express a TCR that binds with an appropriate affinity to a self peptide bound to either a class I MHC protein (using CD8 as a co-receptor) or a class II MHC protein (using CD4 as a co-receptor) will receive an optimal signal to survive and continue to mature, a process called **positive selection** (Figure 24–41). As part of this maturation process, and depending on the TCR’s preference for class I or class II MHC proteins, the CD4 or CD8 co-receptor that is not needed is silenced by DNA methylation of the respective gene; this results in the development of CD4 or CD8 *single-positive thymocytes*, which exit the thymus as *naïve T cells* and enter the recirculating pool of T cells—the CD4 cells as either helper or regulatory T cells and the CD8 cells as cytotoxic T cells.

Although naïve helper and cytotoxic T cells constantly receive survival signals in the form of self peptides bound to MHC proteins that the T cells bind weakly, a T cell is only activated to proliferate and mount an immune response if its TCR binds with high affinity to a peptide-MHC complex and receives co-stimulatory signals at the same time. Generally, this happens only when the T cell encounters an activated dendritic cell (in a peripheral lymphoid organ) that expresses an MHC protein with a foreign peptide derived from a pathogen in its binding groove. Only then will the naïve T cell proliferate and differentiate into an effector or memory T cell.

Negative selection in the thymus is a major mechanism for ensuring that peripheral T cells do not react with host cells expressing MHC proteins with peptides derived from self proteins in their peptide-binding grooves. This mechanism, however, requires that the APCs in the thymus display an array of peptides on their MHC molecules that will reflect the self proteins in peripheral tissues, as well as in the thymus. The thymus, however, would not be expected to produce many of the proteins that are specifically expressed in other organs. As an example, it would not be expected to produce insulin, and yet it is crucial to delete thymocytes with TCRs that could recognize insulin-derived peptides bound to MHC

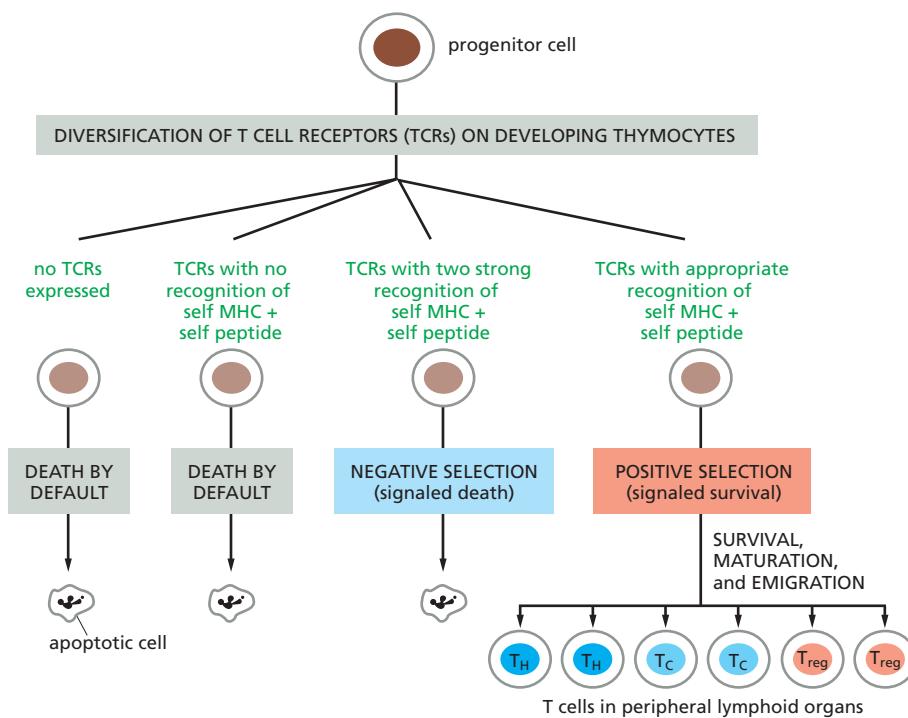


Figure 24–41 Positive and negative selection in the thymus. Developing thymocytes with TCRs that would potentially enable them to respond to peptides in association with self MHC proteins after they leave the thymus are positively selected: the binding of their TCRs to self peptides bound to self MHC proteins in the thymus signals such cells to survive, mature, and migrate to peripheral lymphoid organs. All of the other thymocytes 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.

The regulatory T cells (T_{reg} cells) that are positively selected in the thymus are called *natural T_{reg} cells* to distinguish them from *induced T_{reg} cells*, which develop in peripheral lymphoid organs from naïve helper T cells (T_H cells), as we discuss shortly.

proteins on the surface of insulin-secreting β cells in the pancreas. Any failure to do so would result in the T-cell-dependent destruction of the β cells and, as a consequence, cause *type 1 (or juvenile) diabetes*.

The mechanism that enables the deletion of all such cells in the thymus depends on a subpopulation of epithelial cells in the thymus that express a transcriptional regulator called **AIRE** (autoimmune regulator). By a poorly understood mechanism, the AIRE protein promotes the production of small amounts of mRNA from many genes that encode such “organ-specific” proteins, including the insulin gene. When the peptides derived from the proteins encoded by these genes are bound by MHC proteins and displayed on the surface of the epithelial cells in the thymus medulla, this is sufficient to provoke the deletion of the potentially self-reactive thymocytes. Mutations that inactivate the *AIRE* gene cause a severe multiorgan autoimmune disease in both mice and humans, indicating the importance of AIRE in self-tolerance.

Cytotoxic T Cells Induce Infected Target Cells to Kill Themselves

Cytotoxic T cells (T_C cells), like the NK cells discussed earlier, protect us against intracellular pathogens, including viruses, bacteria, and parasites, that multiply in the cytoplasm of a host cell. T_C cells kill infected host cells before the pathogen can escape to infect neighboring host cells. Before it can kill, however, a naïve T_C cell has to become an effector cell by activation on an APC, usually an activated dendritic cell that has pathogen-derived peptides bound to class I MHC proteins—a process that depends on helper T cells. The effector T_C cell can then recognize any target cell harboring the same pathogen and expressing some of the same peptide-MHC complexes on its surface: its TCRs cluster, along with CD8 co-receptors, adhesion molecules, and intracellular signaling proteins (discussed later), at the interface between the two cells, forming an **immunological synapse**. In this process, the effector T_C cell reorganizes its cytoskeleton to focus its killing apparatus on the target cell, secreting its toxic proteins into a confined space (**Figure 24–42**); in this way, it avoids killing neighboring cells. A similar synapse forms when an effector helper T cell interacts with its target cell, except that the co-receptor is CD4 (**Movie 24.11**).

