



Virtual University of Pakistan
FEDERAL GOVERNMENT UNIVERSITY

COURSE HANDOUTS

Cell Biology

BIO201

Order Your VU Books Now

Get all Virtual University books in high-quality hard copy, delivered right to your doorstep. Visit vubookshoppk.com



Order Now

Part One • THE CELL

1 *Life and Chemistry: Small Molecules*



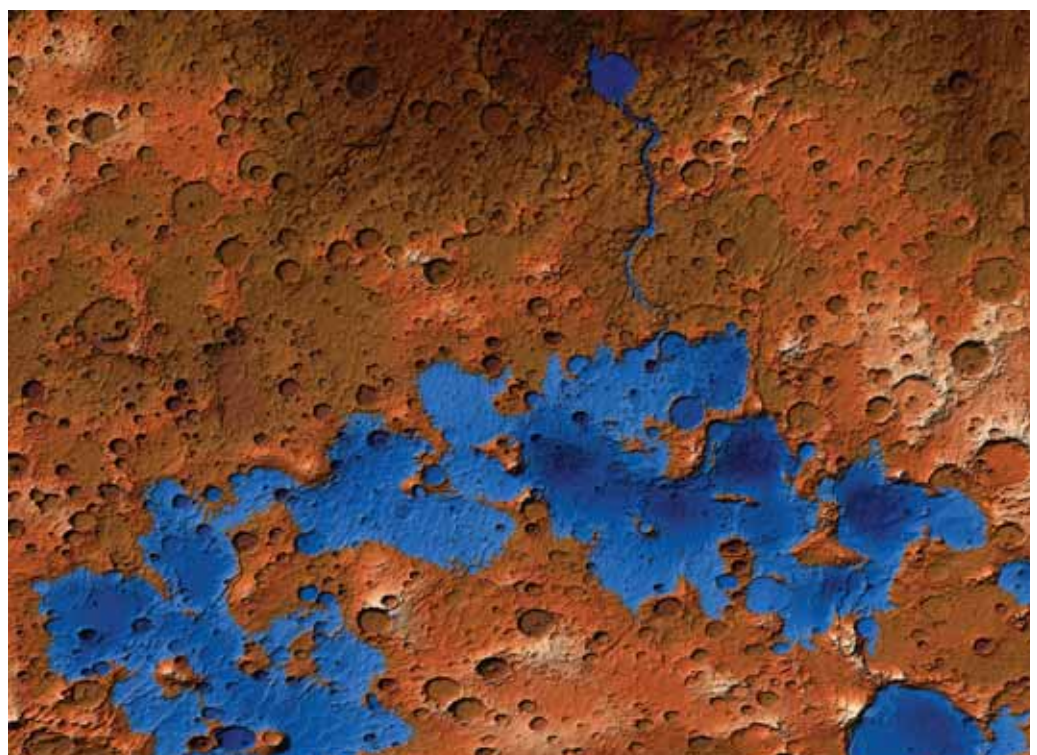
Mars today is a cold, dry place, not suitable for life as we know it. But 3 billion years ago, it was warmer and wetter. An orbiting probe from Earth recently photographed a huge dry lake bed, the size of New Mexico and Texas combined, on the Martian surface. Another probe found evidence of water trapped just below the icy surface of the Martian polar region. These discoveries by geologists have sparked the interest of biologists, for where there is water, there can be life. There is good reason to believe that life as we know it cannot exist without water.

Animals and plants that live on Earth's land masses had to evolve elaborate ways to retain the water that makes up about 70 percent of their bodies. Aquatic organisms living in water do not need these water-retention mechanisms; thus biologists have concluded that the first living things originated in a watery environment. This environment need not have been the lakes, rivers, and oceans with which we are familiar. Living organisms have been found in hot springs at temperatures above the usual boiling point of water, in a lake beneath the frozen Antarctic ice, in water trapped 2 miles below Earth's surface, in water 3 miles below the surface of the sea, in extremely acid and extremely salty water, and even in the water that cools the interiors of nuclear reactors.

With 20 trillion galaxies in the universe, each with 100 billion stars, there are many planets out there, and if our own solar system is typical, some of them have the water needed for life. As biologists contemplate how life could originate from nonliving matter, their attention focuses not just on the presence of water, but on what is dissolved in it.

A major discovery of biology is that living things are composed of the same types of chemical elements as the vast nonliving portion of the universe. This *mechanistic* view—that life is chemically based and obeys universal physicochemical laws—is a relatively recent one in human history. The concept of a “vital force” responsible for life, different from the forces found in physics and chemistry, was common in Western culture until the nineteenth century, and many people still assume such a force exists. However, most scientists adhere to a mechanistic view of life, and it is the cornerstone of medicine and agriculture.

A Grand Canyon on Mars This false color image from the Mars Global Surveyor shows in blue the dry remains of what was once a huge lake on Mars. Just as the Colorado River carved Earth's Grand Canyon, torrents of water from the lake probably carved the mile-deep canyon that is visible as a thin blue line just north of the lake bed.

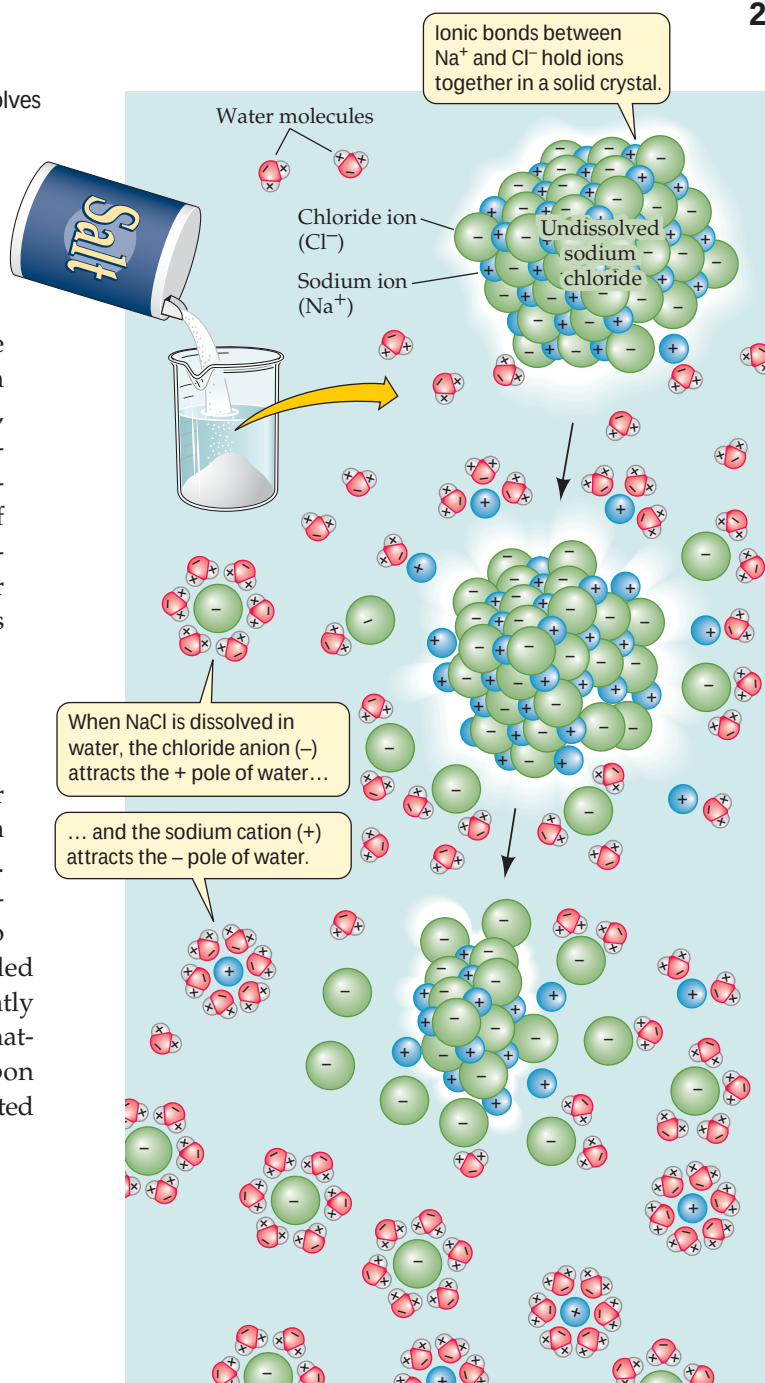
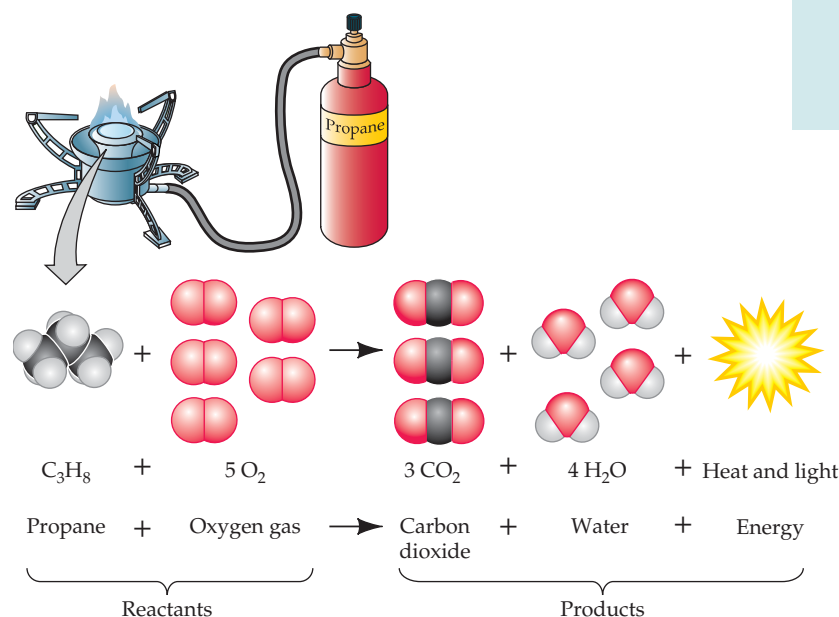


2.14 Water Molecules Surround Ions When an ionic solid dissolves in water, polar water molecules cluster around cations or anions, blocking their reassociation into a solid and forming a solution.

These weak interactions between nonpolar substances are enhanced by **van der Waals forces**, which result when two atoms of nonpolar molecules are in close proximity. These brief interactions are a result of random variations in the electron distribution in one molecule, which create an opposite charge distribution in the adjacent molecule. Although a single van der Waals interaction is brief and weak at any given site, the summation of many such interactions over the entire span of a large nonpolar molecule can produce substantial attraction. van der Waals forces are important in maintaining the structures of many biologically important substances.

Chemical Reactions: Atoms Change Partners

A **chemical reaction** occurs when atoms combine or change their bonding partners. Consider the combustion reaction that takes place in the flame of a propane stove. When propane (C_3H_8) reacts with oxygen gas (O_2), the carbon atoms become bonded to oxygen atoms instead of to hydrogen atoms, and the hydrogen atoms become bonded to oxygen instead of carbon (Figure 2.15). As the covalently bonded atoms change partners, the composition of the matter changes, and propane and oxygen gas become carbon dioxide and water. This chemical reaction can be represented by the equation



2.15 Bonding Partners and Energy May Change in a Chemical Reaction One molecule of propane reacts with five molecules of oxygen gas to give three molecules of carbon dioxide and four molecules of water. This reaction releases energy in the form of heat and light.

2

Life and Chemistry: Large Molecules



In 1984, a rock was found on the ice in the Allan Hills region of Antarctica. ALH 84001, as it came to be called, was a meteorite that came from Mars. We know this because the composition of the gases trapped within the rock was identical to the Martian atmosphere, which is quite different from Earth's atmosphere. Radioactive dating and mineral analyses determined that ALH 84001 was 4.5 billion years old and had been blasted off the Martian surface 16 million years ago, landing on Earth fairly recently, about 11,000 years ago.

Scientists found water trapped below the Martian meteorite's surface. This discovery was not surprising, considering that surface observations of Mars have indicated that liquid water may once have been abundant there (see Chapter 2). Because water is the *sine qua non* for life, scientists wondered whether the meteorite might contain other signs of life as well. Their analysis revealed two substances related to living systems. First, simple carbon-containing molecules called polycyclic aromatic hydrocarbons were present in small but unmistakable amounts; these substances are formed by decaying organisms, such as microbes. And second, crystals of magnetite, an iron oxide mineral made by many living things on Earth, were isolated from the interior of the rock.

ALH 84001 is not the only visitor from outer space that has been shown to contain the chemistry of life. Fragments of a meteorite that fell around the town of Murchison, Australia in 1969 were found to contain molecules that are unique to life, including purines and pyrimidines (the building blocks of DNA) and amino acids (which link together to form proteins). All of the amino acids showed a "handedness" that is unique to life.

These meteorites suggest that life is not found only on Earth, but they do not answer the question of how or where life arose from nonliving matter.

We begin this chapter by presenting two hypotheses for the origin of life on Earth. After discussing these hypotheses, we take a detailed look at the four kinds of large molecules that characterize living organisms: proteins, carbohydrates, lipids (fats), and nucleic acids.

Theories of the Origin of Life

Living things are composed of the same elements as the inanimate universe, the 92 elements of the periodic table (see Figure 2.3). But the arrangements of these atoms into molecules in biological

Was Life Once Here? The meteorite ALH 84001, which came from Mars and landed in Antarctica, contains the chemical signatures of life.



36 CHAPTER THREE

systems are unique. You cannot find DNA in rocks unless it came from a once-living organism.

How life began on Earth sometime during the 600 million years of the Hadean is impossible for us to know for certain, given the vast amount of time that has passed. There are two theories of the origin of life: life from extraterrestrial sources, and chemical evolution.

Could life have come from outside Earth?

As we described in Chapter 2, comets probably brought Earth most of its water. The meteorites described at the beginning of this chapter are evidence that molecules characteristic of life may have traveled to Earth from space. Taken together, these two observations suggest that some of life's complex molecules could have come from space. Although the presence of such molecules in rocks may suggest that those rocks once harbored life, it does not prove that there were living things in the rocks when they landed on Earth. Claims that the spherical objects seen in ALH 84001 are the remnants of ancient Martian organisms are far from accepted by all scientists in the field.

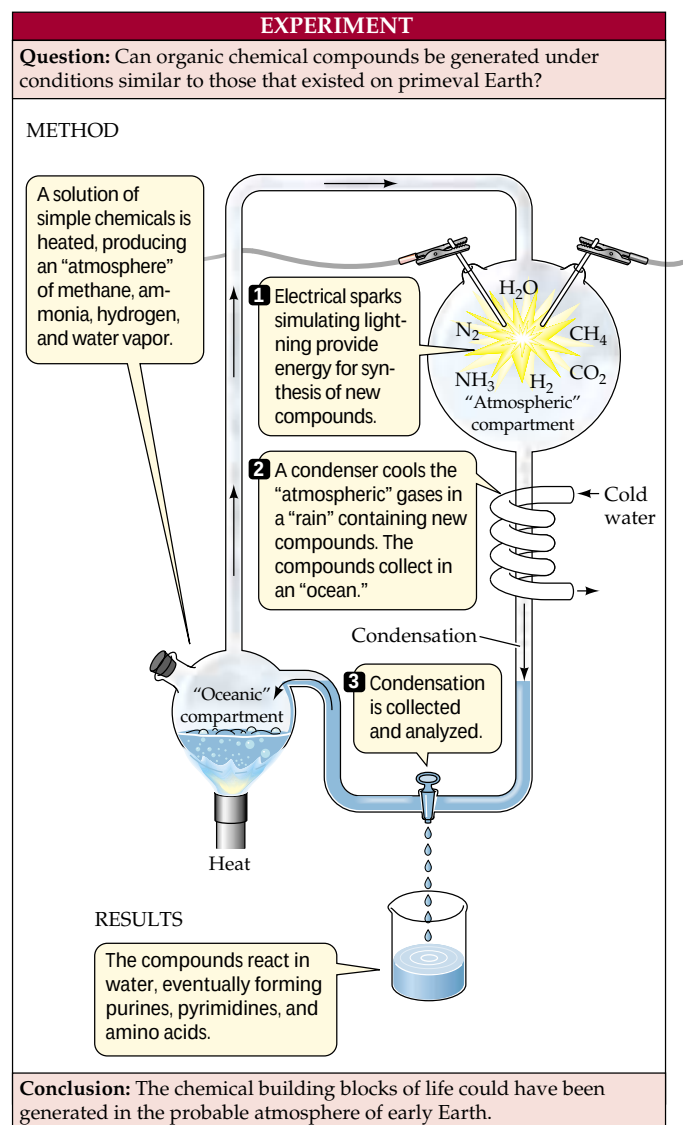
Most scientists find it hard to believe that an organism in a meteorite could survive thousands of years traveling through space, followed by intense heat as it passed through Earth's atmosphere. But there is some evidence that the heat inside some meteorites may not have been severe. When weakly magnetized rock is heated, it reorients its magnetic field to align with the magnetic field around it. In the case of ALH 84001, this would have been Earth's powerful magnetic field, which would have affected the meteorite as it approached our planet. Careful measurements indicate that, while reorientation did occur at the surface of the rock, it did not occur in the inside. The scientists who took these measurements, Benjamin Weiss and Joseph Kirschvink at the California Institute of Technology, claim that the inside of ALH 84001 was never heated over 40°C on its trip to Antarctica, making a long interplanetary trip by living organisms more plausible.

Did life originate on Earth?

Both Earth and Mars once had the water and other simple molecules that could, under the right conditions, form the large molecules unique to life. The second theory of the origin of life on Earth, **chemical evolution**, holds that conditions on the primitive Earth led to the emergence of these molecules. Scientists have sought to reconstruct those primitive conditions.

Early in the twentieth century, researchers proposed that there was little oxygen gas (O_2) in Earth's first atmosphere (unlike today, when it constitutes 21 percent of our atmos-

phere). O_2 is thought to have accumulated in quantity about 2.5 billion years ago as the by-product of the metabolism of single-celled life forms. In the 1950s, Stanley Miller and Harold Urey set up an experimental "primitive" atmosphere, containing hydrogen gas, ammonia, methane gas, and water vapor. Through these gases, they passed a spark to simulate lightning, then cooled the system so the gases would condense and collect in a watery solution, or "ocean" (Figure 3.1). Within days, the system contained numerous complex molecules, including amino acids, purines, and pyrimidines—some of the building blocks of life.



3.1 Synthesis of Prebiotic Molecules in an Experimental Atmosphere The Miller-Urey experiment simulated possible atmospheric conditions on primitive Earth to obtain some of the molecular building blocks of biological systems.

In science, an experiment and its results must be constantly reinterpreted and refined as more knowledge accumulates. The results of the Miller-Urey experiments have undergone several such refinements:

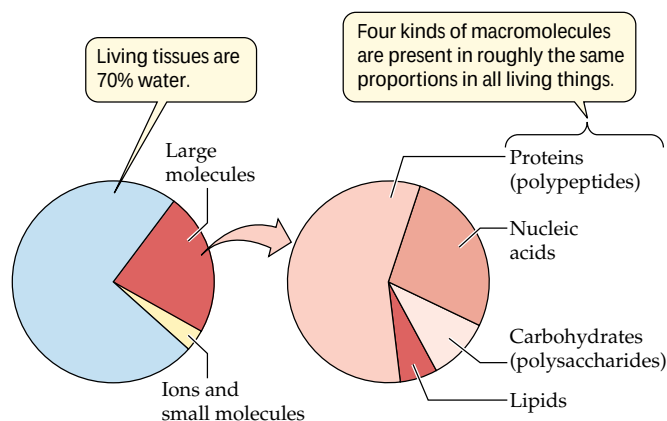
- In living organisms, many molecules have a unique three-dimensional “handedness” (see Figure 2.21). The amino acids, for example, are all in the L-configuration. But the amino acids formed in the Miller-Urey experiments were a mixture of the D- and L-forms. Recent experiments show that natural processes could have selected the L-amino acids from the mixture. Some minerals, especially calcite-based rocks, have unique crystal structures that selectively bind to D- or L-amino acids, separating the two. Such rocks were abundant during the Archean.
- Scientists’ views of the Earth’s original atmosphere have changed since Miller and Urey did their experiment. There is abundant evidence of major volcanic eruptions 4 billion years ago that released carbon dioxide (CO₂), nitrogen (N₂), hydrogen sulfide (H₂S), and sulfur dioxide (SO₂). Prebiotic chemistry experiments using these molecules in addition to the ones in the original “soup” have led to more diverse molecules.
- Long polymers had to be formed from simpler building blocks, called monomers. Scientists have used model systems to try to simulate conditions under which polymers could be made. Solid mineral surfaces, such as finely divided clays, seem to provide the best environment to bind monomers and allow them to polymerize.
- Miller and Urey, as well as others, suggested that life originated in hot pools at the edges of oceans. Because life has been found in many extreme environments on earth, scientists have proposed that such environments—found beneath ice, in deep-sea hydrothermal vents, and within fine clays near the shore—could be the original site of life’s emergence.

In whatever way the earliest stages of chemical evolution occurred, they resulted in the emergence of monomers and polymers that have probably remained unchanged in their general structure and function for 3.8 billion years. We now turn our attention to these large molecules.



Macromolecules: Giant Polymers

The four kinds of large molecules are made the same way and they are present in roughly the same proportions in all living organisms (Figure 3.2). A protein that has a certain role in an apple tree probably has a similar role in a human being, because their basic chemistry is the same. One important advantage of this *biochemical unity* is that organisms acquire needed biochemicals by eating other organisms. When



3.2 Substances Found in Living Tissues The substances shown here make up the nonmineral components of living tissue (bone would be an example of a “mineral tissue”). Most tissues are at least 70 percent water.

you eat an apple, the molecules you take in include carbohydrates, lipids, and proteins that can be refashioned into the special varieties of those molecules used by humans.

Macromolecules are giant **polymers** (*poly-*, “many”; *-mer*, “unit”) constructed by the covalent linking of smaller molecules called **monomers** (Table 3.1). These monomers may or may not be identical, but they always have similar chemical structures. Molecules with molecular weights exceeding 1,000 are usually considered macromolecules, and the proteins, polysaccharides (large carbohydrates), and nucleic acids of living systems certainly fall into this category.

Each type of macromolecule performs some combination of functions: energy storage, structural support, protection, catalysis, transport, defense, regulation, movement, and information storage. These roles are not necessarily exclusive. For example, both carbohydrates and proteins can play structural roles, supporting and protecting tissues and organisms. However, only nucleic acids specialize in information storage and function as hereditary material, carrying both species and individual traits from generation to generation.

3.1 The Building Blocks of Organisms

MONOMER	SIMPLE POLYMER	COMPLEX POLYMER (MACROMOLECULE)
Amino acid	Peptide or oligopeptide	Polypeptide (protein)
Nucleotide	Oligonucleotide	Nucleic acid
Monosaccharide (sugar)	Oligosaccharide	Polysaccharide (carbohydrate)

38 CHAPTER THREE

The functions of macromolecules are directly related to their shapes and to the sequences and chemical properties of their monomers. Some macromolecules fold into compact spherical forms with surface features that make them water-soluble and capable of intimate interaction with other molecules. Other proteins and carbohydrates form long, fibrous systems that provide strength and rigidity to cells and organisms. Still other long, thin assemblies of proteins can contract and cause movement.

Because macromolecules are so large, they contain many different functional groups (see Figure 2.20). For example, a large protein may contain hydrophobic, polar, and charged functional groups that give specific properties to local sites on the macromolecule. As we will see, this diversity of properties determines the shapes of macromolecules and their interactions with both other macromolecules and smaller molecules.

Condensation and Hydrolysis Reactions

Polymers are constructed from monomers by a series of reactions called **condensation reactions** or dehydration reactions (both terms refer to the loss of water). Condensation reactions result in covalently bonded monomers (Figure 3.3a) and release a molecule of water for each bond formed. The condensation reactions that produce the different kinds of polymers differ in detail, but in all cases, polymers form only if energy is added to the system. In living systems, specific energy-rich molecules supply this energy.

The reverse of a condensation reaction is a **hydrolysis reaction** (*hydro-*, “water”; *-lysis*, “break”). Hydrolysis reactions digest polymers and produce monomers. Water reacts with the bonds that link the polymer together, and the products are free monomers. The elements (H and O) of H_2O become part of the products (Figure 3.3b).

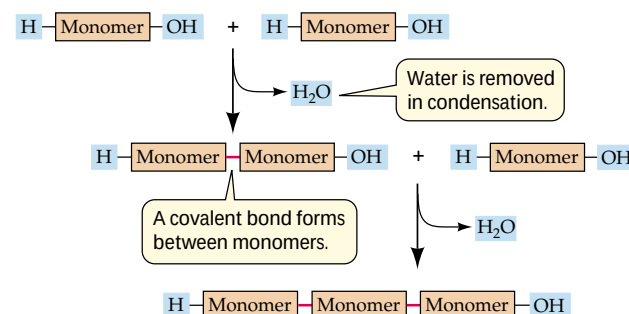
These two types of reactions are universal in living things, and as we have seen, were an important step in the origin of life in an aqueous environment. We begin our study of biological macromolecules with a very diverse group of polymers, the proteins.

Proteins: Polymers of Amino Acids

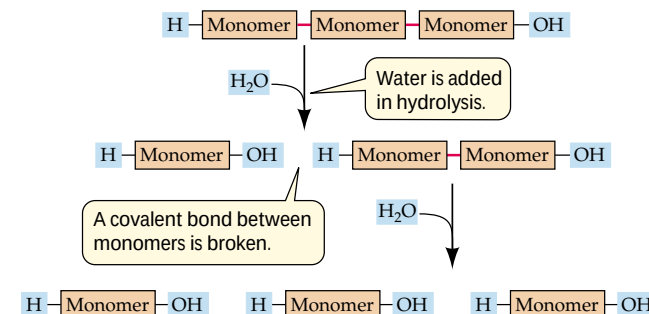
The functions of **proteins** include structural support, protection, transport, catalysis, defense, regulation, and movement. Among the functions of macromolecules listed earlier, only energy storage and information storage are not usually performed by proteins.

Proteins range in size from small ones such as the RNA-digesting enzyme ribonuclease A, which has a molecular weight of 5,733 and 51 amino acid residues, to huge molecules such as the cholesterol transport protein apolipoprotein B, which has a molecular weight of 513,000 and 4,636 amino

(a) Condensation



(b) Hydrolysis



3.3 Condensation and Hydrolysis of Polymers (a) A condensation reaction links monomers into polymers. (b) A hydrolysis reaction digests polymers into individual monomers.

acid residues. (The word *residue* refers to a monomer when it is part of a polymer.) Each of these proteins consists of a single unbranched polymer of amino acids (a *polypeptide chain*), which is folded into a specific three-dimensional shape.

Many proteins require more than one polypeptide chain to make up the functional unit. For example, the oxygen-carrying protein hemoglobin has four chains that are folded separately and associate together to make the functional protein. As we will see later in this book, numerous functional proteins can associate, forming “multi-protein machines” to carry out complex roles such as DNA synthesis.

The *composition* of a protein refers to the relative amounts of the different amino acids it contains. Not every protein contains all kinds of amino acids, nor an equal number of different ones. The diversity in amino acid content and sequence is the source of the diversity in protein structures and functions.

The next several chapters will describe the many functions of proteins. To understand them, we must first explore protein structure. First we will examine the properties of amino acids and see how they link together to form proteins. Then we will systematically examine protein structure and look at how a linear chain of amino acids is consistently folded into a compact three-dimensional shape. Finally, we will see how this three-dimensional structure provides a specific physical and chemical environment that influences how other molecules can interact with the protein.

Proteins are composed of amino acids

In Chapter 2, we looked at the structure of amino acids and identified four different groups attached to a central (α) carbon atom: a hydrogen atom, an amino group (NH_3^+), a carboxyl group (COO^-), and a unique **side chain**, or **R group**.

The R groups of amino acids are important in determining the three-dimensional structure and function of the protein macromolecule. They are highlighted in white in Table 3.2.

As Table 3.2 shows, amino acids are grouped and distinguished by their side chains. Some side chains are electrically

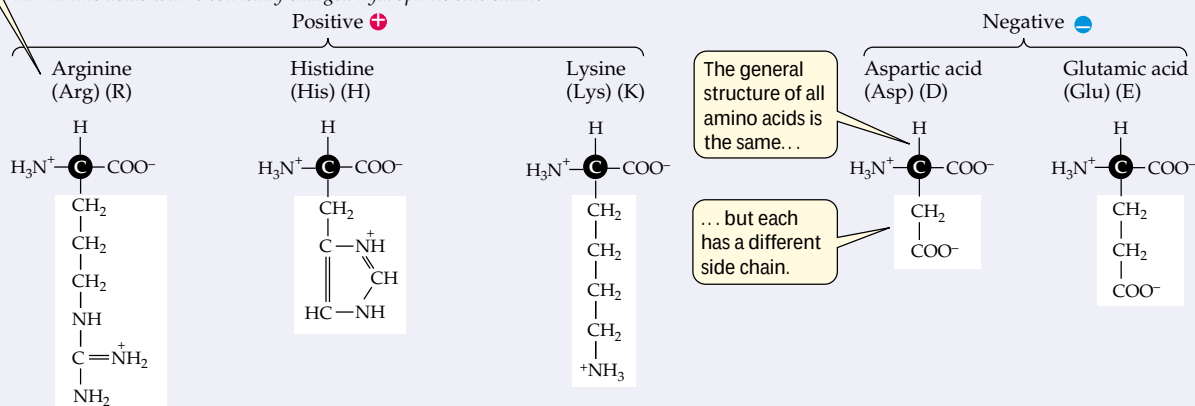
charged (+1, -1), while others are polar (δ^+ , δ^-), and still others are nonpolar and hydrophobic.

- ▶ The five amino acids that have electrically charged side chains attract water (are hydrophilic) and oppositely charged ions of all sorts.
- ▶ The five amino acids that have polar side chains tend to form weak hydrogen bonds with water and with other polar or charged substances. These amino acids are hydrophilic.

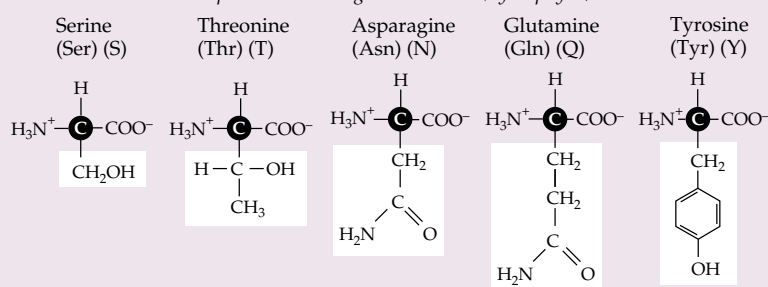
3.2 The Twenty Amino Acids Found in Proteins

Amino acids have both three-letter and single-letter abbreviations.

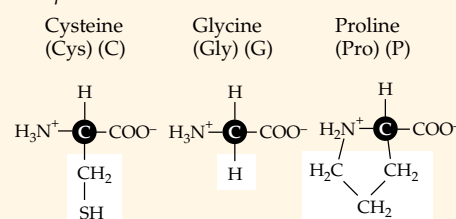
A. Amino acids with electrically charged hydrophilic side chains



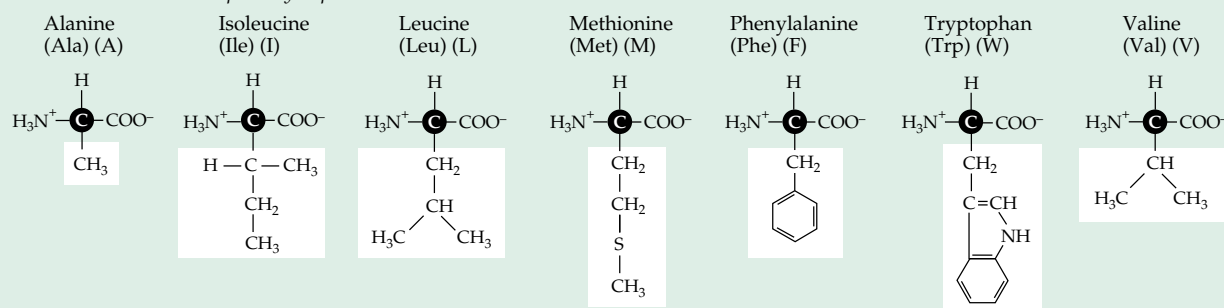
B. Amino acids with polar but uncharged side chains (hydrophilic)



C. Special cases



D. Amino acids with nonpolar hydrophobic side chains



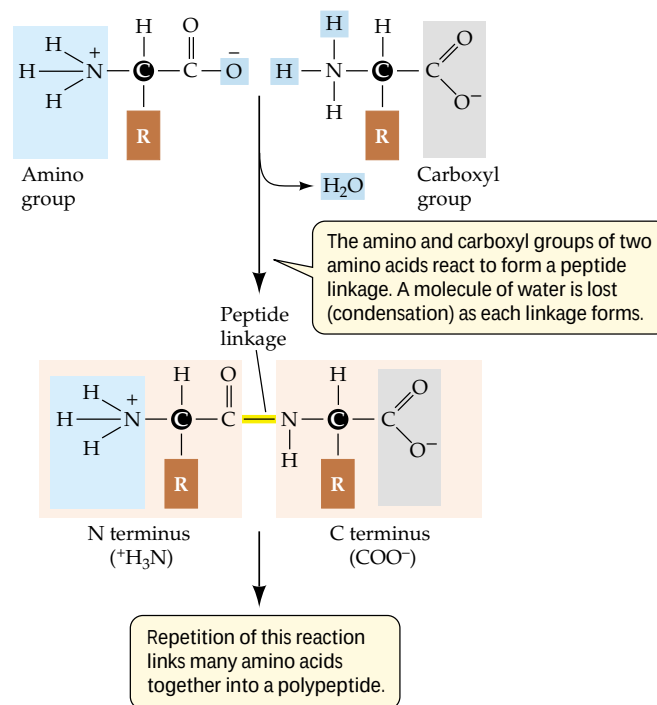
40 CHAPTER THREE

- Seven amino acids have side chains that are nonpolar hydrocarbons or very slightly modified hydrocarbons. In the watery environment of the cell, these hydrophobic side chains may cluster together in the interior of the protein. These amino acids are hydrophobic.
- Three amino acids—cysteine, glycine, and proline—are special cases, although their R groups are generally hydrophobic.

The *cysteine* side chain, which has a terminal —SH group, can react with another cysteine side chain to form a covalent bond called a **disulfide bridge** (—S—S—) (Figure 3.4). Disulfide bridges help determine how a polypeptide chain folds. When cysteine is not part of a disulfide bridge, its side chain is hydrophobic.

The *glycine* side chain consists of a single hydrogen atom and is small enough to fit into tight corners in the interior of a protein molecule, where a larger side chain could not fit.

Proline differs from other amino acids because it possesses a modified amino group lacking a hydrogen on its nitrogen, which limits its hydrogen-bonding ability. Also, the ring sys-



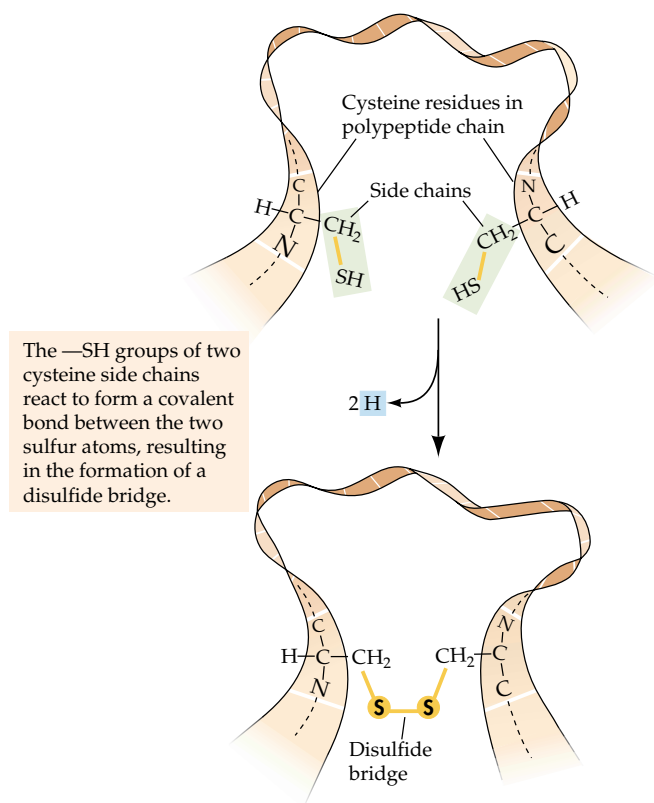
3.5 Formation of Peptide Linkages In living things, the reaction leading to a peptide linkage has many intermediate steps, but the reactants and products are the same as those shown in this simplified diagram.

tem of proline limits rotation about its α carbon, so proline is often found at bends or loops in a protein.

Peptide linkages covalently bond amino acids together

When amino acids polymerize, the carboxyl and amino groups attached to the α carbon are the reactive groups. The carboxyl group of one amino acid reacts with the amino group of another, undergoing a condensation reaction that forms a *peptide linkage*. Figure 3.5 gives a simplified description of this reaction. (In living systems, other molecules must activate the amino acids in order for this reaction to proceed, and there are intermediate steps in the process. We will examine these steps in Chapter 12.)

Just as a sentence begins with a capital letter and ends with a period, polypeptide chains have a linear order. The chemical “capital letter” marking the beginning of a polypeptide is the amino group of the first amino acid in the chain and is known as the *N terminus*. The “punctuation mark” for the end of the chain is the carboxyl group of the last amino acid—the *C terminus*. All the other amino and carboxyl groups in the chain (except those in side chains) are involved in peptide bond formation, so they do not exist in the chain



3.4 A Disulfide Bridge Disulfide bridges (—S—S—) are important in maintaining the proper three-dimensional shapes of some protein molecules.

as “free,” intact groups. Biochemists refer to the “N → C,” or “amino-to-carboxyl” orientation of polypeptides.

The peptide linkage has two characteristics that are important in the three-dimensional structure of proteins:

- ▶ Unlike many single covalent bonds, in which the groups on either side of the bond are free to rotate in space, the C—N peptide linkage is relatively inflexible. The adjacent atoms (the α carbons of the two adjacent amino acids) are not free to rotate because of the partial double-bond character of the peptide bond. This characteristic limits the folding of the polypeptide chain.
- ▶ The oxygen bound to the carbon (C=O) in the carboxyl group carries a slight negative charge (δ^-), whereas the hydrogen bound to the nitrogen (N—H) in the amino group is slightly positive (δ^+). This asymmetry of charge favors hydrogen bonding within the protein molecule itself and with other molecules, contributing to both the structure and the function of many proteins. Before we explore the significance of such hydrogen bonds, we need to examine the importance of the order of amino acids.

The primary structure of a protein is its amino acid sequence

There are four levels of protein structure, called primary, secondary, tertiary, and quaternary. The precise sequence of amino acids in a polypeptide constitutes the **primary structure** of a protein (Figure 3.6a). The peptide backbone of this primary structure consists of a repeating sequence of three atoms (—N—C—C—): the N from the amino group, the α carbon, and the C from the carboxyl group of each amino acid.

Scientists have deduced the primary structure of many proteins. The single-letter abbreviations for amino acids (see Table 3.2) are used to record the amino acid sequence of a protein. Here, for example, are the first 20 amino acids (out of a total of 124) in the protein ribonuclease from a cow:

KETAAAKFERQHMDSTSAA

The theoretical number of different proteins is enormous. Since there are 20 different amino acids, there could be $20 \times 20 = 400$ distinct dipeptides (two linked amino acids), and $20 \times 20 \times 20 = 8,000$ different tripeptides (three linked amino acids). Imagine this process of multiplying by 20 extended to a protein made up of 100 amino acids (which is considered a small protein). There could be 20^{100} such small proteins, each with its own distinctive primary structure. How large is the number 20^{100} ? There aren't that many electrons in the entire universe!

At the higher levels of protein structure, local coiling and folding give the molecule its final functional shape, but all of these levels derive from the primary structure—that is, the precise location of specific amino acids in the polypeptide

chain. The properties associated with a precise sequence of amino acids determine how the protein can twist and fold, thus adopting a specific stable structure that distinguishes it from every other protein.

Primary structure is determined by covalent bonds. But the next level of protein structure is determined by weaker hydrogen bonds.

The secondary structure of a protein requires hydrogen bonding

A protein's **secondary structure** consists of regular, repeated patterns in different regions of a polypeptide chain. There are two basic types of secondary structure, both of them determined by hydrogen bonding between the amino acid residues that make up the primary structure.

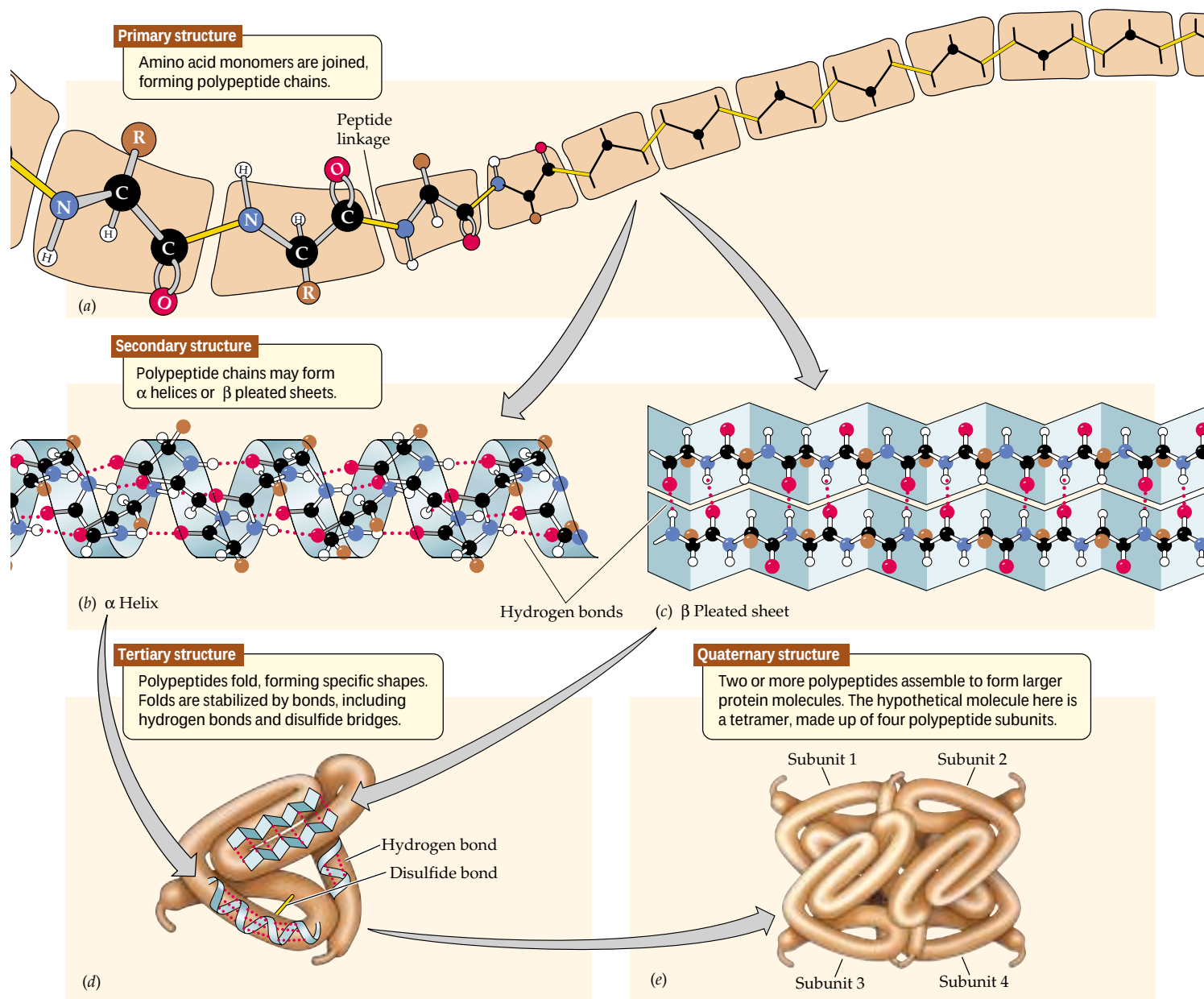
THE α HELIX. The **α (alpha) helix** is a right-handed coil that is “threaded” in the same direction as a standard wood screw (Figure 3.6b). The R groups extend outward from the peptide backbone of the helix. The coiling results from hydrogen bonds that form between the δ^+ hydrogen of the N—H of one amino acid residue and the δ^- oxygen of the C=O of another. When this pattern of hydrogen bonding is established repeatedly over a segment of the protein, it stabilizes the coil, resulting in an α helix. The presence of amino acids with large R groups that distort the coil or otherwise prevent the formation of the necessary hydrogen bonds will keep an α helix from forming.