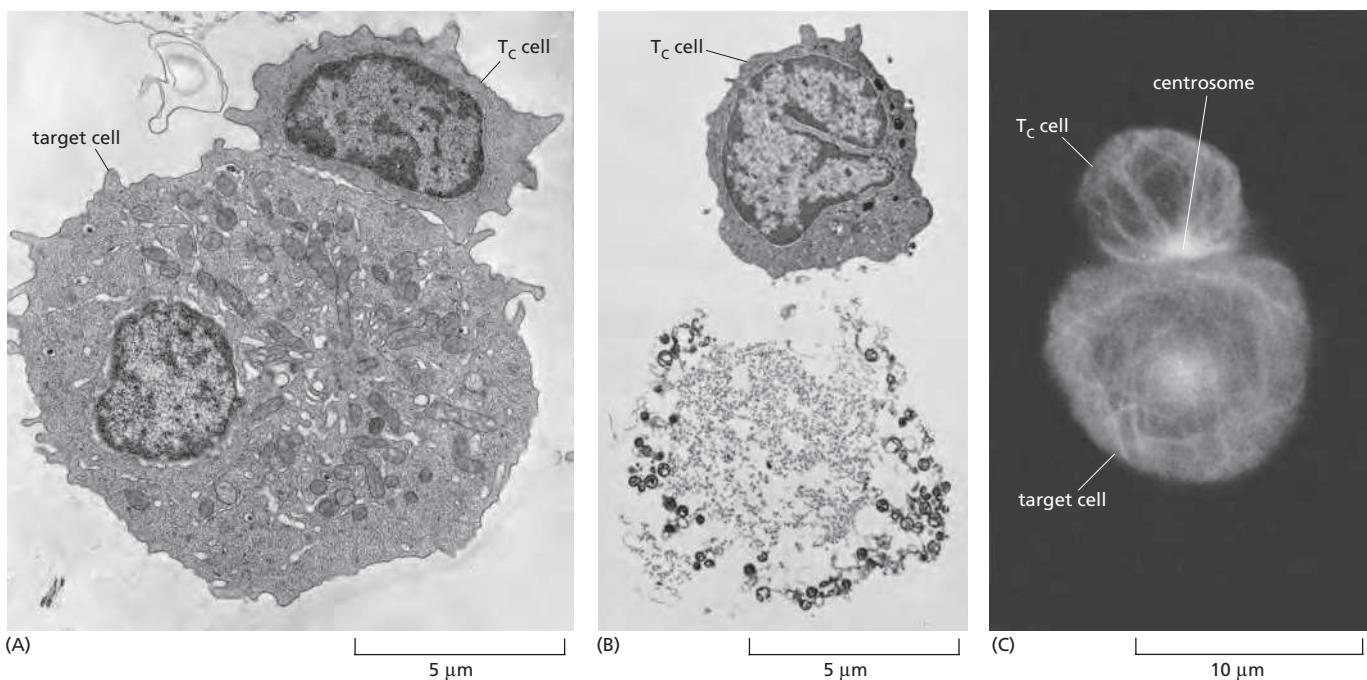


Figure 24–42 Effector cytotoxic T cells killing target cells in culture. (A) Electron micrograph showing an effector cytotoxic T cell (T_c cell) binding to a target cell. The T_c cells were obtained from mice immunized with the target cells, which are foreign tumor cells. (B) Electron micrograph showing a T_c cell and a tumor cell that the T_c cell has killed. In an animal, as opposed to a culture dish, the killed target cell would be phagocytosed by neighboring cells (especially macrophages) long before it disintegrates in the way that it has here. (C) Immunofluorescence micrograph of a T_c cell and tumor cell after immunofluorescence staining with anti-tubulin antibodies. Note that the centrosome in the T_c 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_c 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. (A and B, from D. Zagury et al., *Eur. J. Immunol.* 5:818–822, 1975. With permission from John Wiley & Sons, Inc; C, from B. Geiger, D. Rosen and G. Berke, *J. Cell Biol.* 95:137–143, 1982. With permission from the authors.)

An effector T_c cell (or an NK cell) can employ one of two strategies to kill the target, both of which operate by inducing the target cell to activate caspases and kill itself by undergoing apoptosis. One mechanism uses a protein called *Fas ligand* on the killer-cell surface, which binds to a transmembrane receptor protein called *Fas* on the target cell; this mechanism is discussed in Chapter 18 (see Figure 18–5). The other mechanism is the main one used by both NK cells and T_c cells to kill an infected target cell. The killer cell stores various toxic proteins within secretory vesicles in its cytoplasm that it releases into the synaptic space by exocytosis. The toxic proteins include *perforin* and proteases called *granzymes*. The perforin is homologous to complement component C9 and polymerizes in the target-cell plasma membrane (see Figure 24–8), forming a transmembrane pore that disrupts the membrane and allows the granzymes to enter the target cell. Once in the cytosol, the granzymes help activate caspases, thereby inducing apoptosis (Figure 24–43).