The α -helical secondary structure is common in the fibrous structural proteins called keratins, which make up hair, hooves, and feathers. Hair can be stretched because stretching requires that only the hydrogen bonds of the α helix, not the covalent bonds, be broken; when the tension on the hair is released, both the hydrogen bonds and the helix re-form.

THE β PLEATED SHEET. A **β (beta) pleated sheet** is formed from two or more polypeptide chains that are almost completely extended and lying next to one another. The sheet is stabilized by hydrogen bonds between the N—H groups on one chain and the C=O groups on the other (Figure 3.6c). A β pleated sheet may form between separate polypeptide chains, as in spider silk, or between different regions of the same polypeptide chain that is bent back on itself. Many proteins contain regions of both α helix and β pleated sheet in the same polypeptide chain.

The tertiary structure of a protein is formed by bending and folding

In many proteins, the polypeptide chain is bent at specific sites and then folded back and forth, resulting in the **tertiary**



3.6 The Four Levels of Protein Structure

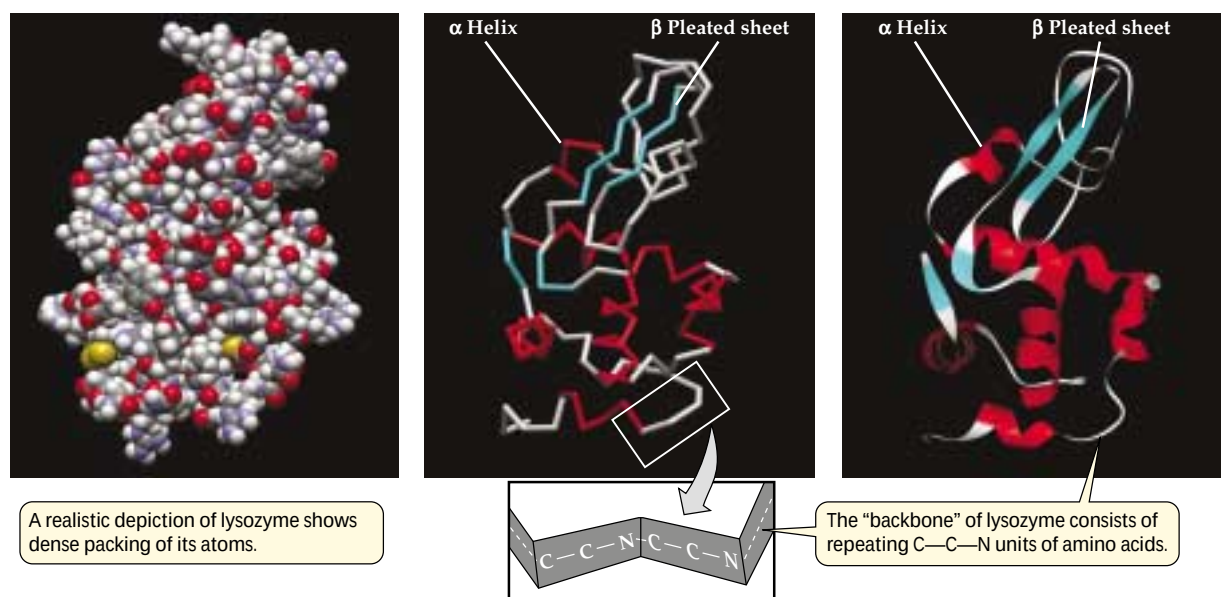
Secondary, tertiary, and quaternary structure all arise from the primary structure of the protein.

structure of the protein (Figure 3.6d). Although the α helices and β pleated sheets contribute to the tertiary structure, only parts of the macromolecule usually have these secondary structures, and large regions consist of structures unique to a particular protein.

While hydrogen bonding between the N—H and C=O groups within and between chains is responsible for secondary structure, the interactions between R groups—the amino acid side chains—determine tertiary structure. We described the various strong and weak interactions between atoms in Chapter 2 (see Table 2.1). Many of these interactions are involved in determining tertiary structure.

- Covalent *disulfide bridges* can form between specific cysteine residues (see Figure 3.4), holding a folded polypeptide in place.
- *Hydrophobic* side chains can aggregate together in the interior of the protein, away from water, folding the polypeptide in the process.
- *van der Waals forces* can stabilize the close interactions between the hydrophobic residues.
- *Ionic bonds* can form between positively and negatively charged side chains buried deep within a protein, away from water, forming a *salt bridge*.

A complete description of a protein's tertiary structure specifies the location of every atom in the molecule in three-dimen-



3.7 Three Representations of Lysozyme Different molecular representations of a protein emphasize different aspects of its tertiary structure. These three representations of lysozyme are similarly oriented.

sional space in relation to all the other atoms. Such a description is available for the protein lysozyme (Figure 3.7). The first tertiary structures to be determined took years to figure out, but today, dozens of new structures are published every week. The major advances making this possible have been the ability to produce large quantities of specific proteins by biotechnology and the use of computers to analyze the atomic data.

Bear in mind that both tertiary structure and secondary structure derive from a protein's primary structure. If lysozyme is heated slowly, the heat energy will disrupt only the weak interactions and cause only the tertiary structure to break down. But the protein will return to its normal tertiary structure when it cools, demonstrating that all the information needed to specify the unique shape of a protein is contained in its primary structure.

The quaternary structure of a protein consists of subunits

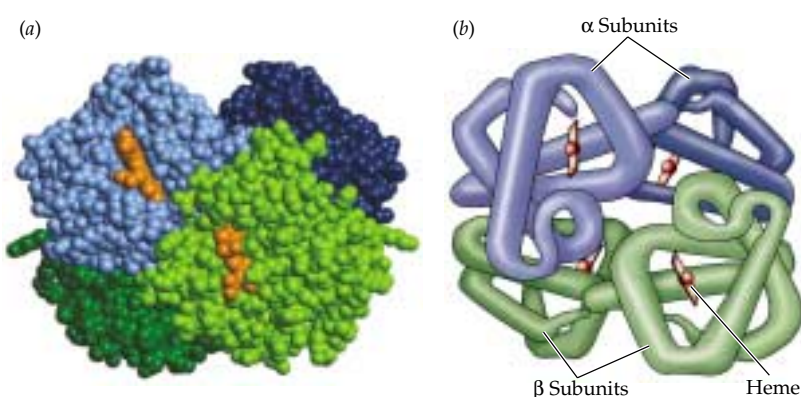
As mentioned earlier, many functional proteins contain two or more polypeptide chains, called *subunits*, each of them folded into its own unique tertiary structure. The protein's **quaternary structure** results from the ways in which these subunits bind together and interact (see Figure 3.6e).

Quaternary structure is illustrated by hemoglobin (Figure 3.8). Hydrophobic interactions,

van der Waals forces, hydrogen bonds, and ionic bonds all help hold the four subunits together to form the hemoglobin molecule. The function of hemoglobin is to carry oxygen in red blood cells. As hemoglobin binds one O_2 molecule, the four subunits shift their relative positions slightly, changing the quaternary structure. Ionic bonds are broken, exposing buried side chains that enhance the binding of additional O_2 molecules. The structure changes again when hemoglobin releases its O_2 molecules to the cells of the body.

The surfaces of proteins have specific shapes

Small molecules in a solution are in constant motion. They vibrate, rotate, and move from place to place like corn in a



3.8 Quaternary Structure of a Protein Hemoglobin consists of four folded polypeptide subunits that assemble themselves into the quaternary structure shown here. In these two graphic representations, each type of subunit is a different color. The heme groups contain iron and are the oxygen-carrying sites.

44 CHAPTER THREE

popper. If two of them collide in the right circumstances, a chemical reaction can occur. The specific shapes of proteins allow them to bind *noncovalently* to other molecules, which in turn allows other important biological events to occur. Here are just a few examples:

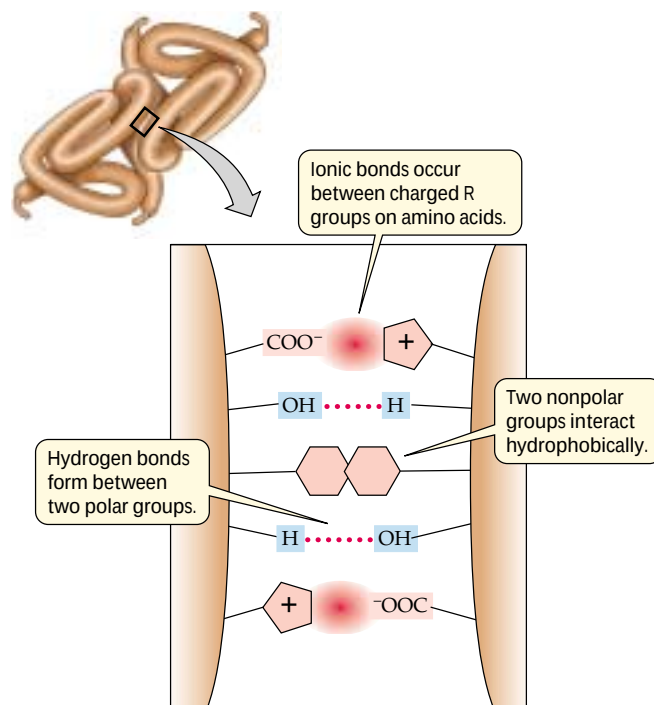
- ▶ Two adjacent cells can stick together because proteins protruding from each of the cells interact with each other (see Chapter 5).
- ▶ A substance can enter a cell by binding to a carrier protein in the cell surface membrane (see Chapter 5).
- ▶ A chemical reaction can be speeded up when an enzyme protein binds to one of the reactants (see Chapter 6).
- ▶ A “multi-protein machine,” DNA polymerase, can bind to and copy DNA (see Chapter 11).
- ▶ Another “multi-protein machine,” RNA polymerase, can synthesize RNA (see Chapter 12).
- ▶ Chemical signals such as hormones can bind to proteins on a cell’s outer surface (see Chapter 15).
- ▶ Defensive proteins called antibodies can recognize the shape of a virus coat and bind to it (see Chapter 18).

The biological specificity of protein function depends on two general properties of the protein: its shape and the chemistry of its exposed surface groups.

- ▶ **Shape.** When a molecule collides with and binds to a much larger protein, it is like a baseball being caught by a catcher’s mitt: The mitt has a shape that binds to the ball and fits around it. A hockey puck or a ping-pong ball would not fit a baseball catcher’s mitt. The binding of a molecule to a protein involves a general “fit” between two three-dimensional objects that becomes even more specific after initial binding.
- ▶ **Chemistry.** The surface of a protein has certain chemical groups that it presents to a substance attempting to bind to it (Figure 3.9). These groups are the R groups of the exposed amino acids, and are therefore a property of the protein’s primary structure.

Look again at the structures of the 20 amino acids in Table 3.2, noting the properties of the R groups. Exposed hydrophobic groups can bind to similarly nonpolar groups in the substance with which the protein interacts (often called the **ligand**). Charged R groups can bind to oppositely charged groups on the ligand. Polar R groups containing a hydroxyl (—OH) group can form a hydrogen bond with the ligand. These three types of interactions—hydrophobic, ionic, and hydrogen bonding—are weak by themselves, but strong when all of them act together. So the exposure of appropriate amino acid R groups on the protein surface allows the binding of a specific ligand to occur.

Knowing the exact shape of a protein and what can bind to it is important not only in understanding basic biology,



3.9 Noncovalent Interactions between Proteins and Other Molecules

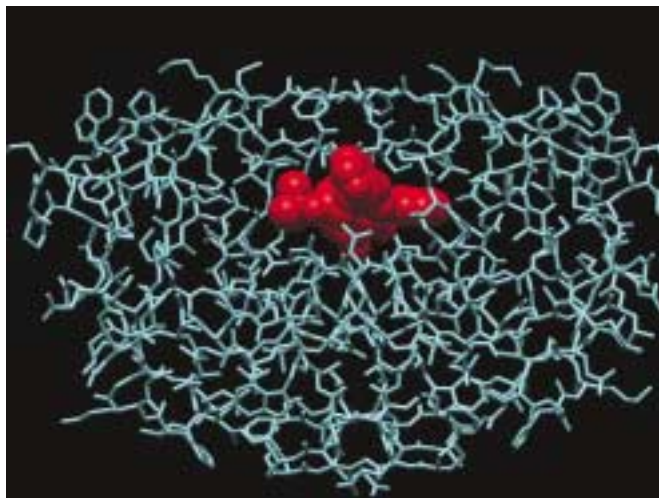
Noncovalent interactions allow a protein to bind tightly to another molecule with specific properties, or allow regions within a protein to interact with one another.

but also in applied fields such as medicine. For example, the three-dimensional structure of a protease, a protein essential for the replication of HIV—the virus that causes AIDS—was first determined, then specific proteins were designed to bind to it and block its action. These protease inhibitors have prolonged the lives of countless people living with HIV (Figure 3.10).

Protein shapes are sensitive to the environment

Because it is determined by weak forces, protein shape is sensitive to environmental conditions that would not break covalent bonds, but do upset the weaker noncovalent interactions that determine secondary and tertiary structure.

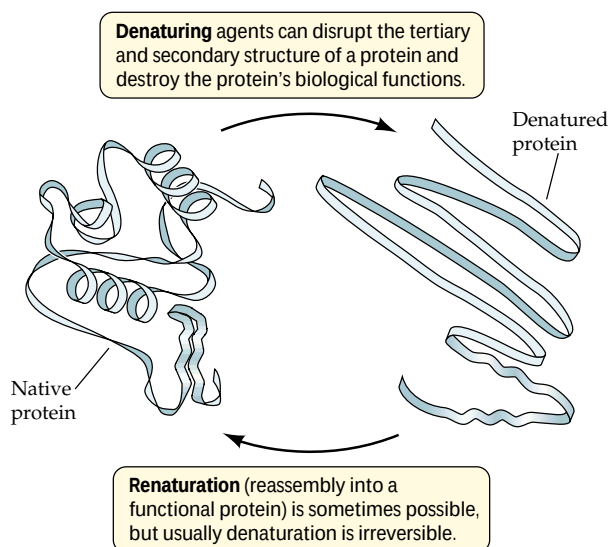
- ▶ Increases in *temperature* cause more rapid molecular movements and thus can break hydrogen bonds and hydrophobic interactions.
- ▶ Alterations in *pH* can change the pattern of ionization of carboxyl and amino groups in the R groups of amino acids, thus disrupting the pattern of ionic attractions and repulsions.
- ▶ High concentrations of polar substances such as urea can disrupt the hydrogen bonding that is crucial to protein



3.10 An HIV Protease Inhibitor After determining the structure of HIV protease (the blue polypeptide chain), a protein essential to the life cycle of HIV, biochemists designed a drug (the red space-filling model) to fit into the protease and block its function. Many people living with HIV and AIDS now take this drug.

structure. Nonpolar solvents may also disrupt normal protein structure.

The loss of a protein's normal three-dimensional structure is called **denaturation**, and it is always accompanied by a loss of the normal biological function of the protein (Figure 3.11).



3.11 Denaturation Is the Loss of Tertiary Protein Structure and Function Agents that can cause denaturation include high temperatures and certain chemicals.

Denaturation is often irreversible, because amino acids that were buried may now be exposed at the surface, and vice versa, causing a new structure to form or different molecules to bind to the protein. Boiling an egg denatures its proteins and is, as you know, not reversible. However, as we saw earlier in the case of lysozyme, denaturation may be reversible in the laboratory. If the protein is allowed to cool or the denaturing chemicals are removed, the protein may return to its "native" shape and normal function.

Chaperonins help shape proteins

There are two occasions when a polypeptide chain is in danger of binding the wrong ligand. First, following denaturation, hydrophobic R groups, previously on the inside of the protein away from water, become exposed on the surface. Since these groups can interact with similar groups on other molecules, the denatured proteins may aggregate and become insoluble, losing their function. Second, when a protein has just been made and has not yet folded completely, it can present a surface that binds the wrong molecule. In the cell, a protein can sometimes fold incorrectly as it is made. This can have serious consequences: In Alzheimer's disease, misfolded proteins accumulate in the brain and bind to one another, forming fibers in the areas of the brain that control memory, mood, and spatial awareness.

Living systems limit inappropriate protein interactions by making a class of proteins called, appropriately, **chaperonins** (recall the chaperones—usually teachers—at school dances who try to prevent "inappropriate interactions" among the students). Chaperonins were first identified in fruit flies as "heat shock" proteins, which prevented denaturing proteins from clumping together when the flies' temperatures were raised.

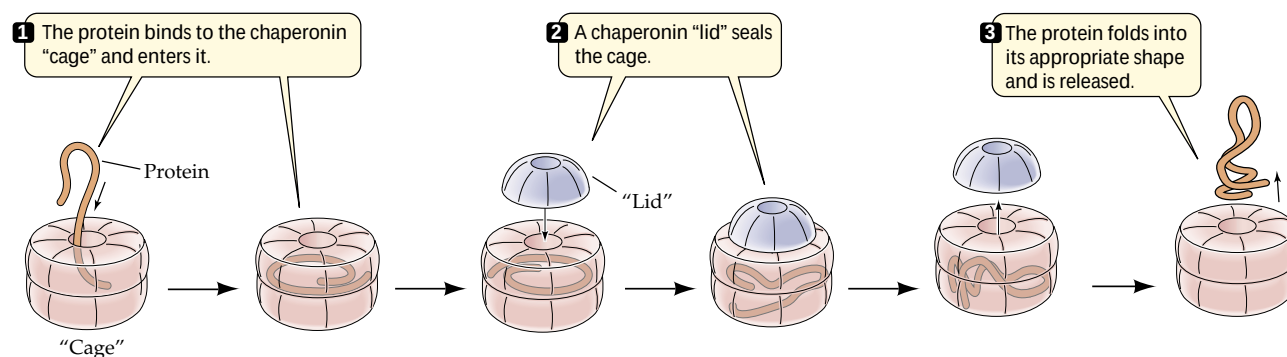
Some chaperonins work by trapping proteins in danger of inappropriate binding inside a molecular "cage" (Figure 3.12). This cage is composed of several identical subunits, and is itself a good example of quaternary protein structure. Inside the cage, the targeted protein folds into the right shape, and then is released at the appropriate time and place.

Carbohydrates: Sugars and Sugar Polymers

The second class of biological molecules, the **carbohydrates**, is a diverse group of compounds. Carbohydrates contain primarily carbon atoms flanked by hydrogen atoms and hydroxyl groups ($\text{H}-\text{C}-\text{OH}$). They have two major biochemical roles:

- ▶ They act as a source of energy that can be released in a form usable by body tissues.

46 CHAPTER THREE

**3.12 Chaperonins Protect Proteins from Inappropriate Binding**

Chaperonins surround new or denatured proteins and prevent them from binding to the wrong ligand.

- They serve as carbon skeletons that can be rearranged to form other molecules that are essential for biological structures and functions.

Some carbohydrates are relatively small, with molecular weights of less than 100. Others are true macromolecules, with molecular weights in the hundreds of thousands.

There are four categories of biologically important carbohydrates, which we will discuss in turn:

- *Monosaccharides* (*mono*-, "one"; *saccharide*-, "sugar"), such as glucose, ribose, and fructose, are simple sugars. They are the monomers out of which the larger carbohydrates are constructed.
- *Disaccharides* (*di*-, "two") consist of two monosaccharides linked together by covalent bonds.
- *Oligosaccharides* (*oligo*-, "several") are made up of several (3 to 20) monosaccharides.
- *Polysaccharides* (*poly*-, "many"), such as starch, glycogen, and cellulose, are large polymers composed of hundreds or thousands of monosaccharides.

The general formula for carbohydrates, CH_2O , gives the relative proportions of carbon, hydrogen, and oxygen in a monosaccharide (i.e., the proportions of these atoms are 1:2:1). In disaccharides, oligosaccharides, and polysaccharides, these proportions differ slightly from the general formula because two hydrogens and an oxygen are lost during each of the condensation reactions that form them.

Monosaccharides are simple sugars

Green plants produce monosaccharides through photosynthesis, and animals acquire them directly or indirectly from plants. All living cells contain the monosaccharide **glucose**. Cells use glucose as an energy source, breaking it down

through a series of reactions that release stored energy and produce water and carbon dioxide.

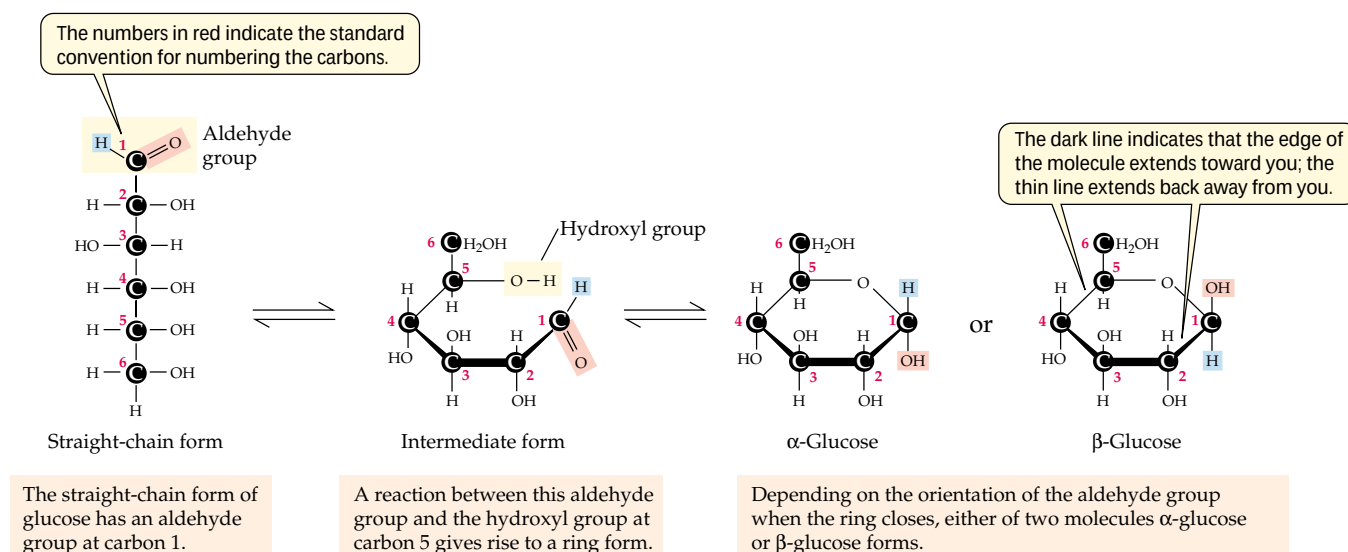
Glucose exists in two forms, the straight chain and the ring. The ring form predominates in more than 99 percent of circumstances because it is more stable under cellular conditions. There are two forms of the ring structure (α -glucose and β -glucose), which differ only in the placement of the —H and —OH attached to carbon 1 (Figure 3.13). The α and β forms interconvert and exist in equilibrium when dissolved in water.

Different monosaccharides contain different numbers of carbons. (The standard convention for numbering carbons in carbohydrates shown in Figure 3.13 is used throughout this book.) Most of the monosaccharides found in living systems belong to the D series of optical isomers (see Chapter 2). But some monosaccharides are structural isomers, which have the same kinds and numbers of atoms, but arranged differently. For example, the *hexoses* (*hex*-, "six"), a group of structural isomers, all have the formula $\text{C}_6\text{H}_{12}\text{O}_6$. Included among the hexoses are glucose, fructose (so named because it was first found in fruits), mannose, and galactose (Figure 3.14).

Pentoses (*pent*-, "five") are five-carbon sugars. Some pentoses are found primarily in the cell walls of plants. Two pentoses are of particular biological importance: Ribose and deoxyribose form part of the backbones of the nucleic acids RNA and DNA, respectively. These two pentoses are not isomers; rather, one oxygen atom is missing from carbon 2 in deoxyribose (*de*-, "absent") (see Figure 3.14). As we will see in Chapter 12, the absence of this oxygen atom has important consequences for the functional distinction of RNA and DNA.

Glycosidic linkages bond monosaccharides together

The disaccharides and polysaccharides described above are all constructed from monosaccharides that are covalently bonded together by condensation reactions that form *glycosidic linkages*. One such linkage between two monosaccharides forms a disaccharide. For example, a molecule of su-



3.13 Glucose: From One Form to the Other All glucose molecules have the formula $C_6H_{12}O_6$, but their structures vary. When dissolved in water, the α and β "ring" forms of glucose interconvert.

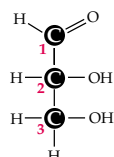
crose (table sugar) is formed from a glucose molecule and a fructose molecule, while lactose (milk sugar) contains glucose and galactose.

The disaccharide maltose contains two glucose molecules, but it is not the only disaccharide that can be made from two glucoses. When glucose molecules form a glycosidic linkage, the linkage will be one of two types, α or β , depending on whether the molecule that bonds its carbon 1 is α -glucose or β -glucose (see Figure 3.13). An α linkage with carbon 4 of a second glucose molecule gives maltose, whereas a β linkage gives cellobiose (Figure 3.15).

Maltose and cellobiose are disaccharide isomers, both having the formula $C_{12}H_{22}O_{11}$. However, they are different compounds with different properties. They undergo different chemical reactions and are recognized by different enzymes. For example, maltose can be hydrolyzed to its monosaccharides in the human body, whereas cellobiose cannot. Certain microorganisms have the chemistry needed to break down cellobiose.

Oligosaccharides contain several monosaccharides bound by glycosidic linkages at various sites. Many oligosaccharides have additional functional groups, which give them special properties. Oligosaccharides are often covalently bonded to proteins and lipids on the outer cell surface, where they serve as cell recognition signals. The human blood groups (such as ABO) get their specificity from oligosaccharide chains.

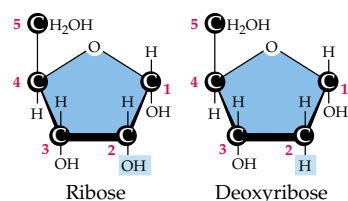
Three-carbon sugar



Glyceraldehyde

Glyceraldehyde is the smallest monosaccharide and exists only as the straight-chain form.

Five-carbon sugars (pentoses)

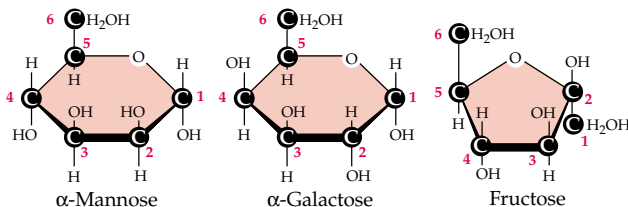


Ribose

Deoxyribose

Ribose and deoxyribose each have five carbons, but very different chemical properties and biological roles.

Six-carbon sugars (hexoses)



α -Mannose

α -Galactose

Fructose

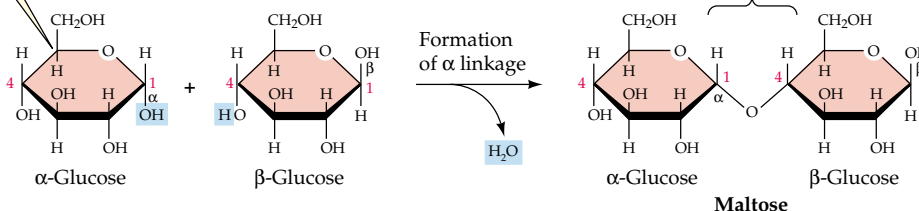
These hexoses are isomers. All have the formula $C_6H_{12}O_6$, but each has distinct chemical properties and biological roles.

3.14 Monosaccharides Are Simple Sugars Monosaccharides are made up of varying numbers of carbons. Some are structural isomers, which have the same number of carbons, but arranged differently. Fructose, for example, is a hexose, but forms a five-sided ring like the pentoses.

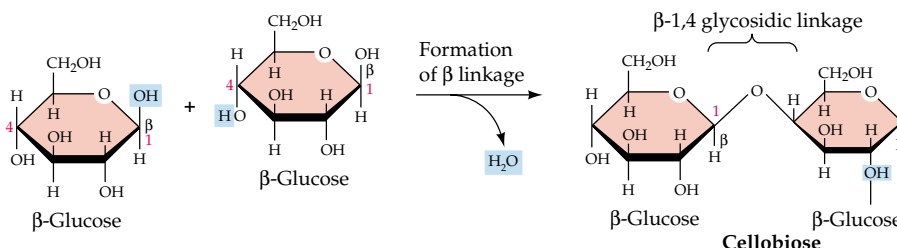
48 CHAPTER THREE

Maltose is produced when an α -1,4 glycosidic linkage forms between two glucose molecules. The hydroxyl group on carbon 1 of one glucose in the α (down) position reacts with the hydroxyl group on carbon 4 of the other glucose.

The presence of a carbon atom (C) at a junction such as this is implied.



In **cellobiose**, two glucoses are linked by a β -1,4 glycosidic linkage.



3.15 Disaccharides Are Formed by Glycosidic Linkages

Glycosidic linkages between two monosaccharides create many different disaccharides. Which disaccharide is formed depends on which monosaccharides are linked, and on the site (which carbon atom is linked) and form (α or β) of the linkage.

Polysaccharides serve as energy stores or structural materials

Polysaccharides are giant polymers of monosaccharides connected by glycosidic linkages (Figure 3.16).

- **Starch** is a polysaccharide of glucose with α -glycosidic linkages.
- **Glycogen** is a highly branched polysaccharide of glucose.
- **Cellulose** is also a polysaccharide of glucose, but its individual monosaccharides are connected by β -glycosidic linkages.

Starch actually comprises a large family of giant molecules of broadly similar structure. While all starches are large polymers of glucose with α linkages (Figure 3.16a), the different starches can be distinguished by the amount of branching that occurs at carbons 1 and 6 (Figure 3.16b). Some plant starches are unbranched, as in plant amylose; others are moderately branched, as in plant amylopectin. Starch readily binds water, and when that water is removed, unbranched starch tends to form hydrogen bonds between the polysaccharide chains, which then aggregate. This is what causes bread to become hard and stale. Adding water and gentle heat separates the chains and the bread becomes softer.

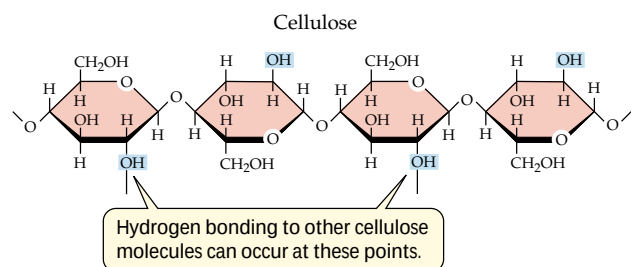
The polysaccharide glycogen stores glucose in animal livers and muscles. Starch and glycogen serve as energy storage compounds for plants and animals, respectively. These polysaccharides are readily hydrolyzed to glucose monomers, which in turn can be further degraded to liberate their stored energy and convert it to forms that can be used for cellular activities. If it is glucose that is actually needed for fuel, why must it be stored as a polymer? The reason is that 1,000 glucose molecules would exert 1,000 times the osmotic pressure (causing water to enter the cells; see Chapter 5) of a single glycogen molecule. If it were not for polysaccharides, many organisms would expend a lot of time and energy expelling excess water.

Cellulose is the predominant component of plant cell walls, and is by far the most abundant **organic** (carbon-containing) compound on Earth. Starch can be easily degraded by the actions of chemicals or enzymes. Cellulose, however, is chemically more stable because of its β -glycosidic linkages (Figure 3.16a). Thus starch is a good storage medium that can be easily broken down to supply glucose for energy-producing reactions, while cellulose is an excellent structural material that can withstand harsh environmental conditions without changing.

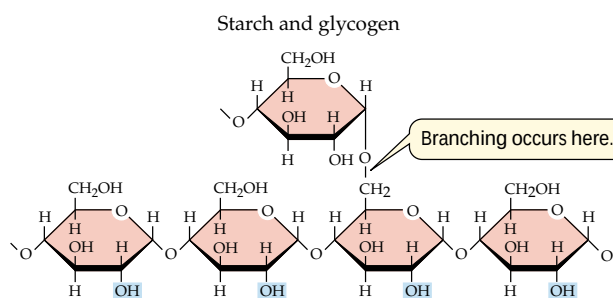
Chemically modified carbohydrates contain other groups

Some carbohydrates are chemically modified by the addition of functional groups, such as phosphate and amino groups

(a) Molecular structure



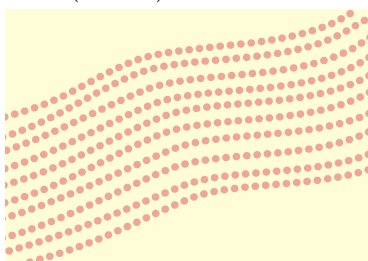
Cellulose is an unbranched polymer of glucose with β -1,4 glycosidic linkages that are chemically very stable.



Glycogen and starch are polymers of glucose with α -1,4 glycosidic linkages. α -1,6 glycosidic linkages produce branching at carbon 6.

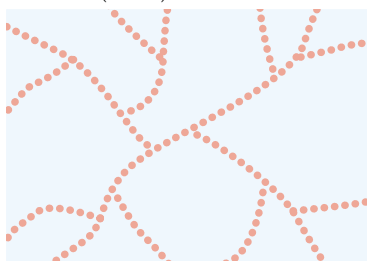
(b) Macromolecular structure

Linear (cellulose)



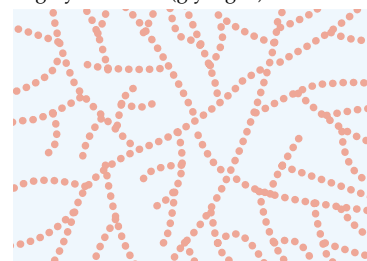
Parallel cellulose molecules from hydrogen-bonds, resulting in thin fibrils.

Branched (starch)



Branching limits the number of hydrogen bonds that can form in starch molecules, making starch less compact than cellulose.

Highly branched (glycogen)



The high amount of branching in glycogen makes its solid deposits more compact than starch.

(c) Polysaccharides in cells



Layers of cellulose fibrils, as seen in this scanning electron micrograph, give plant cell walls great strength.



Dyed purple in this micrograph, starch deposits have a large granular shape within cells.



Colored pink in this electron micrograph of human liver cells, glycogen deposits have a small granular shape.

3.16 Representative Polysaccharides Cellulose, starch, and glycogen demonstrate different levels of branching and compaction in polysaccharides.

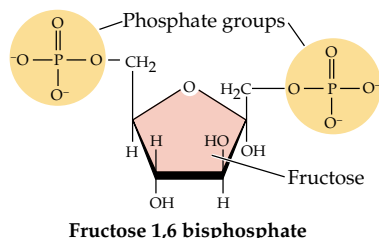
(Figure 3.17). For example, carbon 6 in glucose may be oxidized from $-\text{CH}_2\text{OH}$ to a carboxyl group ($-\text{COOH}$), producing glucuronic acid. Or a phosphate group may be added to one or more of the $-\text{OH}$ sites. Some of the resulting *sugar phosphates*, such as fructose 1,6-bisphosphate, are important intermediates in cellular energy reactions.

When an amino group is substituted for an $-\text{OH}$ group, *amino sugars*, such as glucosamine and galactosamine, are produced. These compounds are important in the extracellular matrix, where they form parts of proteins involved in keeping tissues together. Galactosamine is a major component of cartilage, the material that forms caps on the ends of bones and stiffens the protruding parts of the ears and nose. A derivative of glucosamine produces the polymer *chitin*, which is the principal structural polysaccharide in the skele-

50 CHAPTER THREE

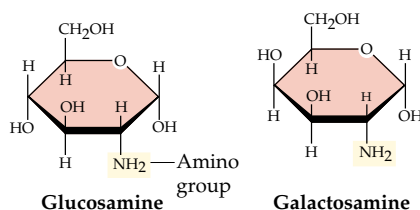
(a) Sugar phosphate

Fructose 1,6 bisphosphate is involved in the reactions that liberate energy from glucose. (The numbers in its name refer to the carbon sites of phosphate bonding; *bis*- indicates that two phosphates are present.)



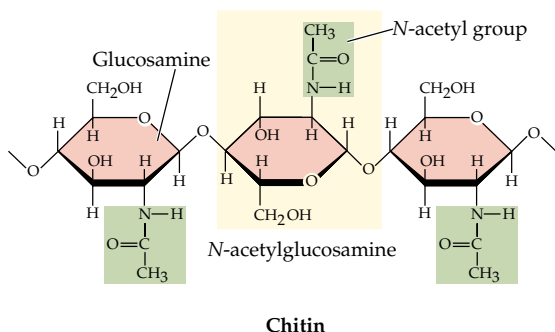
(b) Amino sugars

The monosaccharides glucosamine and galactosamine are amino sugars with an amino group in place of a hydroxyl group.



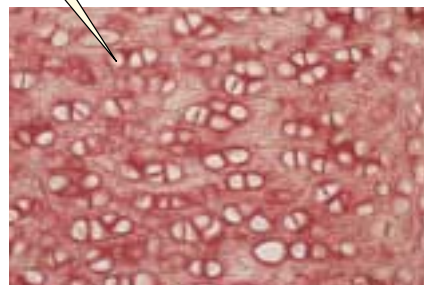
(c) Chitin

Chitin is a polymer of *N*-acetylglucosamine; *N*-acetyl groups provide additional sites for hydrogen bonding between the polymers.



3.17 Chemically Modified Carbohydrates Added functional groups modify the form and properties of a carbohydrate.

Galactosamine is an important component of cartilage, a connective tissue in vertebrates.



The external skeletons of insects are made up of chitin.



tons of insects, crabs, and lobsters, as well as in the cell walls of fungi. Fungi and insects (and their relatives) constitute more than 80 percent of the species ever described, and so chitin is one of the most abundant substances on Earth.

Lipids: Water-Insoluble Molecules

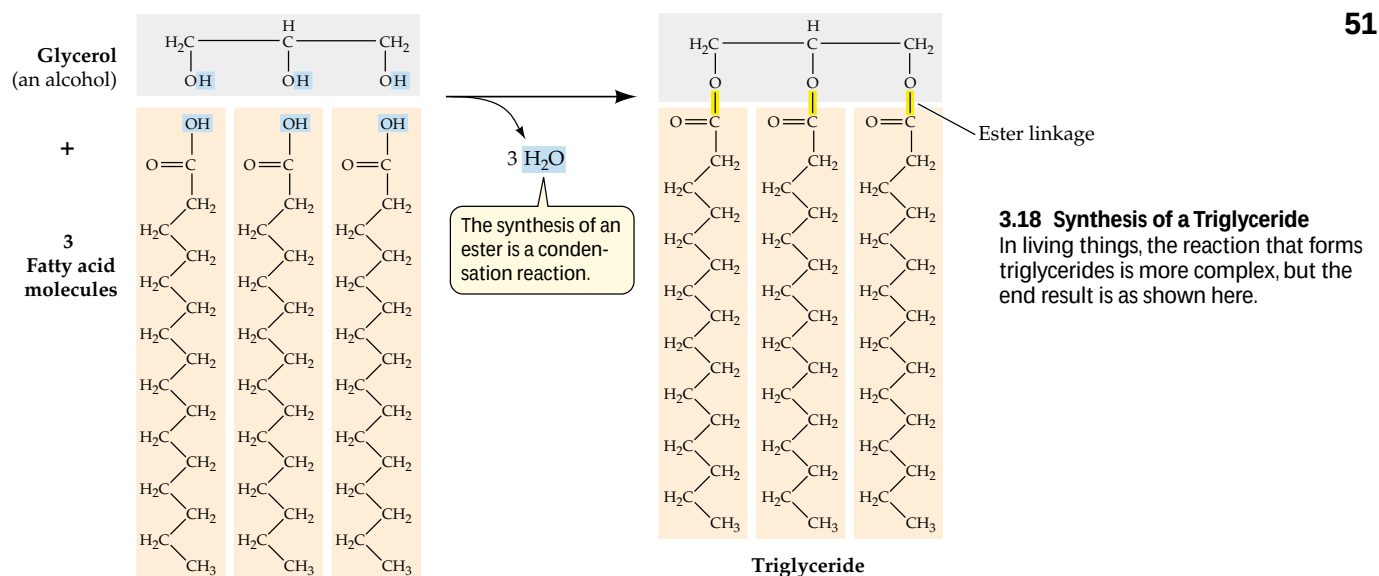
The **lipids** are a chemically diverse group of hydrocarbons. The property they all share is insolubility in water, which is due to the presence of many nonpolar covalent bonds. As we saw in Chapter 2, nonpolar hydrocarbon molecules are hydrophobic and preferentially aggregate among themselves, away from water, which is polar. When these nonpolar molecules are sufficiently close together, weak but additive van der Waals forces hold them together. These huge macromolecular aggregations are not polymers in a strict chemical sense, since their units (lipid molecules) are not held together by covalent bonds, as are, for example, the amino acids in proteins. But they can be considered polymers of individual lipid units.

In this section, we will describe the different types of lipids. Lipids have a number of roles in living organisms:

- ▶ Fats and oils store energy.
- ▶ Phospholipids play important structural roles in cell membranes.
- ▶ The carotenoids help plants capture light energy.
- ▶ Steroids and modified fatty acids play regulatory roles as hormones and vitamins.
- ▶ The fat in animal bodies serves as thermal insulation.
- ▶ A lipid coating around nerves acts as electrical insulation.
- ▶ Oil or wax on the surfaces of skin, fur, and feathers repels water.

Fats and oils store energy

Chemically, fats and oils are **triglycerides**, also known as *simple lipids*. Triglycerides that are solid at room temperature (20°C) are called *fats*; those that are liquid at room temperature are called *oils*. Triglycerides are composed of two types of building blocks: fatty acids and glycerol. *Glycerol* is a small molecule with three hydroxyl (—OH) groups (an alcohol). A *fatty acid* is made up of a long nonpolar hydrocarbon chain and a polar carboxyl group (—COOH). A triglyceride contains three fatty acid molecules and one molecule of glycerol.