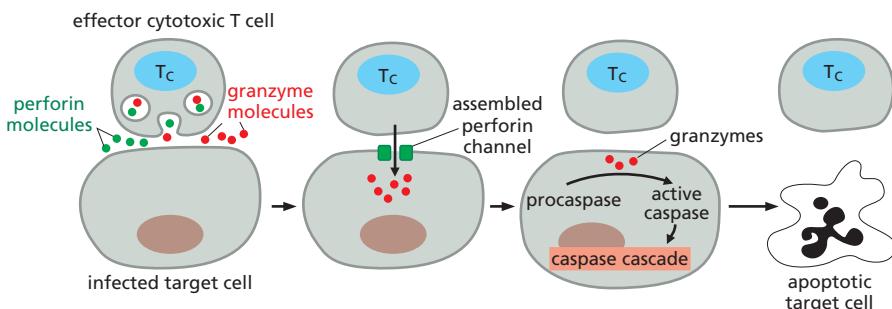


Figure 24–43 The main way that an effector T_c cell (or NK cell) kills an infected target cell. This simplified drawing shows how the killer cell releases perforin and granzymes onto the surface of an infected target cell by localized exocytosis at an immunological synapse. The high concentration of Ca²⁺ in the extracellular fluid causes the perforin to assemble into transmembrane channels in the target-cell plasma membrane, allowing the granzymes to enter the target-cell cytosol. The granzymes cleave and activate pro-caspases to initiate a caspase cascade, leading to apoptosis (see Figure 18–3). A single cytotoxic cell can kill multiple target cells in sequence. It remains a mystery why the released perforins do not form pores in the plasma membrane of the killer cell itself (Movie 24.12 and Movie 24.13).

Effector Helper T Cells Help Activate Other Cells of the Innate and Adaptive Immune Systems

In contrast to T_C cells, **helper T cells (T_H cells)** are crucial for defense against both extracellular and intracellular pathogens, and they express CD4 rather than CD8 co-receptors and recognize foreign peptides bound to class II rather than class I MHC proteins. Once naïve T_H cells are induced on activated dendritic cells to become effector cells, they can help activate other cells: they help activate B cells to become antibody-secreting cells and later to undergo Ig class switching and somatic hypermutation; they help activate macrophages to destroy any intracellular pathogens multiplying within the macrophage's phagosomes; they help induce naïve T_C cells to become effector cells that can kill infected target cells; and they stimulate the activated dendritic cell that activated them to maintain the dendritic cell in an activated state. In each case, the effector T_H cell recognizes the same complex of foreign peptide and class II MHC protein on the target-cell surface that it initially recognized on the activated dendritic cell. As discussed later, the T_H cell stimulates the target cell both by secreting a variety of cytokines and by displaying co-stimulatory proteins on its surface.

Naïve Helper T Cells Can Differentiate Into Different Types of Effector T Cells

When activated by binding to a foreign peptide bound to a class II MHC protein on an activated dendritic cell, a naïve T_H cell can differentiate into several distinct types of effector T cells, depending on the nature of the pathogen and the cytokines they encounter. These cells include four subtypes of helper cells— T_{H1} , T_{H2} , T_{FH} , and T_{H17} cells—and regulatory (suppressor) T cells. **Figure 24–44** summarizes both the cytokines that induce these effector T cells and some of the cytokines the effector cells secrete, as well as the master transcription regulators that control the effector cell's development.

Naïve T_H cells activated by dendritic cells secreting the cytokine interleukin-12 ($IL12$) develop into **T_{H1} cells**. These effector cells produce *interferon-γ (IFNγ)*,

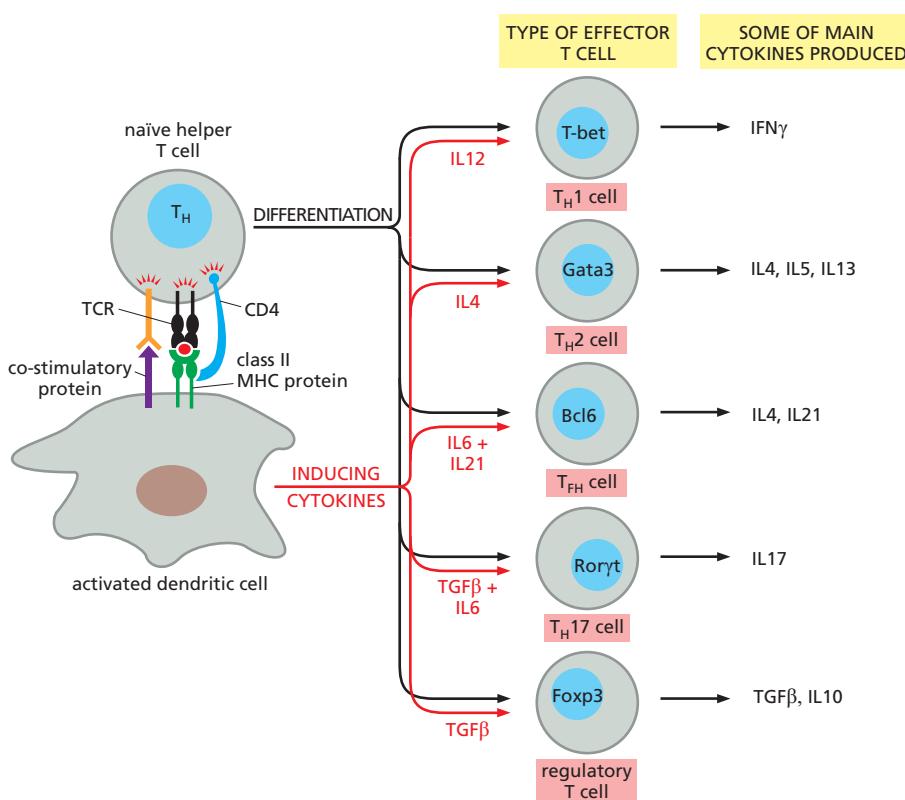


Figure 24–44 Differentiation of naïve helper T cells into different types of effector helper cells or regulatory T cells in a peripheral lymphoid organ. The cytokines produced by the activating dendritic cell (and by other cells in the environment) mainly determine which type of effector T cell develops, as indicated. Some of the main cytokines produced by each type of effector cell are also shown, and the master transcription regulator for each subset is indicated in the nucleus. There is increasing evidence that some of the effector cells are plastic and can change the cytokines they produce in response to changes in their environment (not shown).

which is critical for the activation of macrophages to destroy pathogens that either invaded the macrophage or were ingested by it; the IFN γ can also induce B cells to switch the class of Ig they are making. Naïve T_H cells activated in the presence of *IL4* develop into **T_H2 cells**. These effector cells are important for the control of extracellular pathogens, including parasites. They stimulate B cells to undergo somatic hypermutation and to switch the class of Ig they produce: for example, the T_H2 cells themselves produce IL4, which can induce B cells to switch from making IgM and IgD to making IgE antibodies, which can bind to mast cells, as discussed earlier. Naïve T_H cells activated in the presence of *IL6* and *IL21* develop into **follicular helper T cells (T_{FH})**, which are located in lymphoid follicles and secrete a variety of cytokines, including IL4 and IL21; these cells are especially important for stimulating B cells to undergo Ig class switching and somatic hypermutation. Naïve T_H cells activated in the presence of *IL6* and *TGF β* develop into **T_H17 cells**. These effector cells secrete *IL17*, which recruits neutrophils and stimulates epithelial cells and fibroblasts in the skin and gut to produce pro-inflammatory cytokines. T_H17 cells are important in controlling extracellular bacterial and fungal infections and in wound healing, but they can also have a major role in autoimmune diseases and allergy.