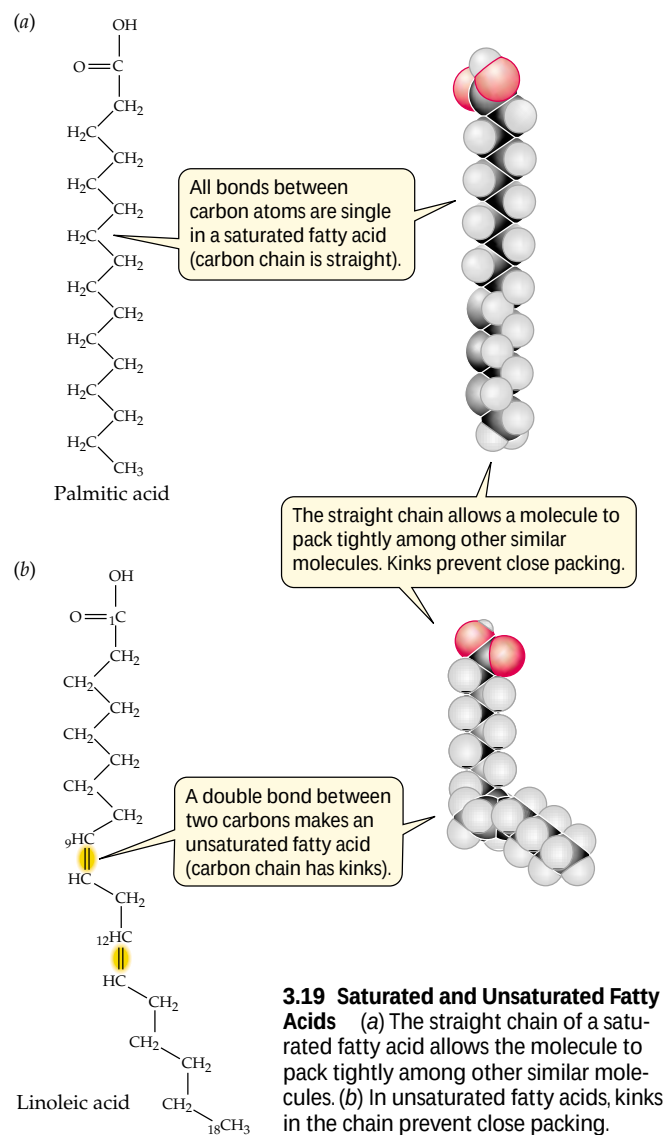
The carboxyl group of a fatty acid can form a covalent bond with the hydroxyl group of glycerol, resulting a functional group called an *ester* and water (Figure 3.18).

The three fatty acids in a triglyceride molecule need not all have the same hydrocarbon chain length or structure:

- In **saturated** fatty acids, all the bonds between the carbon atoms in the hydrocarbon chain are single bonds—there are no double bonds. That is, all the bonds are saturated with hydrogen atoms (Figure 3.19a). These fatty acid molecules are relatively rigid and straight, and they pack together tightly, like pencils in a box.
- In **unsaturated** fatty acids, the hydrocarbon chain contains one or more double bonds. Oleic acid, for example, is a *monounsaturated* fatty acid that has one double bond near the middle of the hydrocarbon chain, which causes a kink in the molecule (Figure 3.19b). Some fatty acids have more than one double bond—are *polyunsaturated*—and have multiple kinks. These kinks prevent the molecules from packing together tightly.

The kinks in fatty acid molecules are important in determining the fluidity and melting point of a lipid. The triglycerides of animal fats tend to have many long-chain saturated fatty acids, packed tightly together; these fats are usually solids at room temperature and have a high melting point. The triglycerides of plants, such as corn oil, tend to have short or unsaturated fatty acids. Because of their kinks, these fatty acids pack together poorly and have a low melting point, and these triglycerides are usually liquids at room temperature.

Fats and oils are marvelous storehouses for energy. When they take in excess food, many animal species deposit fat droplets in their cells as a means of storing energy. Some plant species, such as olives, avocados, sesame, castor beans, and all nuts, have substantial amounts of lipids in their seeds



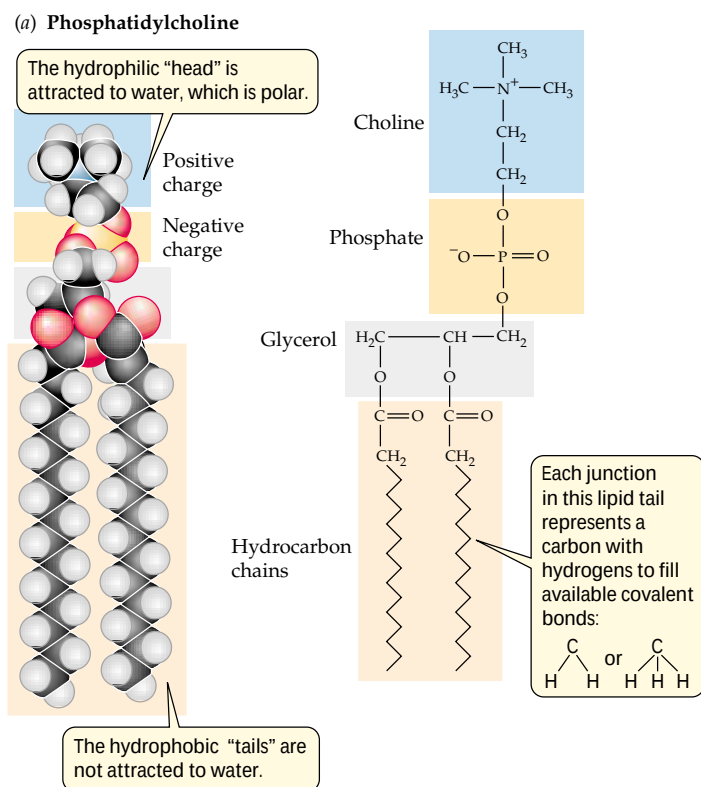
52 CHAPTER THREE

or fruits that serve as energy reserves for the next generation. This energy can be tapped by people who eat these plant oils or use them for fuel. Indeed, the famous German engineer Rudolf Diesel used peanut oil to power one of his early automobile engines in 1900.

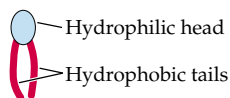
Phospholipids form the core of biological membranes

Because lipids and water do not interact, a mixture of water and lipids forms two distinct layers. Many biologically important substances—such as ions, sugars, and free amino acids—that are soluble in water are insoluble in lipids.

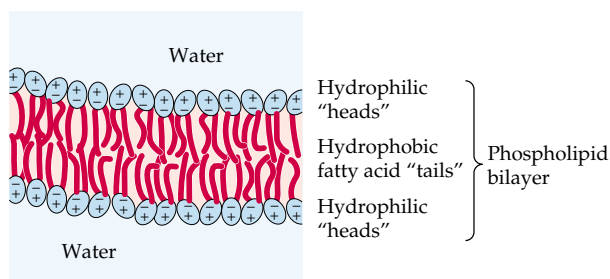
Like triglycerides, **phospholipids** contain fatty acids bound to glycerol by ester linkages. In phospholipids, however, any one of several phosphate-containing compounds



(b) **Membrane phospholipid, generalized symbol**



3.20 Phospholipid Structure (a) Phosphatidylcholine (lecithin) demonstrates the structure of a phospholipid molecule. In other phospholipids, the amino acid serine, the sugar alcohol inositol, or other compounds replace choline. (b) This generalized symbol is used throughout this book to represent a membrane phospholipid.



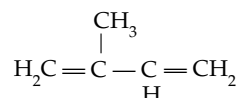
3.21 Phospholipids Form a Bilayer In an aqueous environment, hydrophobic interactions bring the "tails" of phospholipids together in the interior of a phospholipid bilayer. The hydrophilic "heads" face outward on both sides of the bilayer, where they interact with the surrounding water molecules.

replaces one of the fatty acids (Figure 3.20). The phosphate functional group has a negative electric charge, so this portion of the molecule is hydrophilic, attracting polar water molecules. But the two fatty acids are hydrophobic, so they tend to aggregate away from water.

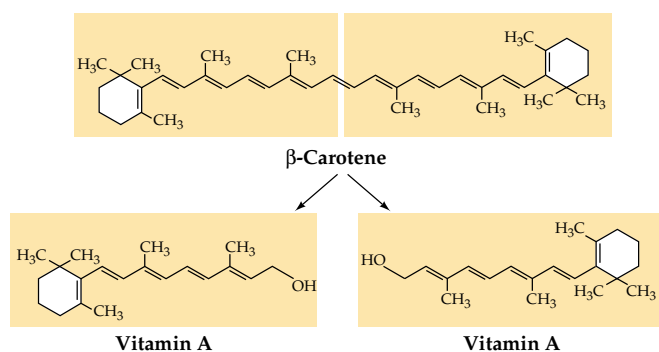
In an aqueous environment, phospholipids line up in such a way that the nonpolar, hydrophobic "tails" pack tightly together and the phosphate-containing "heads" face outward, where they interact with water. The phospholipids thus form a *bilayer*, a sheet two molecules thick, with water excluded from the core (Figure 3.21). Biological membranes have this kind of phospholipid bilayer structure, and we will devote all of Chapter 5 to their biological functions.

Carotenoids and steroids

The next two lipid classes we'll discuss—the carotenoids and the steroids—have chemical structures very different from those of triglycerides and phospholipids and from each other. Both carotenoids and steroids are synthesized by covalent linking and chemical modification of isoprene to form a series of isoprene units:



CAROTENOIDS TRAP LIGHT ENERGY. The **carotenoids** are a family of light-absorbing pigments found in plants and animals. Beta-carotene (β -carotene) is one of the pigments that traps light energy in leaves during photosynthesis. In humans, a molecule of β -carotene can be broken down into two vitamin A molecules (Figure 3.22), from which we make the pigment rhodopsin, which is required for vision. Carotenoids are responsible for the colors of carrots, tomatoes, pumpkins, egg yolks, and butter.



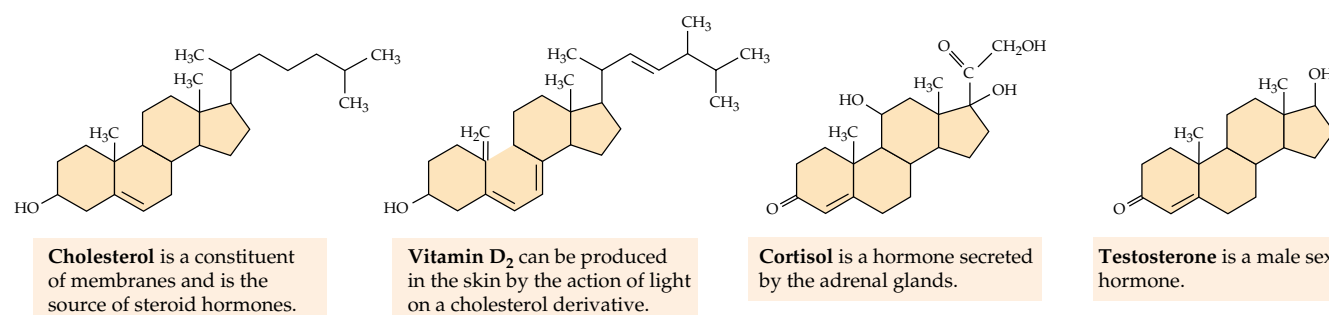
3.22 β-Carotene is the Source of Vitamin A The carotenoid β-carotene is symmetrical around its central double bond; when split, β-carotene becomes two vitamin A molecules. The simplified structural formula used here is standard chemical shorthand for large organic molecules with many carbon atoms. Structural formulas are simplified by omitting the C (indicating a carbon atom) at the intersections of the lines representing covalent bonds. Hydrogen atoms (H) to fill all the available bonding sites on each C are assumed.

STERIODS ARE SIGNAL MOLECULES. The **steroids** are a family of organic compounds whose multiple rings share carbons (Figure 3.23). The steroid cholesterol is an important constituent of membranes. Other steroids function as hormones, chemical signals that carry messages from one part of the body to another. Testosterone and the estrogens are steroid hormones that regulate sexual development in vertebrates. Cortisol and related hormones play many regulatory roles in the digestion of carbohydrates and proteins, in the maintenance of salt balance and water balance, and in sexual development.

Cholesterol is synthesized in the liver and is the starting material for making testosterone and other steroid hormones, as well as the bile salts that help break down dietary fats so that they can be digested. Cholesterol is absorbed from foods such as milk, butter, and animal fats.

Some lipids are vitamins

Vitamins are small molecules that are not synthesized by the body, but are necessary for its normal functioning. Vitamins must be acquired from dietary sources.



- ▶ **Vitamin A** is formed from the β-carotene found in green and yellow vegetables (see Figure 3.22). In humans, a deficiency of vitamin A leads to dry skin, eyes, and internal body surfaces, retarded growth and development, and night blindness, which is a diagnostic symptom for the deficiency.
- ▶ **Vitamin D** regulates the absorption of calcium from the intestines. It is necessary for the proper deposition of calcium in bones; a deficiency of vitamin D can lead to rickets, a bone-softening disease.
- ▶ **Vitamin E** seems to protect cells from the damaging effects of oxidation–reduction reactions. For example, it has an important role in preventing unhealthy changes in the double bonds in the unsaturated fatty acids of membrane phospholipids. Commercially, vitamin E is added to some foods to slow spoilage.
- ▶ **Vitamin K** is found in green leafy plants and is also synthesized by bacteria normally present in the human intestine. This vitamin is essential to the formation of blood clots.

Inadequate vitamin intake can lead to deficiency diseases.

Wax coatings repel water

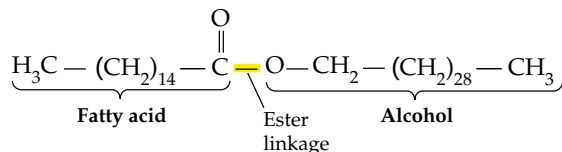
The sheen on human hair is not there only for cosmetic purposes. Glands in the skin secrete a waxy coating that repels water and keeps the hair pliable. Birds that live near water have a similar waxy coating on their feathers. The shiny leaves of holly plants, familiar during winter holidays, also have a waxy coating. Finally, bees make their honeycombs out of wax.

All waxes have the same basic structure: They are formed by an ester linkage between a saturated, long-chain fatty acid and a saturated, long-chain alcohol. The result is a very long

3.23 All Steroids Have the Same Ring Structure The steroids shown here, all important in vertebrates, are composed of carbon and hydrogen and are highly hydrophobic. However, small chemical variations, such as the presence or absence of a methyl or hydroxyl group, can produce enormous functional differences.

54 CHAPTER THREE

molecule, with 40–60 CH₂ groups. For example, here is the structure of beeswax:



This highly nonpolar structure accounts for the impermeability of wax to water.

Nucleic Acids: Informational Macromolecules That Can Be Catalytic

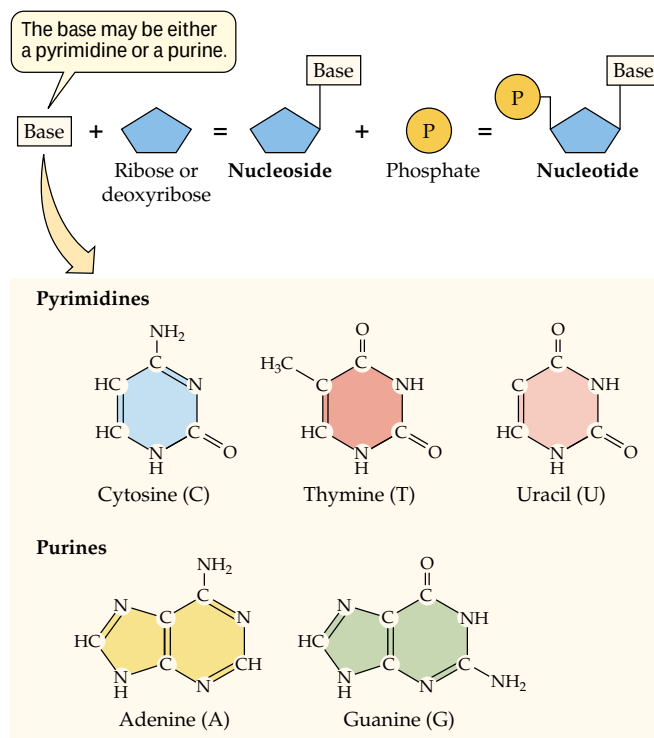
The **nucleic acids** are polymers specialized for the storage, transmission, and use of information. There are two types of nucleic acids: **DNA** (deoxyribonucleic acid) and **RNA** (ribonucleic acid). DNA molecules are giant polymers that encode hereditary information and pass it from generation to generation. Through an RNA intermediate, the information encoded in DNA is also used to specify the amino acid sequence of proteins. Information flows from DNA to DNA in reproduction, but in the nonreproductive activities of the cell, information flows from DNA to RNA to proteins, which ultimately carry out these functions. In addition, certain RNAs act as catalysts for important reactions in cells.

The nucleic acids have characteristic chemical properties

Nucleic acids are composed of monomers called **nucleotides**, each of which consists of a pentose sugar, a phosphate group, and a nitrogen-containing **base**—either a pyrimidine or a purine (Figure 3.24). (Molecules consisting of a pentose sugar and a nitrogenous base, but no phosphate group, are called **nucleosides**.) In DNA, the pentose sugar is deoxyribose, which differs from the ribose found in RNA by one oxygen atom (see Figure 3.14).

In both RNA and DNA, the backbone of the macromolecule consists of alternating pentose sugars and phosphates (sugar—phosphate—sugar—phosphate). The bases are attached to the sugars and project from the chain (Figure 3.25). The nucleotides are joined by *phosphodiester linkages* between the sugar of one nucleotide and the phosphate of the next (*-diester* refers to the two covalent bonds formed by —OH groups reacting with acidic phosphate groups). The phosphate groups link carbon 3 in one pentose sugar to carbon 5 in the adjacent sugar.

Most RNA molecules consist of only one polynucleotide chain. DNA, however, is usually double-stranded; it has two polynucleotide strands held together by hydrogen bonding between their nitrogenous bases. The two strands of DNA run in opposite directions. You can see what this means by



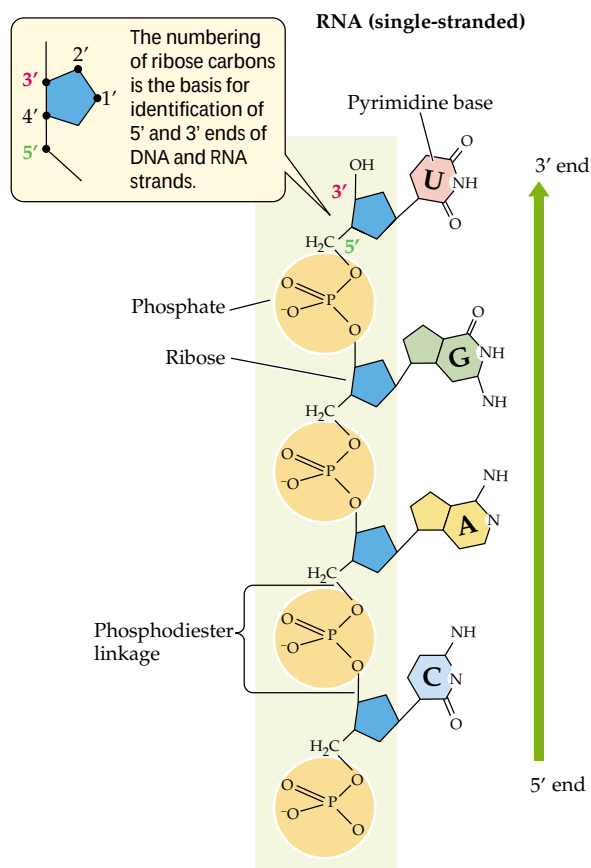
3.24 Nucleotides Have Three Components A nucleotide consists of a phosphate group, a pentose sugar (ribose or deoxyribose), and a nitrogen-containing base, all linked together by covalent bonds. The nitrogenous bases fall into two categories: Purines have two fused rings, and the smaller pyrimidines have a single ring.

drawing an arrow through the phosphate group from carbon 5' to carbon 3' in the next ribose. If you do this for both strands of the DNA in Figure 3.25, the arrows will point in opposite directions. This antiparallel orientation is necessary for the strands to fit together in three-dimensional space.

The uniqueness of a nucleic acid resides in its nucleotide sequence

Only four nitrogenous bases—and thus only four nucleotides—are found in DNA. The DNA bases and their abbreviations are adenine (A), cytosine (C), guanine (G), and thymine (T). A key to understanding the structure and function of nucleic acids is the principle of **complementary base pairing**. In double-stranded DNA, adenine and thymine always pair (A-T), and cytosine and guanine always pair (C-G).

Base pairing is complementary because of three factors: the sites for for hydrogen bonding on each base, the geometry of the sugar-phosphate backbone, which brings opposite bases near each other, and the molecular sizes of the paired bases. Adenine and guanine are both purines, consisting of two fused rings. Thymine and cytosine are both pyrimidines, consisting of only one ring. The pairing of a large purine with



In RNA, the bases are attached to ribose. The bases in RNA are the purines adenine (A) and guanine (G) and the pyrimidines cytosine (C) and uracil (U).



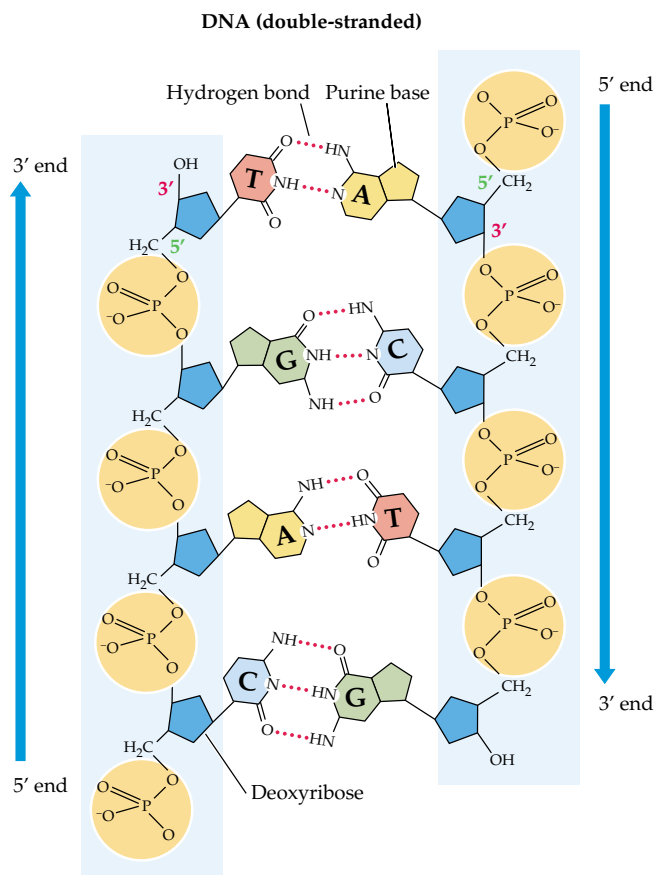
3.25 Distinguishing Characteristics of DNA and RNA

RNA is usually a single strand. DNA usually consists of two strands running in opposite directions.

a small pyrimidine ensures stability and consistency in the double-stranded molecule of DNA.

Ribonucleic acids are also made up of four different monomers, but their nucleotides differ from those of DNA. In RNA the nucleotides are termed *ribonucleotides* (the ones in DNA are *deoxyribonucleotides*). They contain ribose rather than deoxyribose, and instead of the base thymine, RNA uses the base uracil (U) (Table 3.3). The other three bases are the same as in DNA.

Although RNA is generally single-stranded, complementary hydrogen bonding between ribonucleotides can take place. These bonds play important roles in determining the shapes of some RNA molecules and in associations between RNA molecules during protein synthesis (Figure 3.26). When the base sequence of DNA is copied in the synthesis of RNA, complementary base pairing also takes place



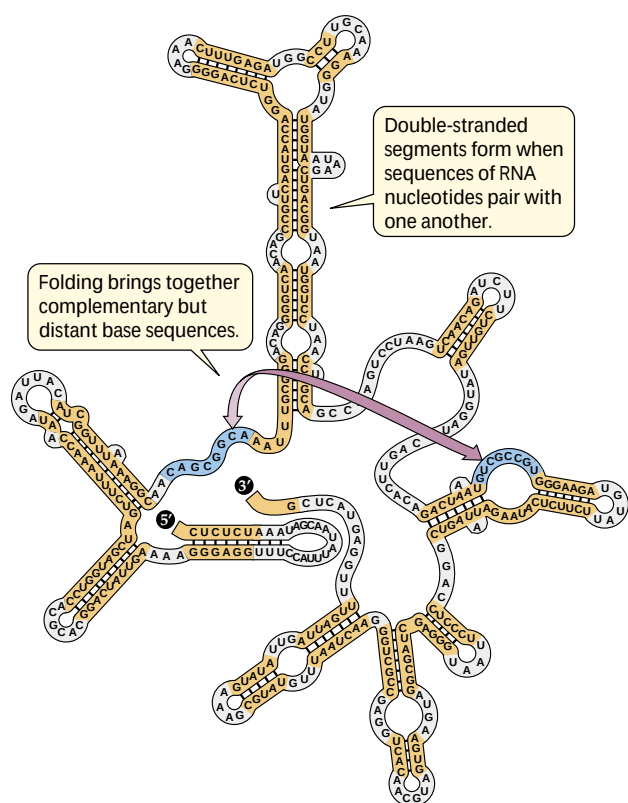
In DNA, the bases are attached to deoxyribose, and the base thymine (T) is found instead of uracil. Hydrogen bonds between purines and pyrimidines hold the two strands of DNA together.

between ribonucleotides and deoxyribonucleotides. In RNA, guanine and cytosine pair (G-C), as in DNA, but adenine pairs with uracil (A-U). Adenine in an RNA strand can pair either with uracil (in another RNA strand) or with thymine (in a DNA strand).

3.3 Distinguishing RNA from DNA

NUCLEIC ACID	SUGAR	BASES
RNA	Ribose	Adenine Cytosine Guanine Uracil
DNA	Deoxyribose	Adenine Cytosine Guanine Thymine

56 CHAPTER THREE



3.26 Hydrogen Bonding in RNA When a single-stranded RNA folds in on itself, hydrogen bonds can stabilize it into a three-dimensional shape.

DNA is a purely *informational* molecule. The information in DNA is encoded in the sequence of bases carried in its strands—the information encoded in the sequence TCAG is different from the information in the sequence CCAG. The information can be read easily and reliably, in a specific order.

The three-dimensional appearance of DNA is strikingly uniform. The segment shown in Figure 3.27 could be from any DNA molecule. The variations in DNA—the different sequences of bases—are strictly “internal.” Through hydrogen bonding, the two complementary polynucleotide strands

pair and twist to form a double helix. When compared with the complex and varied tertiary structures of different proteins, this uniformity is surprising. But this structural contrast makes sense in terms of the functions of these two classes of macromolecules.

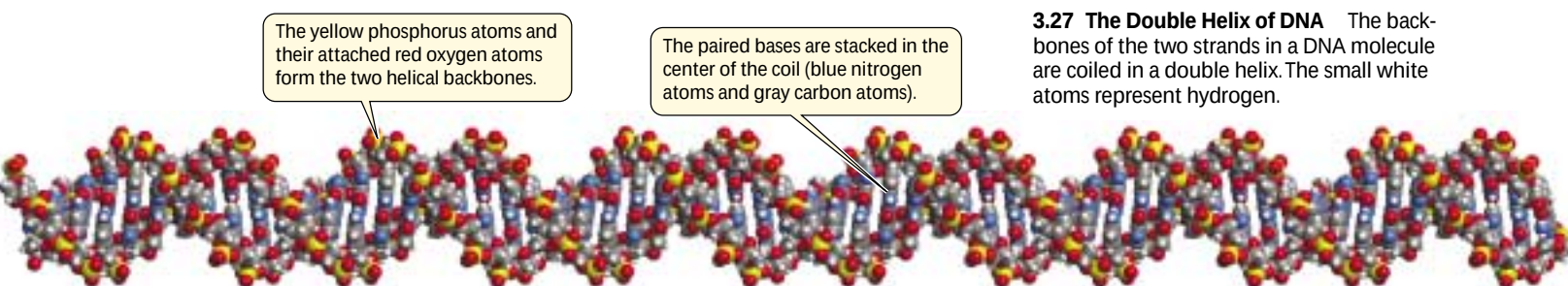
It is their different and unique shapes that permit proteins to recognize specific “target” molecules. The unique three-dimensional form of each protein matches at least a portion of the surface of the target molecule. In other words, structural diversity in the molecules to which proteins bind requires corresponding diversity in the structure of the proteins themselves.

In DNA, then, the information is in the sequence of the bases; in proteins, the information is in the shape of the molecule.

DNA is a guide to evolutionary relationships

Because DNA carries hereditary information between generations, a theoretical series of DNA molecules with changes in base sequences stretches back through evolutionary time. Of course, we cannot study all of these DNA molecules, because many of their organisms have become extinct. However, we can study the DNA of living organisms, which are judged to have changed little over millions of years. Comparisons and contrasts of these DNA molecules can be added to evidence from fossils and other sources to reveal the evolutionary record, as we will see in Chapter 24.

Closely related living species should have more similar base sequences than species judged by other criteria to be more distantly related. The examination of base sequences has confirmed many of the evolutionary relationships that have been inferred from the more traditional study of body structures, biochemistry, and physiology. For example, the closest living relative of humans (*Homo sapiens*) is the chimpanzee (genus *Pan*), which shares more than 98 percent of its DNA base sequence with human DNA. This confirmation of well-established evolutionary relationships gives credibility to the use of DNA to elucidate relationships when studies of structure are not possible or are not conclusive. For example,



3.27 The Double Helix of DNA The backbones of the two strands in a DNA molecule are coiled in a double helix. The small white atoms represent hydrogen.

DNA studies revealed a close evolutionary relationship between starlings and mockingbirds that was not expected on the basis of their anatomy or behavior.

DNA studies support the division of the prokaryotes into two domains, Bacteria and Archaea. Each of these two groups of prokaryotes is as distinct from the other as either is from the Eukarya, the third domain into which living things are classified (see Chapter 1). In addition, DNA comparisons support the hypothesis that certain subcellular compartments of eukaryotes (the organelles called mitochondria and chloroplasts) evolved from early bacteria that established a stable and mutually beneficial way of life inside larger cells.

RNA may have been the first biological catalyst

The three-dimensional structure of a folded RNA molecule presents a unique surface to the external environment (see Figure 3.26). These surfaces are every bit as specific as those of proteins. We noted above that an important role of proteins in biology is to act as catalysts, speeding up reactions that would ordinarily take place too slowly to be biologically useful, and that the spatial property of proteins is vital to this role.

As we will see, certain RNA molecules can also act as catalysts, using their three-dimensional shapes and other chemical properties. They can catalyze reactions on their own nucleotides as well as in other cellular substances. These catalytic RNAs are called **ribozymes**. Their discovery had implications for theories of the origin of life.

The Miller-Urey experiment and other such experiments in prebiotic chemistry yielded both amino acids and nucleotides. Organisms can synthesize both RNA and proteins from these monomers. As we noted above, in current organisms on Earth, protein synthesis requires DNA and RNA, and nucleic acid synthesis requires proteins (as enzymes). So the question is, when life originated, which came first, the proteins or the nucleic acids?

The discovery of catalytic RNAs provided a solution to this dilemma and led to the hypothesis that early life was part of an “RNA world.” RNA can be informational (in its nucleotide sequence) as well as catalytic. So when RNA was first made, it could have acted as a catalyst for its own replication, as well as for the synthesis of proteins. Then DNA could have eventually evolved by being made from RNA. There is some laboratory evidence supporting this scenario:

- ▶ RNAs of different sequences have been put in a test tube and made to replicate on their own. Such self-replicating ribozymes speed up the synthesis of RNA 7 million-fold.
- ▶ In living organisms today, the formation of peptide linkages (see Figure 3.5) is catalyzed by a ribozyme.

- ▶ In certain viruses called retroviruses, there is an enzyme called reverse transcriptase that catalyzes the synthesis of DNA from RNA.

Nucleotides have other important roles

Nucleotides are more than just the building blocks of nucleic acids. As we will see in later chapters, there are several nucleotides with other functions:

- ▶ ATP (adenosine triphosphate) acts as an energy transducer in many biochemical reactions (see Chapter 6).
- ▶ GTP (guanosine triphosphate) serves as an energy source, especially in protein synthesis. It also has a role in the transfer of information from the environment to the body tissues (see Chapters 12 and 15).
- ▶ cAMP (cyclic adenosine monophosphate), a special nucleotide in which a bond forms between the sugar and phosphate groups within adenosine monophosphate, is essential in many processes, including the actions of hormones and the transmission of information by the nervous system (see Chapter 15).

All Life from Life

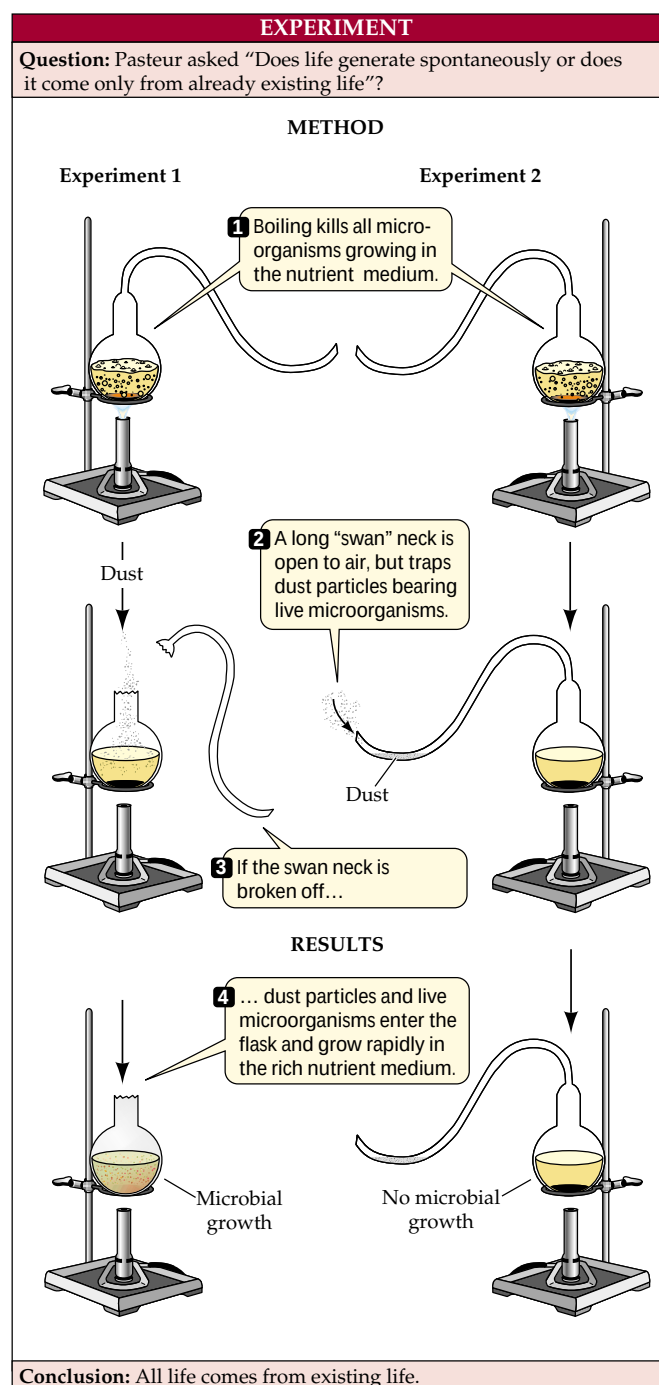
The concepts conveyed throughout this chapter—that large molecules obey the mechanistic laws of physics and chemistry, and that life could have arisen from inanimate, self-replicating macromolecules—have come to be generally accepted by the scientific community. So should we expect to see new life forms arise at any time from the biochemical environment?

During the Renaissance (a period from about 1350 to 1700 A.D., marked by the birth of modern science), most people thought that at least some forms of life arose directly from inanimate or decaying matter by *spontaneous generation*. For instance, it was suggested that mice arose from sweaty clothes placed in dim light, frogs came from moist soil, and flies were produced from meat. These ideas were attacked by scientists such as the Italian doctor and poet Francisco Redi using the relatively new idea of using experiments to test an idea. In 1668, Redi proposed that flies arose not by some mysterious transformation of decaying meat, but from other flies, who laid their eggs on the meat. The eggs developed into wormlike maggots (the immature form of flies). Redi set out several jars containing chunks of meat.

- ▶ One jar contained meat exposed both to the air and to flies.
- ▶ A second jar contained meat in a container wrapped in a fine cloth so that the meat was exposed to the air, but not to flies.
- ▶ The meat in the third jar was in a sealed container and thus was not exposed to either air or flies.

58 CHAPTER THREE

As he had hypothesized, Redi found maggots, which then hatched into flies, only in the first container. The idea that a complex organism like a fly could come from a totally different substance was laid to rest.



3.28 Disproving the Spontaneous Generation of Life
Louis Pasteur's classic experiments showed that, under today's conditions, an inanimate solution remains lifeless unless a living organism contaminates it.

With the invention of the microscope in the 1660s, a vast new biological world was unveiled. Under microscopic observation, virtually every environment on Earth was found to be teeming with tiny organisms such as bacteria. Some scientists believed that these organisms arose spontaneously from their rich chemical environment.

The experiments that disproved this idea were done by the great French scientist Louis Pasteur. His experiments showed that microorganisms come only from other microorganisms, and that an environment without life remains lifeless unless contaminated by living creatures (Figure 3.28).

These experiments by Redi, Pasteur, and others provided solid evidence that neither small (bacteria) nor large (flies) organisms come from inanimate matter, but instead come from living parent organisms.

Indeed, life on Earth no longer arises from nonliving materials. This is because the atmospheric and planetary conditions that exist on Earth today are vastly different from those on the prebiotic, anaerobic planet. The oxygen present in today's atmosphere would break down the prebiotic molecules before they could accumulate. In addition, the necessary energy sources—including constant lightning strikes, immense volcanic eruptions, and bombardment by intense ultraviolet light—are no longer present with anything like their primeval force.

Chapter Summary

Theories of the Origin of Life

- ▶ Life may have come from outside Earth. The evidence for this proposal comes primarily from chemicals contained in meteorites that have landed on Earth.
- ▶ The theory of chemical evolution proposes that life on Earth originated on Earth. Experiments using model systems that attempt to duplicate the ancient Earth have shown that chemical evolution could have produced the four types of macromolecules that distinguish living things. **Review Figure 3.1. See Web/CD Tutorial 3.1**

Macromolecules: Giant Polymers

- ▶ Macromolecules are polymers constructed by the formation of covalent bonds between smaller molecules called monomers. Macromolecules in living organisms include polysaccharides, proteins, and nucleic acids. **Review Figure 3.2 and Table 3.1**
- ▶ Macromolecules have specific, characteristic three-dimensional shapes that depend on the structure, properties, and sequence of their monomers.
- ▶ Different functional groups give local sites on macromolecules specific properties that are important for their biological functioning and their interactions with other macromolecules. **See Web/CD Tutorial 3.2**

Condensation and Hydrolysis Reactions

- ▶ Monomers are joined by condensation reactions, which release a molecule of water for each bond formed. Hydrolysis

reactions use water to break polymers into monomers. **Review Figure 3.3**

Proteins: Polymers of Amino Acids

- ▶ The functions of proteins include support, protection, catalysis, transport, defense, regulation, and movement. Protein function sometimes requires an attached prosthetic group.
- ▶ There are 20 amino acids found in proteins. Each amino acid consists of an amino group, a carboxyl group, a hydrogen, and a side chain bonded to the α carbon atom. **Review Table 3.2**
- ▶ The side chains, or R groups, of amino acids may be charged, polar, or hydrophobic; there are also special cases, such as the —SH groups of cysteine, which can form disulfide bridges. The side chains give different properties to each of the amino acids. **Review Table 3.2 and Figure 3.4**
- ▶ Amino acids are covalently bonded together into polypeptide chains by peptide linkages, which form by condensation reactions between the carboxyl and amino groups. **Review Figure 3.5**
- ▶ Polypeptide chains are folded into specific three-dimensional shapes to form functional proteins. Four levels of protein structure are possible: primary, secondary, tertiary, and quaternary.
- ▶ The primary structure of a protein is the sequence of amino acids bonded by peptide linkages. This primary structure determines both the higher levels of structure and protein function. **Review Figure 3.6a**
- ▶ The two types of secondary structure— α helices and β pleated sheets—are maintained by hydrogen bonds between atoms of the amino acid residues. **Review Figure 3.6b,c**
- ▶ The tertiary structure of a protein is generated by bending and folding of the polypeptide chain. **Review Figures 3.6d, 3.7**
- ▶ The quaternary structure of a protein is the arrangement of two or more polypeptides into a single functional protein consisting of two or more polypeptide subunits. **Review Figures 3.6e, 3.8**
- ▶ Weak chemical interactions are important in the three-dimensional structure of proteins and in their binding to other molecules. **Review Figure 3.9, 3.10**
- ▶ Proteins denatured by heat, alterations in pH, or certain chemicals lose their tertiary and secondary structure as well as their biological function. Renaturation is not often possible. **Review Figure 3.11**
- ▶ Chaperonins assist protein folding by preventing binding to inappropriate ligands. **Review Figure 3.12**

Carbohydrates: Sugars and Sugar Polymers

- ▶ All carbohydrates contain carbon bonded to hydrogen atoms and hydroxyl groups.
- ▶ Hexoses are monosaccharides that contain six carbon atoms. Examples of hexoses include glucose, galactose, and fructose, which can exist as chains or rings. **Review Figures 3.13, 3.14.**
See Web/CD Activity 3.1
- ▶ The pentoses are five-carbon monosaccharides. Two pentoses, ribose and deoxyribose, are components of the nucleic acids RNA and DNA, respectively. **Review Figure 3.14**
- ▶ Glycosidic linkages may have either α or β orientation in space. They covalently link monosaccharides into larger units such as disaccharides, oligosaccharides, and polysaccharides. **Review Figure 3.15**
- ▶ Cellulose, a very stable glucose polymer, is the principal component of the cell walls of plants. It is formed by glucose units linked together by β -glycosidic linkages between carbons 1 and 4. Starches, less dense and less stable than cellulose, store energy in plants. Starches and glycogen are formed by α -glycosidic

linkages between carbons 1 and 4 and are distinguished by the amount of branching they exhibit. **Review Figure 3.16**

- ▶ Chemically modified monosaccharides include the sugar phosphates and amino sugars. A derivative of the amino sugar glucosamine polymerizes to form the polysaccharide chitin, which is found in the cell walls of fungi and the exoskeletons of insects. **Review Figure 3.17**

Lipids: Water-Insoluble Molecules

- ▶ Although lipids can form gigantic structures, these aggregations are not chemically macromolecules because the individual units are not linked by covalent bonds.
- ▶ Fats and oils are triglycerides, composed of three fatty acids covalently bonded to a glycerol molecule by ester linkages. **Review Figure 3.18**
- ▶ Saturated fatty acids have a hydrocarbon chain with no double bonds. The hydrocarbon chains of unsaturated fatty acids have one or more double bonds that bend the chain, making close packing less possible. **Review Figure 3.29**
- ▶ Phospholipids have a hydrophobic hydrocarbon “tail” and a hydrophilic phosphate “head.” **Review Figure 3.20**
- ▶ In water, the interactions of the hydrophobic tails and hydrophilic heads of phospholipids generate a phospholipid bilayer that is two molecules thick. The head groups are directed outward, where they interact with the surrounding water. The tails are packed together in the interior of the bilayer. **Review Figure 3.21**
- ▶ Carotenoids trap light energy in green plants. Carotene can be split to form vitamin A, a lipid vitamin. **Review Figure 3.22**
- ▶ Some steroids, such as testosterone, function as hormones. Cholesterol is synthesized by the liver and has a role in cell membranes, as well as in the digestion of fats. **Review Figure 3.23**
- ▶ Vitamins are substances that are required for normal functioning, but must be acquired from the diet.