In some cases, naïve T_H cells that encounter their antigen in a peripheral lymphoid organ in the presence of TGF β and the absence of IL6 develop into **induced regulatory T cells (T_{reg} cells)**, which suppress rather than help immune cells; as mentioned earlier, **natural T_{reg} cells** develop in the thymus during thymocyte development (see Figure 24–41). In either case, the T_{reg} cells suppress the development, activation, or function of most other types of immune cells, by means of both secreted suppressive cytokines such as *IL10* and TGF β and inhibitory proteins on the T_{reg} cell surface. Induced T_{reg} cells seem mainly to suppress immune responses to foreign antigens—preventing responses to harmless ingested or inhaled antigens and limiting responses against pathogens to avoid excessive responses that cause unwanted pathology; natural T_{reg} cells are needed to prevent immune responses to self molecules (see Figure 24–21). T_{reg} cells express the transcription regulator *FoxP3*, which serves as both a marker of these cells and a master controller of their development: if the gene encoding this protein is inactivated in mice or humans, the individuals fail to produce T_{reg} cells and develop a fatal autoimmune disease involving multiple organs—findings that establish the crucial importance of T_{reg} cells in self-tolerance.

Both T and B Cells Require Multiple Extracellular Signals For Activation

Foreign antigen binding to BCRs or TCRs initiates the process whereby the T and B cells are stimulated to proliferate and differentiate into effector or memory cells. As mentioned earlier, these antigen receptors do not act on their own: they are stably associated with invariant transmembrane polypeptide chains that are required to relay the signal into the cell. In B cells, these are called *Ig α* and *Ig β* (Figure 24–45A), while in T cells they exist in a complex called CD3, composed of four types of polypeptide chains (Figure 24–45B). In both cases, the associated proteins help convert extracellular antigen binding to the TCR or BCR into intracellular signals, and they do so in similar ways.

Antigen binding to BCRs or TCRs clusters these receptors and their associated invariant chains (and CD4 or CD8 co-receptors in the case of TCRs). This clustering activates a Src family cytoplasmic tyrosine kinase to phosphorylate tyrosines on the cytoplasmic tails of some of the invariant chains. The phosphotyrosines then serve as docking sites for a second cytoplasmic tyrosine kinase, which becomes phosphorylated and activated by the first kinase; the second kinase then relays the signal downstream by phosphorylating other intracellular signaling proteins on tyrosines. Some of these early events in the signaling pathway activated by BCRs are shown in Figure 24–46.

Signaling through BCRs or TCRs and their associated proteins alone is not sufficient to activate a lymphocyte to proliferate and differentiate. Extracellular

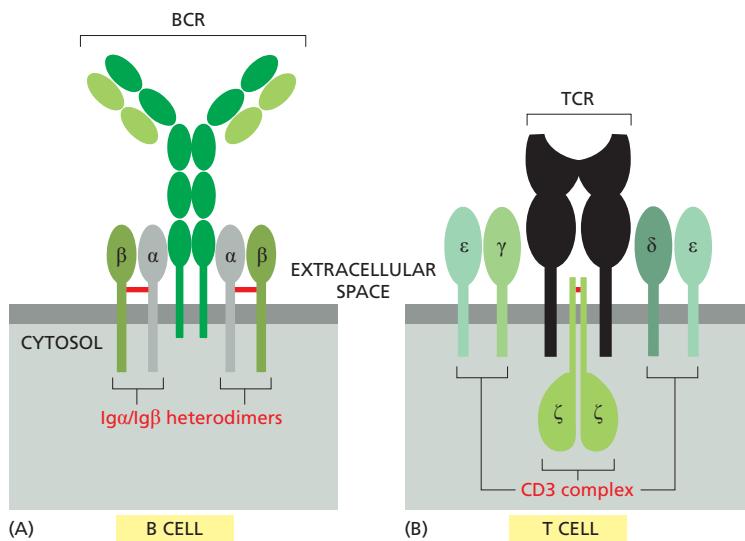


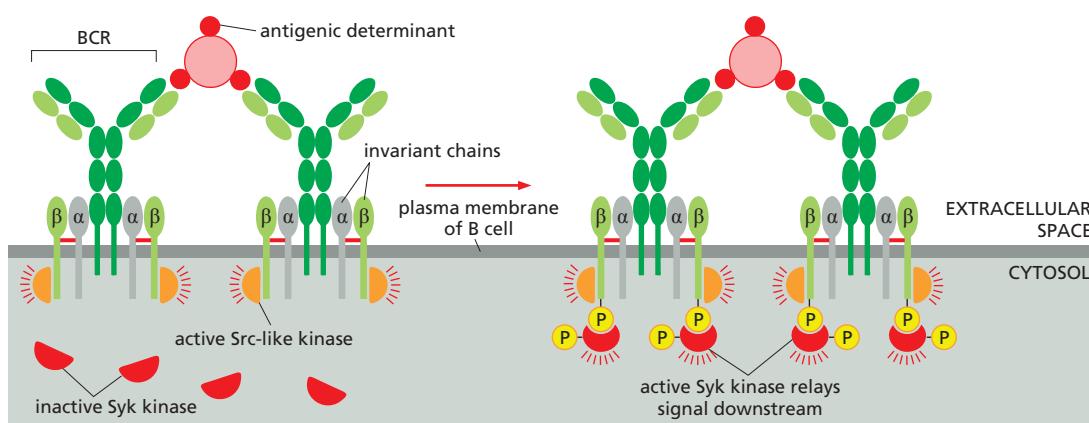
Figure 24–45 The invariant chains associated with BCRs and TCRs. (A) Each BCR is associated with two invariant heterodimers, each composed of an Ig α and an Ig β polypeptide chain linked by a sulfide bond (red). (B) Each TCR is associated with an invariant CD3 complex composed of two disulfide-bonded ζ chains, two ϵ chains, and one δ and one γ chain; these chains form homodimers or heterodimers, as shown.