Nucleic Acids: Informational Macromolecules

- ▶ DNA is the hereditary material. Both DNA and RNA play roles in the formation of proteins. Information flows from DNA to RNA to protein.
- ▶ Nucleic acids are polymers made up of nucleotides. A nucleotide consists of a phosphate group, a sugar (ribose in RNA and deoxyribose in DNA), and a nitrogen-containing base. In DNA the bases are adenine, guanine, cytosine, and thymine, but in RNA uracil substitutes for thymine. **Review Figure 3.24 and Table 3.3. See Web/CD Activity 3.2**
- ▶ In the nucleic acids, the bases extend from a sugar-phosphate backbone. The information content of DNA and RNA resides in their base sequences. RNA is single-stranded. DNA is a double-stranded helix in which there is complementary, hydrogen-bonded base pairing between adenine and thymine (A-T) and guanine and cytosine (G-C). The two strands of the DNA double helix run in opposite directions. **Review Figures 3.25, 3.27. See Web/CD Activity 3.3**
- ▶ Base pairing of single-stranded RNAs can lead to three-dimensional structures, which can be catalytic. This finding has led to the proposal that in the origin of life, RNA preceded protein. **Review Figure 3.26**
- ▶ Comparing the DNA base sequences of different living species provides information on their evolutionary relationships.

All Life from Life

- ▶ One of the earliest conclusions from biology as a modern experimental science was that even the tiniest microbe comes from others of the same type—that is, that life begets life. **Review Figure 3.28. See Web/CD Tutorial 3.3**

60 CHAPTER THREE

► The conditions on primeval Earth that may have enabled life to arise from inanimate self-replicating chemicals no longer exist. Today all life comes from pre-existing life.

Self-Quiz

- The most abundant molecule in the cell is
 - carbohydrate.
 - lipid.
 - nucleic acid.
 - protein.
 - water.
- All lipids are
 - triglycerides.
 - polar.
 - hydrophilic.
 - polymers of fatty acids.
 - more soluble in nonpolar solvents than in water.
- All carbohydrates
 - are polymers.
 - are simple sugars.
 - consist of one or more simple sugars.
 - are found in biological membranes.
 - are more soluble in nonpolar solvents than in water.
- Which of the following is *not* a carbohydrate?
 - Glucose
 - Starch
 - Cellulose
 - Hemoglobin
 - Deoxyribose
- All proteins
 - are enzymes.
 - consist of one or more polypeptides.
 - are amino acids.
 - have quaternary structures.
 - are more soluble in nonpolar solvents than in water.
- Which of the following statements about the primary structure of a protein is *not* true?
 - It may be branched.
 - It is determined by the structure of the corresponding DNA.
 - It is unique to that protein.
 - It determines the tertiary structure of the protein.
 - It is the sequence of amino acids in the protein.
- The amino acid leucine (see Table 3.2)
 - is found in all proteins.
 - cannot form peptide linkages.
 - is likely to appear in the part of a membrane protein that lies within the phospholipid bilayer.
 - is likely to appear in the part of a membrane protein that lies outside the phospholipid bilayer.
 - is identical to the amino acid lysine.
- The quaternary structure of a protein
 - consists of four subunits—hence the name *quaternary*.
 - is unrelated to the function of the protein.
 - may be either alpha or beta.
 - depends on covalent bonding among the subunits.
 - depends on the primary structures of the subunits.
- All nucleic acids
 - are polymers of nucleotides.
 - are polymers of amino acids.
 - are double-stranded.
 - are double-helical.
 - contain deoxyribose.
- Which of the following statements about condensation reactions is *not* true?
 - Protein synthesis results from them.
 - Polysaccharide synthesis results from them.
 - Nucleic acid synthesis results from them.
 - They consume water as a reactant.
 - Different condensation reactions produce different kinds of macromolecules.

For Discussion

- Phospholipids make up a major part of most biological membranes; cellulose is the major constituent of the cell walls of plants. How do the chemical structures and physical properties of phospholipids and cellulose relate to their functions in cells?
- Suppose that, in a given protein, one lysine is replaced by aspartic acid (see Table 3.2). Does this change occur in the primary structure or in the secondary structure? How might it result in a change in tertiary structure? In quaternary structure?
- If there are 20 different amino acids commonly found in proteins, how many different dipeptides are there? How many different tripeptides? How many different trinucleotides? How many different single-stranded RNAs composed of 200 nucleotides?
- Contrast the following three structures, emphasizing the surfaces they present to their environment: hemoglobin; a DNA molecule; a protein that spans a biological membrane.
- Why might RNA have preceded proteins in the evolution of biological macromolecules?

3

Cells: The Basic Units of Life



Charles Darwin faced a dilemma. In his great book *On the Origin of Species*, published in 1859, he proposed the theory of natural selection to explain the gradual appearance and disappearance of different forms of animals and plants over long periods of time. But he realized that the fossil record, on which he based his theory, was incomplete, especially for the beginning of life. The oldest fossils that had been found in Darwin's time were complex organisms in rocks dated at about 550 million years ago (the Cambrian period). Where were the missing Precambrian fossils? These would surely provide a link to the origin of life.

As we saw in Chapter 3, conditions on Earth were probably suitable for the emergence of life by 4 billion years ago, about 600 million years after Earth began to form. But until recently there was no evidence for life older than the Cambrian. By the turn of the twentieth century, there was evidence for fossilized clumps of algae (simple aquatic photosynthetic organisms) in rocks at the base of the Grand Canyon that were close to 1 billion years old.

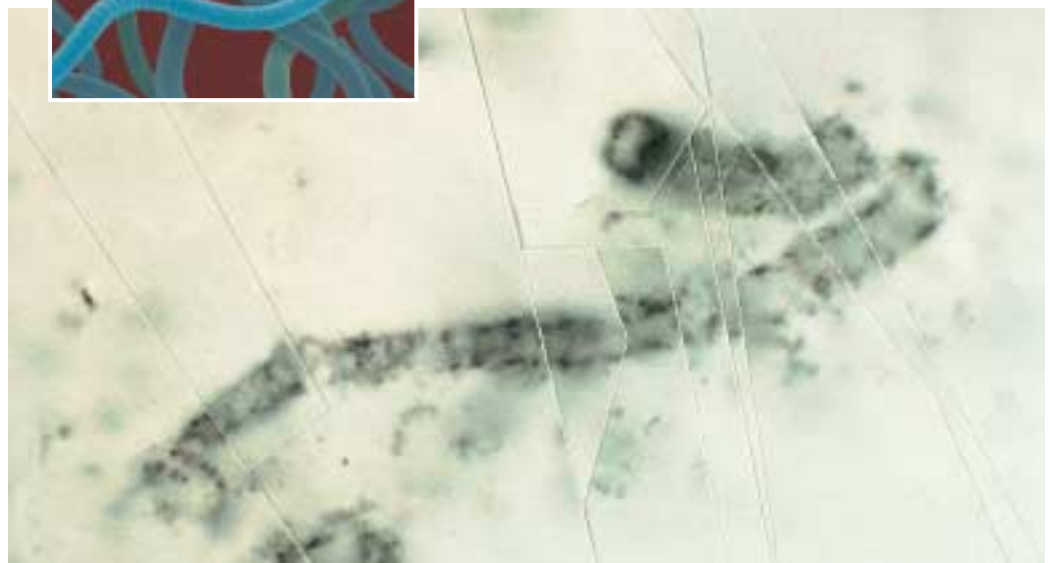
It took nearly another century to push the clock of life back nearer to its origins. In 1993, geologist J. William Schopf found fossilized chains of cylindrical objects, quite similar in size and shape to contemporary cyanobacteria ("blue-green algae"), in rocks in Western Australia that he dated at an astonishing 3.5 billion years old. He then used a chemical analysis method called laser Raman spectroscopy to show that these objects apparently contain carbon deposits that are chemical signatures of life.

Rounded or cylindrical objects in Earth's rocks or in a meteorite from Mars (see Chapter 3) get scientists excited because they realize that life is not just a bunch of macromolecules. Rather, life is macromolecules that can perform unique functions because they are enclosed in a structural compartment that is separate from the external environment. This separation allows living things to maintain a constant internal environment (homeostasis).

The "living compartment" is the cell, the subject of this chapter. The water-insoluble phospholipid structure (see Figure 3.21) that defines and contains cells is called the plasma membrane. It and its functions are so important that we will devote the entire next chapter to membranes. Subsequent chapters will be devoted to the chemical activities that take place inside all cells.



The Earliest Trace of Life? This fossil from Western Australia is 3.5 billion years old and shows carbon traces that indicate life. Its form is similar to that of modern filamentous cyanobacteria (inset).



4 Cellular Membranes



During his years as an undergraduate at Oxford, physics student Stephen Hawking took up rowing. Although he had never been much of an athlete, he was doing a passable job. But he noticed he was getting increasingly clumsy, and by the time he went to Cambridge as a graduate student, he was falling over for no particular reason. After weeks of tests, his physicians told him that he had motor neuron disease, an incurable condition in which the nerve cells that stimulate muscles gradually die and the patient loses all muscular control.

As the years progressed, Hawking made major contributions to theoretical physics, especially to the study of black holes and the origin of the universe. He ascended the ladder of academic success and now holds a professorship at Cambridge once held by Isaac Newton. But his disease has gotten worse, and he has lost almost all muscular control.

A hallmark of living cells is their ability to regulate the substances that enter and leave them. This regulation is a function of the plasma membrane, which is composed of a hydrophobic lipid bilayer with associated proteins. Muscle cells respond to stimulation by nerve cells by opening protein-lined channels in their plasma membranes. Because his nerves cannot stimulate them, the channels of Hawking's muscle cells do not open, and his muscles do not contract. Channels in plasma membranes underlie the biological activities of many organisms, ranging from the beating of an animal's heart to the opening of tiny pores in leaves to let outside air in.

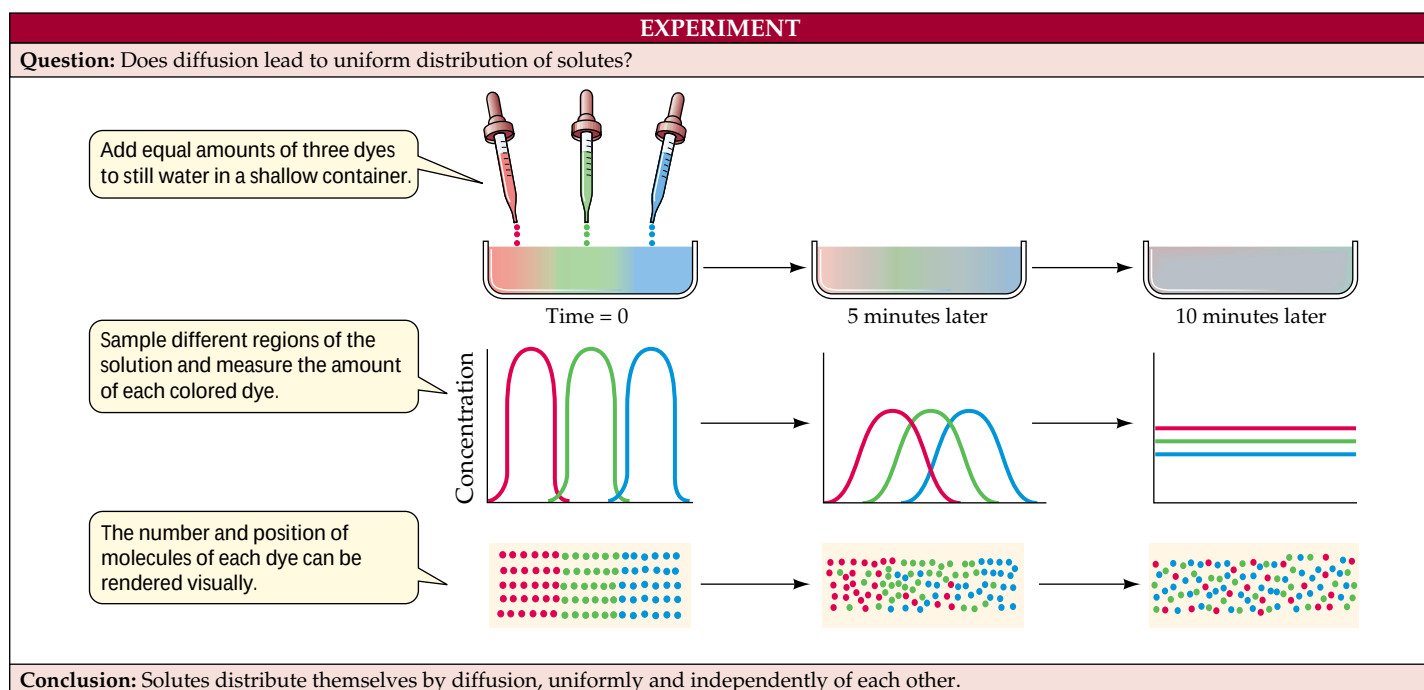
Membranes are dynamic structures whose components move and change. They perform their vital physiological roles by allowing cells to interact with other cells and with molecules in the environment. We describe the structural aspects of those interactions here. Membranes also regulate the traffic of chemicals into and out of the cell. The selective permeability of membranes, which we describe in this chapter, is an important characteristic of life. Later in this book, we will see it in action in such diverse situations as the transduction of light energy into chemical energy in the chloroplast and the retention of water and ions in the mammalian kidney.

Membrane Composition and Structure

The physical organization and functioning of all biological membranes depend on their constituents: lipids, proteins, and carbohydrates. The lipids establish the physical integrity of the membrane

Stephen Hawking The effects of motor neuron disease have confined the famous physicist to a wheelchair. A major cellular manifestation of this disease is the lack of ability of the nerve cells to stimulate the opening of channels through muscle cell membranes that would result in normal muscle function.





5.7 Diffusion Leads to Uniform Distribution of Solutes Diffusion is the net movement of a solute from regions of greater concentration to regions of lesser concentration. The speed of diffusion varies with the substances involved, but the process continues until the solution reaches equilibrium.

Diffusion is the process of random movement toward a state of equilibrium. Although the motion of each individual particle is absolutely random, in diffusion the *net* movement of particles is directional until equilibrium is reached. Diffusion is thus net movement from regions of greater concentration to regions of lesser concentration (Figure 5.7).

In a complex solution (one with many different solutes), the diffusion of each substance is independent of that of the others. How fast a substance diffuses depends on four factors:

- ▶ the *diameter* of the molecules or ions: smaller molecules diffuse faster
- ▶ the *temperature* of the solution: higher temperatures lead to faster diffusion
- ▶ the *electric charge*, if any, of the diffusing material: electric charge has a variable effect on diffusion
- ▶ the *concentration gradient* in the system—that is, the change in concentration with distance in a given direction. The greater the concentration gradient, the more rapidly a substance diffuses.

DIFFUSION WITHIN CELLS AND TISSUES. Within cells, or wherever distances are very short, solutes distribute themselves rapidly by diffusion. Small molecules and ions may move from one end of an organelle to another in a millisecond

(10^{-3} s). On the other hand, the usefulness of diffusion as a transport mechanism declines drastically as distances become greater. In the absence of mechanical stirring, diffusion across more than a centimeter may take an hour or more, and diffusion across meters may take years! Diffusion would not be adequate to distribute materials over the length of the human body, but within our cells or across layers of one or two cells, diffusion is rapid enough to distribute small molecules and ions almost instantaneously.

DIFFUSION ACROSS MEMBRANES. In a solution without barriers, all the solutes diffuse at rates determined by temperature, their physical properties, and the concentration gradient of each solute. If a biological membrane divides the solution into separate compartments, the movement of the different solutes can be affected by the properties of the membrane. The membrane is said to be *permeable* to solutes that can cross it more or less easily, but *impermeable* to substances that cannot move across it.

Molecules to which the membrane is impermeable remain in separate compartments, and their concentrations remain different on the two sides of the membrane. Molecules to which the membrane is permeable diffuse from one compartment to the other until their concentrations are equal on both sides of the membrane. When the concentrations of the diffusing substance are identical on both sides of the permeable membrane, equilibrium is reached. Individual molecules are still passing through the membrane after equilibrium is established, but equal numbers of molecules are moving in each direction, so there is *no net change* in concentration.

5

Energy, Enzymes, and Metabolism

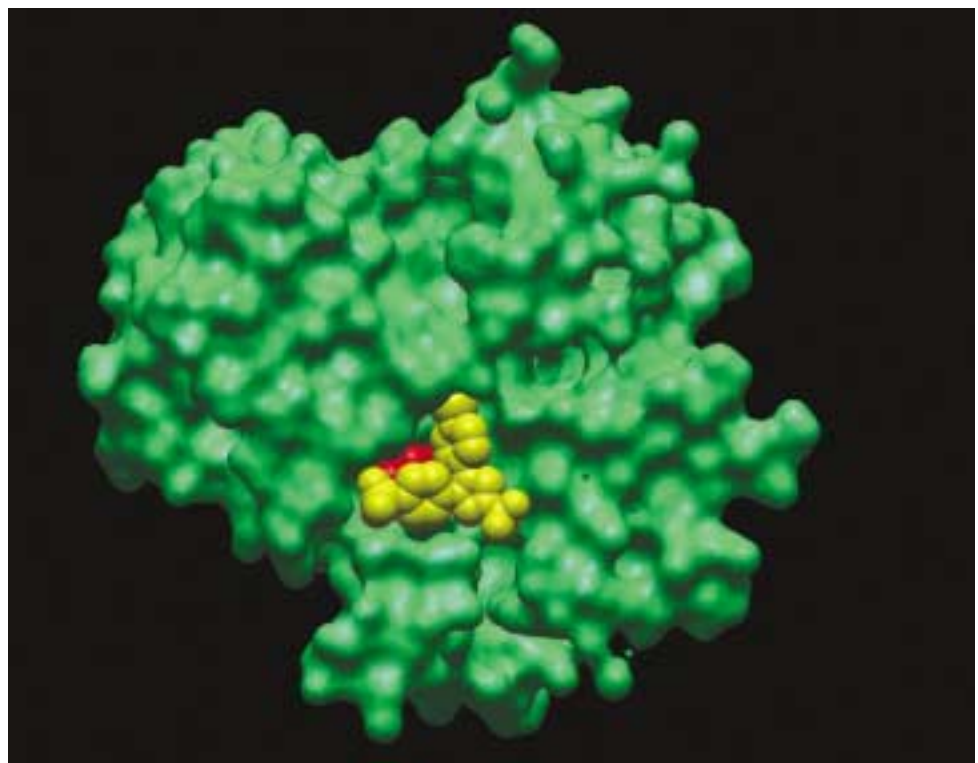


Millions of people, including famous athletes such as hockey star Wayne Gretzky, baseball great Hank Aaron, and Olympic decathlete Bruce Jenner, suffer from crippling arthritis. Until recently, physicians prescribed aspirin to calm the swelling that plagues arthritic joints. Aspirin has a long tradition in medicine. For thousands of years, healers in many cultures knew that the bark of a willow tree had anti-inflammatory properties that reduced swelling and pain. In 1829, German chemists isolated the active ingredient in willow bark, and later in that century others modified it chemically to make an even more effective drug—aspirin. While this drug was effective, it had several negative side effects, such as severe stomach irritation and reduced blood clotting.

It was only when biochemists discovered how aspirin works that its beneficial and undesirable effects could be explained. They discovered that aspirin binds to and adds an acetyl group to a particular amino acid (a serine) in a protein called cyclooxygenase, or COX. The normal role of COX in the body is to act as an enzyme: a catalyst to speed up the conversion of a linear fatty acid to a ring structure. The fatty acid enters a channel in the enzyme macromolecule, where it undergoes a specific chemical conversion, then departs from the enzyme. The ring form of the fatty acid stimulates inflammation in joints, repairs damage in the stomach wall, and helps blood clotting. When the serine residue in COX is acetylated by reactions with aspirin, the enzyme no longer speeds the production of the ring structure. Without COX, the formation of the ring structure still occurs, but at an exceedingly slow rate. So when COX is inhibited by aspirin, inflammation of joints is reduced, but stomach damage and a reduction in blood clotting also occur.

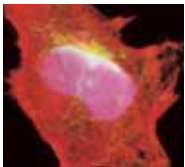
The search went on for a “better” anti-inflammatory drug: one that would block COX only in the joints. During the 1990s, biochemists hit paydirt. They found that there are actually *two* COX enzymes, one that acts in the stomach and blood cells (COX-1) and another that acts in the joints (COX-2). When they determined the primary structures of these two enzymes, they found that only one amino acid differs between them: COX-1 has a bulky

An Inflammatory Enzyme The COX-2 enzyme (green), represented here in a cut-away view, catalyzes the formation of a molecule that stimulates inflammation. Certain drugs (shown in yellow) block the substrate from binding to the active site. Such drugs inhibit the enzyme and depress inflammation. The valine residue that distinguishes COX-2 from the COX-1 enzyme is shown in red.