co-stimulatory signals produced by another cell are also required, and they are provided by membrane-bound proteins (see Figure 24–34) and secreted cytokines. Indeed, signaling through the BCR or TCR with insufficient co-stimulation can either eliminate the lymphocyte (clonal deletion) or inactivate it, with both of these mechanisms contributing to self-tolerance (see Figure 24–21). For a naïve T cell, an activated dendritic cell provides the co-stimulatory signals; these include the transmembrane *B7 proteins*, which are recognized by the co-receptor protein *CD28* on the surface of the T cell (Figure 24–47A). For a B cell, an effector T_H cell provides the co-stimulatory signals; these include the transmembrane *CD40 ligand*, which binds to *CD40 receptors* on the B cell (Figure 24–47B). The CD40 ligand on effector T_H cells acts in two other situations: (1) it 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; and (2) it acts as a co-stimulatory signal on the surface of an effector T_{H1} cell, allowing the T cell to help activate an infected macrophage to destroy the pathogens it harbors.

In addition to receptors for co-stimulatory proteins, both B and T cells have inhibitory proteins on their surface that help regulate the cell's activity, preventing excessive or inappropriate responses. Two such proteins expressed by T cells have attracted great attention because of their roles in suppressing the ability of T cells to inhibit cancer progression: CTLA4 and PD1 proteins inhibit T cell activity in different ways, and monoclonal antibodies against either or especially both can relieve the inhibition and allow T cells to dramatically destroy the tumors in some patients with metastatic cancer (see Figure 20–45).

Figure 24–46 Early signaling events in a B cell activated by the binding of specific foreign antigen to its BCRs. If the antigen is on the surface of a pathogen or is a soluble macromolecule with two or more identical antigenic determinants (as shown), it cross-links adjacent BCRs, causing them and their associated invariant chains to cluster, as shown. A Src-like cytoplasmic tyrosine kinase (which can be *Fyn* or *Lyn*) is associated with the cytosolic tail of Ig β ; it joins the cluster and phosphorylates both the Ig α and Ig β invariant chains (for simplicity, only the phosphorylation on Ig β is shown). A transmembrane protein tyrosine phosphatase called *CD45* is also required to remove inactivating phosphates from these Src-like kinases (not shown). The resulting phosphotyrosines on Ig α and Ig β serve as docking sites for another Src-like tyrosine kinase called *Syk*, which becomes phosphorylated and thereby activated to relay the signal downstream.

The pathway from TCRs is similar (including a requirement for CD45), except that the first Src-like kinase is *Lck*, which is associated with a CD4 or CD8 co-receptor and phosphorylates tyrosines on all the CD3 polypeptide chains shown in Figure 24–45B; the second Src-like kinase is *ZAP70*, which is homologous to the *Syk* kinase in B cells (Movie 24.14).



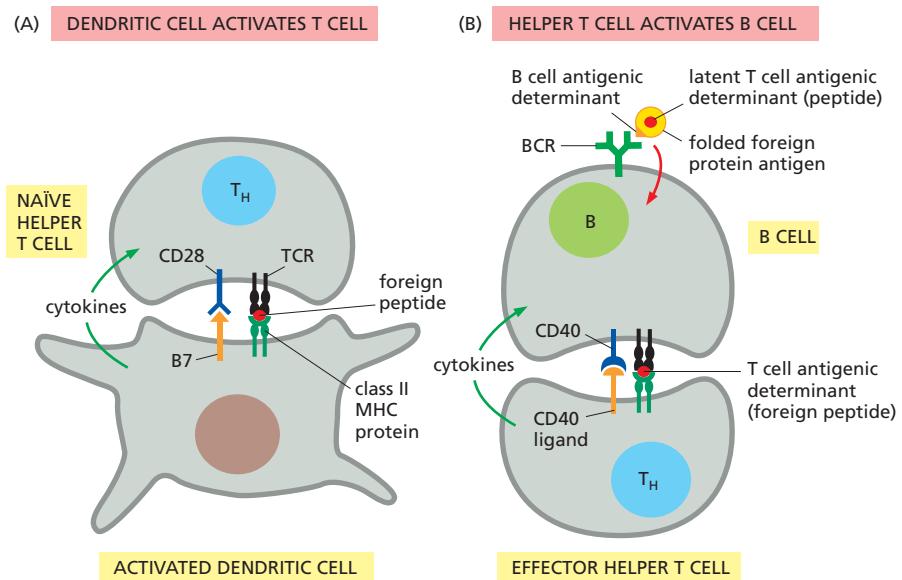


Figure 24-47 Comparison of the co-stimulatory proteins required to activate a helper T cell and a B cell in response to the same foreign protein. (A) A naïve helper T cell is activated by a peptide fragment of a foreign protein bound to a class II MHC protein on the surface of an activated dendritic cell. The co-stimulatory protein on the dendritic cell (a B7 protein—either CD80 or CD86) binds to the CD28 co-receptor on the T cell, providing a necessary co-stimulatory signal to the T cell; in addition, cytokines secreted by the dendritic cell (or other nearby cells) influence what subtype of effector helper cell the T cell becomes (see Figure 24-44). (B) Once activated to become an effector cell, the helper T cell can help activate B cells that have the same peptide–MHC protein complexes on their surface as the dendritic cell that activated the T cell. These B cells have BCRs that bind an antigenic determinant on the surface of a folded foreign protein and endocytose the protein (red arrow); the protein is then cleaved into peptides, which are carried to the B cell surface by class II MHC proteins, where some of them can be recognized by the TCRs on the helper T cell (see Figure 24-39). Note that the BCRs and TCRs recognize different antigenic determinants of the protein. As indicated, the co-stimulatory protein used by the effector helper T cell is CD40 ligand, which binds to the CD40 co-receptor on the B cell; the T cell also secretes cytokines such as IL4 to help stimulate the B cell to undergo somatic hypermutation and class switching (not shown). The CD4 co-receptor on T_H cells is omitted in both (A) and (B) for simplicity.

Many Cell-Surface Proteins Belong to the Ig Superfamily

Most of the proteins that mediate antigen recognition and cell-cell recognition in the immune system contain one or more Ig or Ig-like domains, suggesting that the proteins have a common evolutionary history. Included in this very large **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 leukocytes. Many of these proteins are dimers or higher oligomers, in which Ig or Ig-like domains of one chain interact with those in another (Figure 24-48).

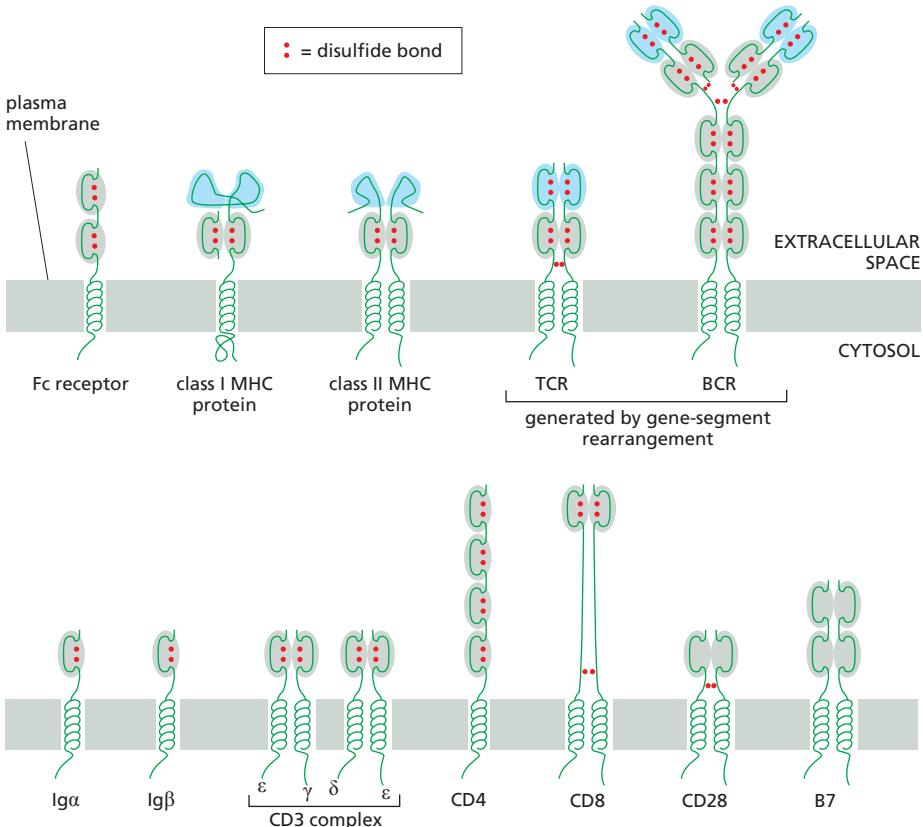


Figure 24-48 Some of the cell-surface proteins discussed in this chapter that belong to the Ig superfamily. The Ig and Ig-like domains are shaded in gray, except for the antigen-binding domains (not all of which are Ig domains—the class I and class II MHC proteins are the exception), which are shaded in blue. 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 (N-CAM) 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.

In both vertebrates and invertebrates, many proteins in the Ig superfamily are also found outside immune systems, where they often function in cell-cell recognition and adhesion processes, both during development and in adult tissues. It seems likely that the entire gene superfamily evolved from a primordial gene coding for a single Ig-like domain, similar to that encoding β_2 -microglobulin (see Figure 24–36). In present-day family members, a separate exon usually encodes the amino acids in each Ig-like domain, consistent with the likelihood that new family members arose during evolution by exon and gene duplications.

Summary

There are three main functionally distinct classes of T cells. Cytotoxic T cells (T_C cells) directly kill infected cells by secreting perforins and granzymes that induce the infected cells to undergo apoptosis. Helper T cells (T_H cells) help activate cytotoxic T cells to kill their target cells, B cells to make antibody responses, macrophages to destroy the microorganisms they harbor, and dendritic cells to activate T cells. Regulatory T cells (T_{reg} cells) produce suppressive proteins (such as the cytokines IL10 and TGF β) to inhibit other immune cells.

All T cells express cell-surface antigen receptors (TCRs), which are encoded by genes that are assembled from multiple gene segments during T cell development in the thymus. TCRs recognize peptide fragments of foreign proteins that are displayed in association with MHC proteins on the surface of antigen-presenting cells (APCs) and target cells. Naïve T cells are activated in peripheral lymphoid organs by activated dendritic cells, which secrete cytokines and express peptide-MHC complexes, co-stimulatory proteins, and various cell-cell adhesion molecules on their cell surface.

Class I MHC proteins present foreign peptides to T_C cells, whereas class II MHC proteins present foreign peptides to T_H cells and T_{reg} cells. Whereas class I MHC proteins are expressed on almost all nucleated vertebrate cells, class II MHC proteins are normally restricted to APCs, including dendritic cells, macrophages, and B lymphocytes. Both classes of MHC proteins have a single peptide-binding groove, which binds a large set of small peptide fragments produced intracellularly by normal protein-degradation processes: class I MHC proteins mainly bind fragments produced in the cytosol, whereas class II MHC proteins mainly bind fragments produced in endocytic compartments. The peptide-MHC complexes are transported to the cell surface, where 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. T cells also express CD4 or CD8 co-receptors, which recognize invariant regions of MHC proteins: T_H cells and T_{reg} cells express CD4, which recognizes class II MHC proteins; T_C cells express CD8, which recognizes class I MHC proteins.

A combination of positive and negative selection operates during T cell development in the thymus to help ensure that only T cells with potentially useful TCRs survive, mature, and emigrate, while all of the others die by apoptosis. The naïve T_H and T_C cells that leave the thymus constantly receive survival signals when their TCRs recognize self-peptide-MHC complexes, but they can only be activated when their TCRs encounter foreign peptides in the grooves of MHC proteins on an activated dendritic cell. The natural T_{reg} cells that leave the thymus suppress self-reactive lymphocytes to help maintain self-tolerance.

The production of an effector T cell from a naïve 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 a T_H or T_{reg} cell. Co-stimulatory proteins on the dendritic cell surface and secreted cytokines are the other signals. When naïve T_H cells are initially activated on a dendritic cell, they differentiate into T_{H1} , T_{H2} , T_{FH} or T_{H17} effector helper cells or into induced T_{reg} cells, depending mainly on the cytokines in their environment. T_{H1} cells secrete interferon- γ (IFN γ) to activate macrophages and to induce B cells to switch the class of Ig they make; T_{H2} and T_{FH} cells secrete other cytokines that also induce B cells to switch Ig class; and T_{H17} cells secrete IL17 to promote inflammatory responses

WHAT WE DON'T KNOW

- What initiates an autoimmune disease such as type 1 diabetes or multiple sclerosis?
- When a naïve or memory T or B cell is activated by antigen and co-stimulatory signals, how does it decide whether to become an effector cell or memory cell? Are there cells that are pre-committed to becoming either effector or memory cells, for example, or is the decision determined solely by extracellular signals?
- Why do some of us make IgE antibodies against harmless antigens and thereby develop hay fever and allergic asthma, while most of us do not, and why is the proportion of such allergic individuals increasing?
- How does a cytotoxic T cell (or NK cell) avoid being killed by the perforin and granzymes that it secretes to kill a target cell?

and wound healing. The effector helper T_H 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 producing a combination of membrane-bound and secreted co-stimulatory proteins. T_{reg} cells suppress immune cells using cell-surface and secreted inhibitory proteins.

Both T cells and B cells require multiple signals for activation. Antigen binding to the TCRs or BCRs provides one signal, while co-stimulatory proteins binding to co-receptors and cytokines binding to their complementary receptors provide the others. Effector T_H cells provide the co-stimulatory signals for B cells, whereas APCs provide them for T cells.

PROBLEMS

Which statements are true? Explain why or why not.

24–1 T cells whose receptors strongly bind a self-peptide-MHC complex are killed off in peripheral lymphoid organs when they encounter the self peptide on an antigen-presenting dendritic cell.

24–2 To guarantee that the antigen-presenting cells in the thymus will display a complete repertoire of self peptides to allow elimination of self-reactive T cells, the thymus recruits dendritic cells from all over the body.

24–3 The antibody diversity created by the combinatorial joining of V, D, and J segments by V(D)J recombination pales in comparison to the enormous diversity created by the random gain and loss of nucleotides at V, D, and J joining sites.

Discuss the following problems.

24–4 Why do living trees not rot? Redwood trees, for example, can live for centuries, but once they die they decay fairly quickly. What might this suggest?

24–5 It would be disastrous if a complement attack were not confined to the surface of the pathogen that is the target of the attack. Yet, the proteolytic cascade involved in the attack liberates biologically active molecules at several steps: one that diffuses away and one that remains bound to the target surface. How does the complement reaction remain localized when active products leave the surface?

24–6 Based on its sequence similarity to Apobec1, which deaminates Cs to Us in RNA, activation-induced deaminase (AID) was originally proposed to work on RNA. But definitive experiments in *E. coli* demonstrated that AID deaminates Cs to Us in DNA. The authors of the paper expressed AID in bacteria and followed mutations in a selectable gene. They found that AID expression increased mutations about fivefold above the background level in the absence of AID expression. More importantly, they found that 80% of the induced mutations were G→A or C→T. Does this fit with your expectation if AID-induced mutations arose by deamination of C to U in the DNA?

[Hint: imagine what would happen if the G:U mismatch created by AID was replicated several times; how would the sequences of the final mutations relate to the original G-C base pair?]

24–7 For many years it was a complete mystery how cytotoxic T cells could see a viral protein that seemed to be present only in the nucleus of the virus-infected cell. The answer was revealed in a classic paper that took advantage of a clone of T cells whose T cell receptor was directed against an antigen associated with the nuclear protein of the 1968 strain of influenza virus. The authors of the paper found that when they incubated high concentrations of certain peptides derived from the viral nuclear protein, the cells became sensitive to lysis by subsequent incubation with the cytotoxic T cells. Using various peptides from the 1968 strain and the 1934 strain (with which the cytotoxic T cells did not react), the authors defined the particular peptide responsible for the T cell response (Figure Q24–1).

A. Which part of the viral protein gives rise to the peptide that is recognized by the clone of cytotoxic T cells?

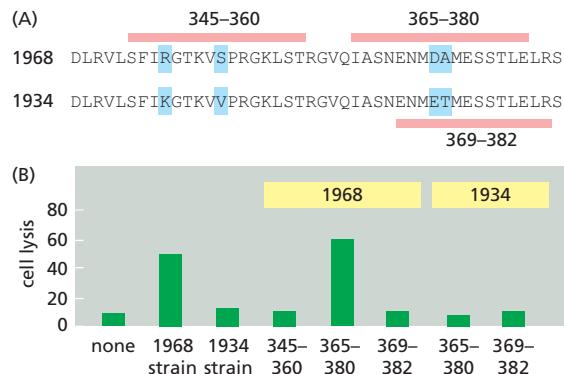


Figure Q24–1 Viral nuclear protein recognition by cytotoxic T cells (Problem 24–7). (A) Sequences of a segment of the nuclear protein from the 1968 and 1934 strains of influenza virus. Peptides used in the experiments in (B) are highlighted by pink bars. The amino acid differences between the viral proteins are highlighted in blue. (B) Cytotoxic T-cell-mediated lysis of target cells. The target cells were untreated (none), infected with virus (1968 or 1934 strain), or preincubated with high concentrations of the indicated viral peptide.

Why do not all viral peptides sensitize the target cells for lysis by the cytotoxic T cells?

B. It is thought the MHC molecules come to the cell surface with peptides already bound. If that is so, how do you imagine that these experiments worked?

24–8 Working out the rules by which T cells interact with their target cells was complicated. Some of the key observations came from studying the way cytotoxic T cells killed cells infected with choriomeningitis virus (LCMV). Cytotoxic T cells derived from mice expressing “k-type” class I MHC proteins lysed LCMV-infected cells expressing the same k-type MHC protein, but they did not lyse infected cells from mice expressing “d-type” class I MHC proteins (Figure Q24–2). Similarly, cytotoxic T cells from d-type mice lysed infected d-type cells, but not infected k-type cells. LCMV can kill both k-type and d-type mice.

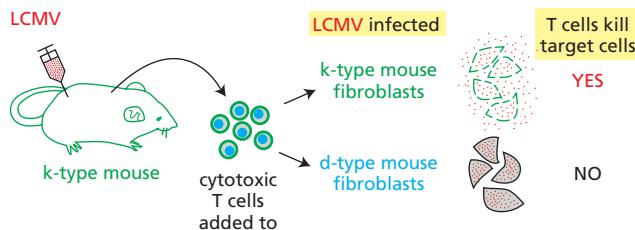


Figure Q24–2 Pattern of killing of LCMV-infected fibroblasts by cytotoxic T cells from an LCMV-infected k-type mouse (Problem 24–8).

A. If homozygous d-type mice were bred to homozygous k-type mice to generate d-type/k-type heterozygous progeny, would you expect that cytotoxic T cells from these heterozygotes, when infected with LCMV, to be able to lyse infected d-type cells? How about infected k-type cells? Explain your answers.

B. Oddly enough, LCMV infection does not kill mice that lack a thymus—such as “nude” mice, so called because they also lack hair. If a thymus is transplanted back into a nude mouse, it will die when infected with LCMV. Suppose that a d-type/k-type heterozygous nude mouse was given a thymus from an d-type donor. Would you expect its cytotoxic T cells to be able to lyse infected d-type cells? How about infected k-type cells? Explain your answers.

24–9 Before exposure to a foreign antigen, T cells with receptors specific for the antigen are a tiny fraction of the T cells—on the order of 1 in 10^5 or 1 in 10^6 T cells. After exposure to the antigen, only a small number of dendritic cells typically display the antigen on their surface. How long does it take for such antigen-presenting dendritic cells to interact with the antigen-specific T cells, which is the key first step in T cell activation and clonal expansion? The dynamics of the search process were examined by labeling dendritic cells red and T cells green, so that contacts in an intact lymph node could be scored visually using two-photon fluorescence microscopy (Figure Q24–3A). The frequency of contacts between dendritic cells and T cells from such experiments is given in Figure 24–3B. Assuming that 100 dendritic cells present the specific antigen, how long would it take them to scan 10^5 T cells? How long for 10^6 T cells?

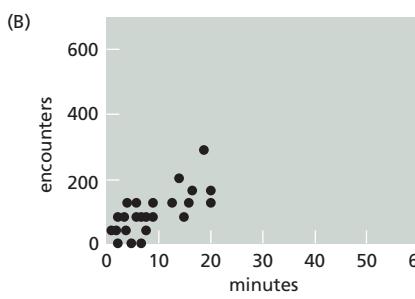
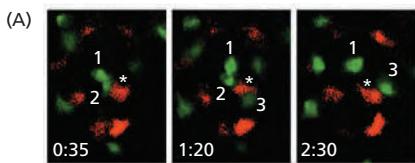


Figure Q24–3 Scanning of the T cell repertoire by dendritic cells (Problem 24–9). (A) Contacts between different T cells and one dendritic cell. T cells are green and dendritic cells are red. The dendritic cell labeled with an asterisk contacts a total of three T cells (numbered) over time in this sequence of images. Times are shown as hours: minutes. (B) Plot of T cell contacts for individual dendritic cells over time. (A, from P Bousso and E. Robey, *Nat. Immunol.* 4:579–581, 2003. With permission from Macmillan Publishers Ltd.)

24–10 At first glance, it would seem a dangerous strategy for the thymus to actively promote the survival, maturation, and emigration of developing T cells that bind weakly to self peptides bound to self MHC molecules. Would it not be safer to get rid of these T cells, along with those that bind strongly to such self-peptide-MHC complexes, as this would seem a more secure way to avoid autoimmune reactions?

24–11 CD4 proteins on helper and regulatory T cells serve as co-receptors that bind to invariant parts of class II MHC proteins. CD4 is thought to increase the adhesion between T cells and antigen-presenting cells (APCs) that are initially connected only weakly by the T cell receptor bound to its specific peptide-MHC complex. To test this possibility, you label cell-surface MHC molecules with a fluorescently labeled peptide so that you can detect individual peptide-MHC complexes at the interface between the APCs and the T cells in a culture dish. To detect T cell responses—the sign of a productive contact—you load them with a Ca^{2+} indicator dye, as cytosolic Ca^{2+} increases when lymphocytes are active. You now count the peptide-MHC complexes at a large number of interfaces (immunological synapses) and measure the resulting uptake of Ca^{2+} in the adherent T cells (Figure Q24–4, red circles). When you repeat the experiment in the presence of blocking antibodies against CD4, you get a different result (blue circles). Do these results support or refute the notion that CD4 augments T cell receptor binding? Explain your answer.

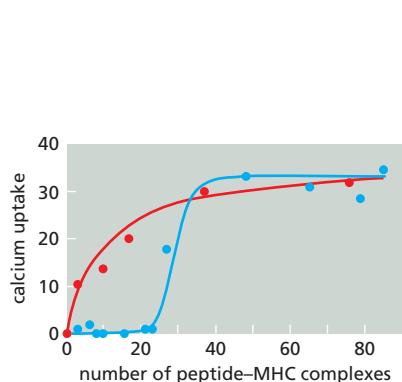


Figure Q24–4 Role of CD4 in the T cell response (Problem 24–11). The uptake of Ca^{2+} in cells with different numbers of fluorescently labeled peptide-MHC complexes at the interface between the T cells and the antigen-presenting cells. The results in the absence of CD4-blocking antibodies are shown by the red curve; results in the presence of CD4 antibodies are shown by the blue curve.