Part Two • INFORMATION AND HEREDITY

6 Chromosomes, the Cell Cycle, and Cell Division



In 1951, 31-year-old Henrietta Lacks entered Johns Hopkins Hospital to be treated for a cancerous tumor. Although she died a few months later, her tumor cells are still alive today. Scientists found that, given adequate nourishment, cancerous cells from the tumor could reproduce themselves indefinitely in a laboratory dish, where they grew as a formless mass. These “HeLa cells” became a test-tube model for many studies of human biology. Over the past half-century, tens of thousands of research articles have been published using information obtained from Henrietta’s cells. But are these “immortal” cells really a good model for human biology?

In one sense, they are indeed a good model. Most multicellular organisms come from a single cell: the fertilized egg. This cell reproduces itself to make two cells, which in turn divide to become four cells, and so on until all the cells of a new organism have been produced. An organism is not just a mass of cells like the HeLa culture, however; its cells must form specialized tissues and organs, each with specific roles to perform.

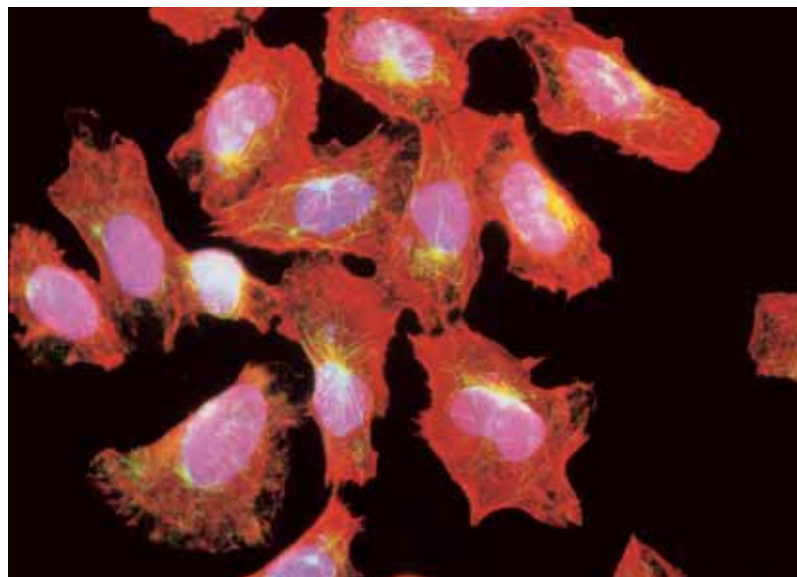
In normal tissues, cell reproduction (cell “births”) is offset by cell loss (cell “deaths”). We know that cell death is important from careful studies of a tiny worm, *Caenorhabditis elegans*, in which 1,090 cells are produced from the fertilized egg and exactly 131 of them die before the worm is born. If the cells that are programmed to die do not do so, the worm’s organs are severely malformed. Another example occurs in the mammalian brain. Young mice, for instance, lose hundreds of thousands of brain cells each day; if these cells do not die, the mouse’s overcrowded brain simply does not work.

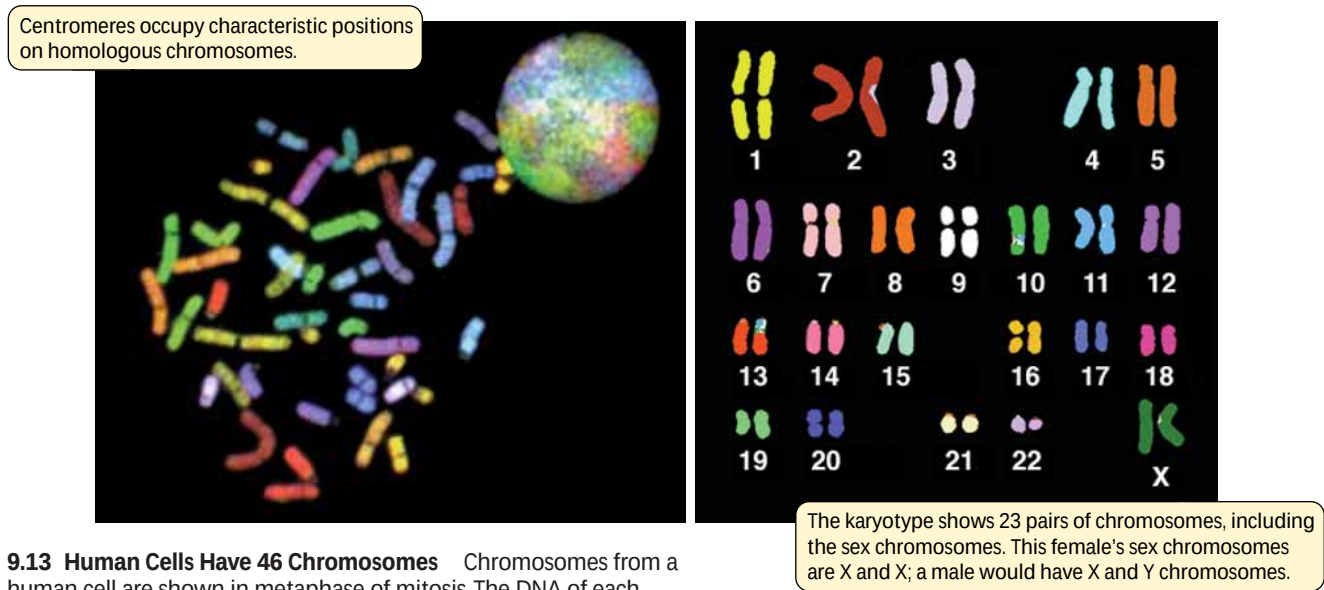
A cell’s death is often programmed into its genetic instructions: normal cells “sacrifice” themselves for the greater good of the organism. Once an organism reaches its adult size, it stays that way through a combination of cell division and programmed cell death. Unlike most normal cells, but like most cancerous cells, HeLa cells keep growing because they have a genetic imbalance that heavily favors cell reproduction over cell death.

In this chapter, we will first describe how prokaryotic cells produce two new organisms from the original single-celled organism. Then we will describe two types of cell and nuclear division—mitosis and meiosis—and relate them to asexual and sexual reproduction in eukaryotic organisms. Finally, to

HeLa Cells: More Births Than Deaths

These tumor cells grow and reproduce as an unspecialized mass on the surface of a solid medium. They have been cultured in a laboratory since 1951. They are the source of much data relating to the reproduction of human cells.





9.13 Human Cells Have 46 Chromosomes Chromosomes from a human cell are shown in metaphase of mitosis. The DNA of each chromosome has a specific nucleotide sequence that is stained by a specific colored dye, so that each homologous pair shares a distinctive color. Each chromosome at this stage is composed of two chromatids, but these cannot be distinguished by this “chromosome painting” technique. The multicolored globe is an interphase nucleus. The karyotype on the right is produced by computerized analysis of the image on the left.

techniques that can capture cells in metaphase and spread out the chromosomes. A photograph of the entire set of chromosomes can then be made, and the images of the individual chromosomes can be placed in an orderly arrangement. Such a rearranged photograph reveals the number, shapes, and sizes of chromosomes in a cell, which together constitute its **karyotype** (Figure 9.13).

Individual chromosomes can be recognized by their lengths, the positions of their centromeres, and characteristic banding when they are stained and observed at high magnification. When the cell is diploid, the karyotype consists of homologous pairs of chromosomes—23 pairs for a total of 46 chromosomes in humans, and greater or smaller numbers of pairs in other diploid species. There is no simple relationship between the size of an organism and its chromosome number (Table 9.1).



Meiosis: A Pair of Nuclear Divisions

Meiosis consists of two nuclear divisions that reduce the number of chromosomes to the haploid number in preparation for sexual reproduction. Although the *nucleus divides twice* during meiosis, the *DNA is replicated only once*. Unlike the products of mitosis, the products of meiosis are different from one another and from the parent cell. To understand the process of meiosis and its specific details, it is useful to keep in mind the overall functions of meiosis:

- ▶ To reduce the chromosome number from diploid to haploid

- ▶ To ensure that each of the haploid products has a complete set of chromosomes
- ▶ To promote genetic diversity among the products

The first meiotic division reduces the chromosome number

Two unique features characterize the first of the two meiotic divisions, **meiosis I**. The first is that homologous chromosomes pair along their entire lengths. No such pairing occurs in mitosis. The second is that after metaphase I, the homologous chromosomes separate. The individual chromosomes, each consisting of two sister chromatids, remain intact until the end of metaphase II in the second meiotic division. Throughout this discussion, refer to Figure 9.14 on the two following pages to help you visualize each step.

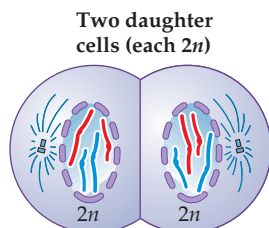
9.1 Numbers of Pairs of Chromosomes in Some Plant and Animal Species

COMMON NAME	SPECIES	NUMBER OF CHROMOSOME PAIRS
Mosquito	<i>Culex pipiens</i>	3
Housefly	<i>Musca domestica</i>	6
Toad	<i>Bufo americanus</i>	11
Rice	<i>Oryza sativa</i>	12
Frog	<i>Rana pipiens</i>	13
Alligator	<i>Alligator mississippiensis</i>	16
Rhesus monkey	<i>Macaca mulatta</i>	21
Wheat	<i>Triticum aestivum</i>	21
Human	<i>Homo sapiens</i>	23
Potato	<i>Solanum tuberosum</i>	24
Donkey	<i>Equus asinus</i>	31
Horse	<i>Equus caballus</i>	32
Dog	<i>Canis familiaris</i>	39
Carp	<i>Cyprinus carpio</i>	52

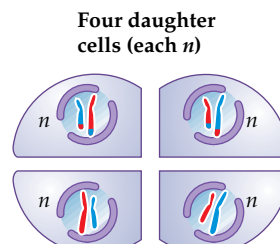
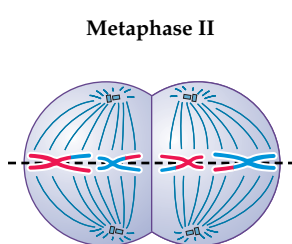
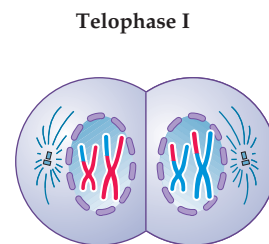
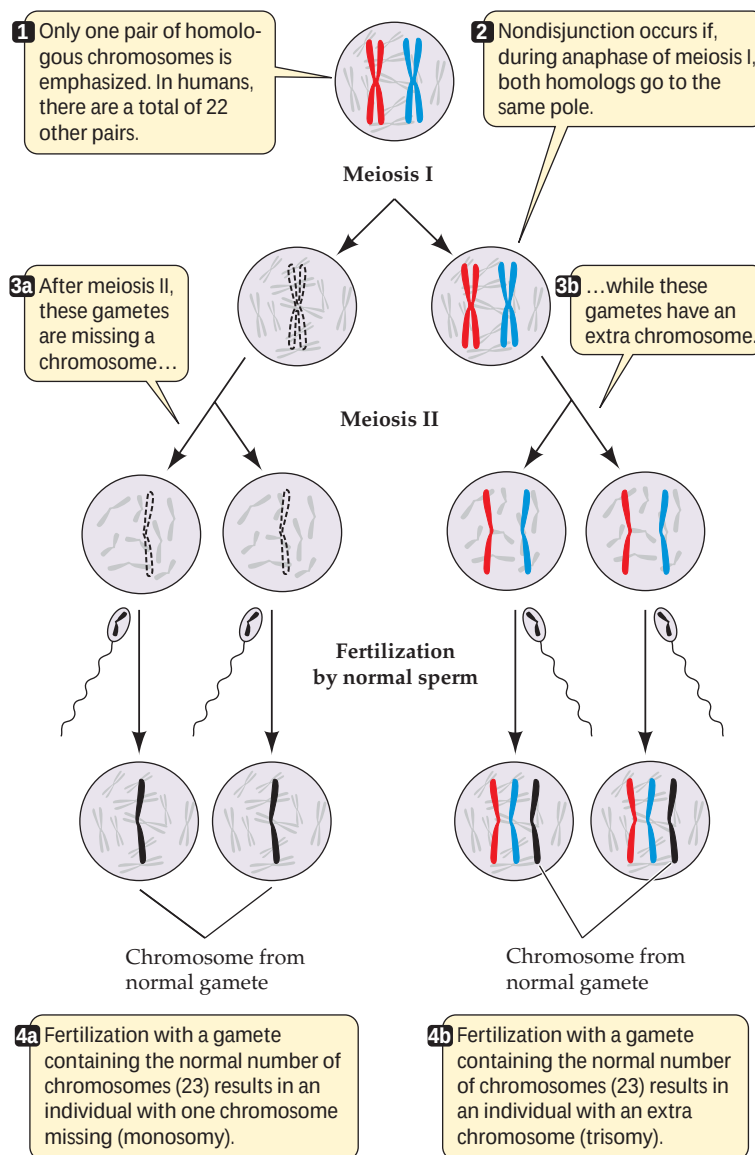
9.18 Nondisjunction Leads to Aneuploidy Nondisjunction occurs if homologous chromosomes fail to separate during meiosis I. The result is aneuploidy: One or more chromosomes are either lacking or present in excess.

break away and become attached to another chromosome. For example, a particular large part of one chromosome 21 may be translocated to another chromosome. Individuals who inherit this translocated piece along with two normal chromosomes 21 will have Down syndrome.

Trisomies (and the corresponding monosomies) are surprisingly common in human zygotes, with 10–30 percent of all conceptions showing aneuploidy. But most of the embryos that develop from such zygotes do not survive to birth, and those that do often die before the age of 1 year. Trisomies and monosomies for most chromosomes other than chromosome 21 are lethal to the embryo. At least one-fifth of all recognized pregnancies are spontaneously terminated during the first 2 months, largely because of such trisomies and monosomies. (The actual proportion of spontaneously terminated pregnancies is certainly higher, because the earliest ones often go unrecognized.)



Mitosis is a mechanism for constancy: The parent nucleus produces two identical daughter nuclei.



Chromatids separate.

Meiosis is a mechanism for diversity: The parent nucleus produces four different haploid daughter nuclei.

7

Genetics: Mendel and Beyond



In the Middle Eastern desert 1,800 years ago, the rabbi faced a dilemma. A Jewish woman had given birth to a son. As required by the laws set down by God's commandment to Abraham almost 2,000 years previously and later reiterated by Moses, the mother brought her 8-day-old son to the rabbi for ritual penile circumcision. The rabbi knew that the woman's two previous sons had bled to death when their foreskins were cut. Yet the biblical commandment remained: Unless he was circumcised, the boy could not be counted among those with whom God had made His solemn covenant. After consultation with other rabbis, it was decided to exempt this, the third son.

Almost a thousand years later, in the twelfth century, the physician and biblical commentator Moses Maimonides reviewed this and numerous other cases in the rabbinical literature and stated that in such instances the third son should not be circumcised. Furthermore, the exemption should apply whether the mother's son was "from her first husband or from her second husband." The bleeding disorder, he reasoned, was clearly carried by the mother and passed on to her sons.

Knowing nothing of our modern vision of genetics, these rabbis linked a human disease (which we now know as hemophilia A) to a pattern of inheritance (which we know as sex linkage). Only in the past few decades have the precise biochemical nature of hemophilia A and its genetic determination been worked out.

How do we account for, and predict, such patterns of inheritance? In this chapter, we will discuss how the units of inheritance, called genes, are transmitted from generation to generation, and we will show how many of the rules that govern genetics can be explained by the behavior of chromosomes during meiosis. We will also describe the interactions of genes with one another and with the environment, and we will examine the consequences of the fact that genes occupy specific positions on chromosomes.

The Foundations of Genetics

Much of the early study of biological inheritance was done with plants and animals of economic importance. Records show that people were deliberately cross-breeding date palm trees and horses as early as 5,000 years ago. By the early nineteenth century, plant breed-

An Ancient Ritual

A male infant undergoes ritual circumcision in accordance with Jewish laws. Sons of Jewish mothers who carry the gene for hemophilia may be exempted from the ritual.



8

DNA and Its Role in Heredity

In the novel and film *Jurassic Park*, fictional scientists were depicted using biotechnology to produce dinosaurs. In the story, the scientists isolated the DNA of dinosaurs from fossilized insects that had sucked the reptiles' blood. The insects, preserved intact in amber (fossilized tree resin), yielded DNA that could be used to produce living individuals of long-extinct organisms such as *Tyrannosaurus rex*. This premise of Michael Crichton's novel was based on an actual scientific paper that claimed to show reptilian DNA sequences in a fossil insect. Unfortunately, the scientific report was not correct; the "preserved" DNA turned out to be a contaminant from modern organisms.

Despite the fact that the preservation of intact DNA over millions of years is highly improbable, the popular success of *Jurassic Park* did bring the idea of DNA as the genetic material to the attention of millions of readers and viewers. Indeed, the double helix of DNA has become a familiar secular icon in the 50 years since that structure was first proposed.

In this and the next several chapters, we will focus on the structure, replication, and function of DNA. As you will see, the structure of DNA determines its function. This chapter will first describe the key experiments that led to the determination that the genetic material is DNA. Then the structure and replication of the molecule will be described. Finally, we will present two practical applications that have arisen from our knowledge of DNA replication: the polymerase chain reaction and DNA sequencing.

Jurassic Park In Michael Crichton's novel, DNA isolated from a fossil is used to produce living dinosaurs. Although such a procedure is strictly fiction, the movie based on the book did bring the role of DNA as the genetic material to the attention of millions of people.

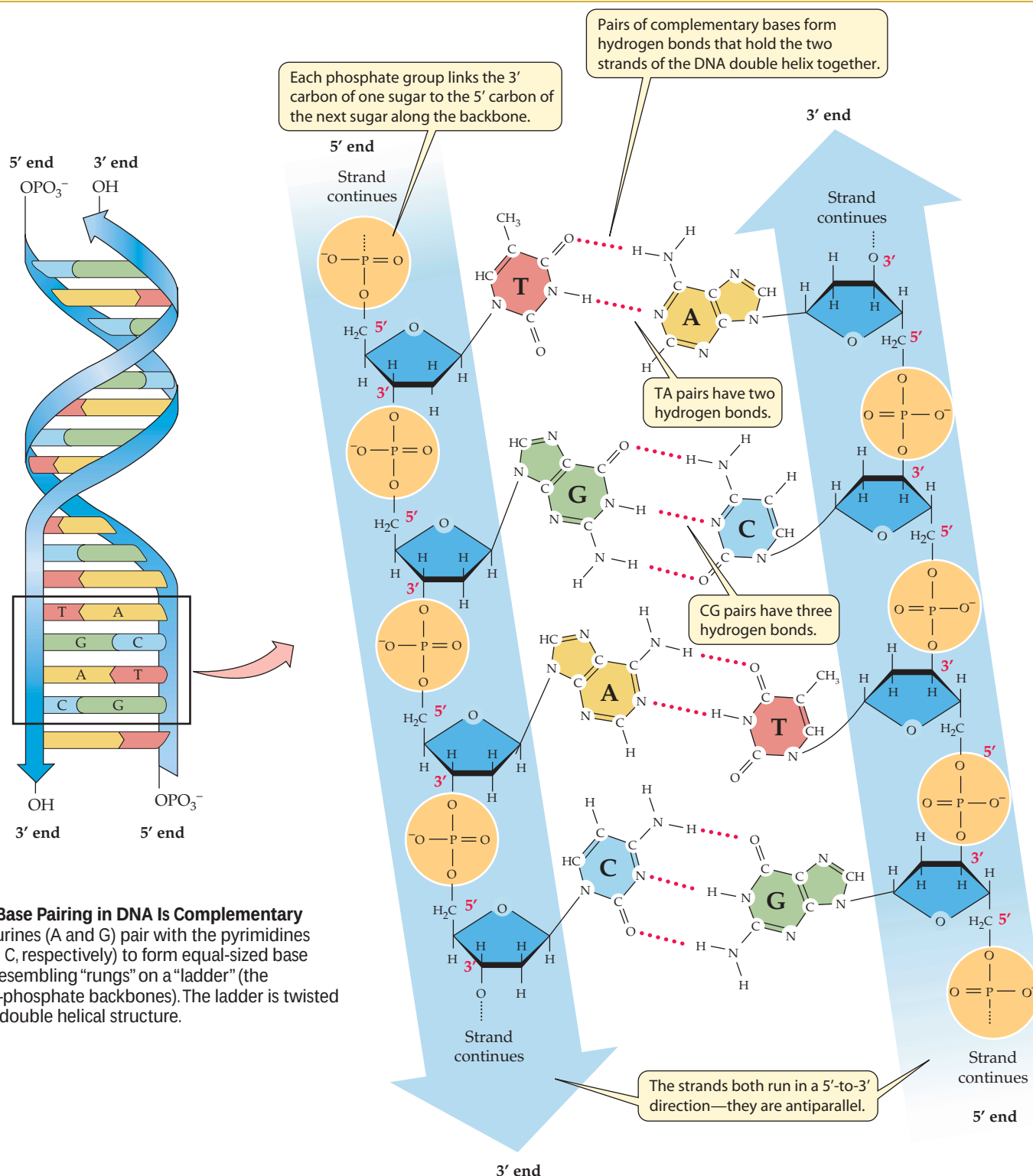
DNA: The Genetic Material

As we saw in Chapter 10, by the early twentieth century geneticists had associated the presence of genes with chromosomes. Research began to focus on exactly which chemical component of chromosomes comprised this genetic material.

By the 1920s, scientists knew that chromosomes were made up of DNA and proteins. At this time a new dye was developed that could bind specifically to DNA and turned red in direct proportion to the amount of DNA present in a cell. This technique provided circumstantial evidence that DNA was the genetic material:

- It was in the right place, since it was an important component of the nucleus and the chromosomes, which were known to carry genes.





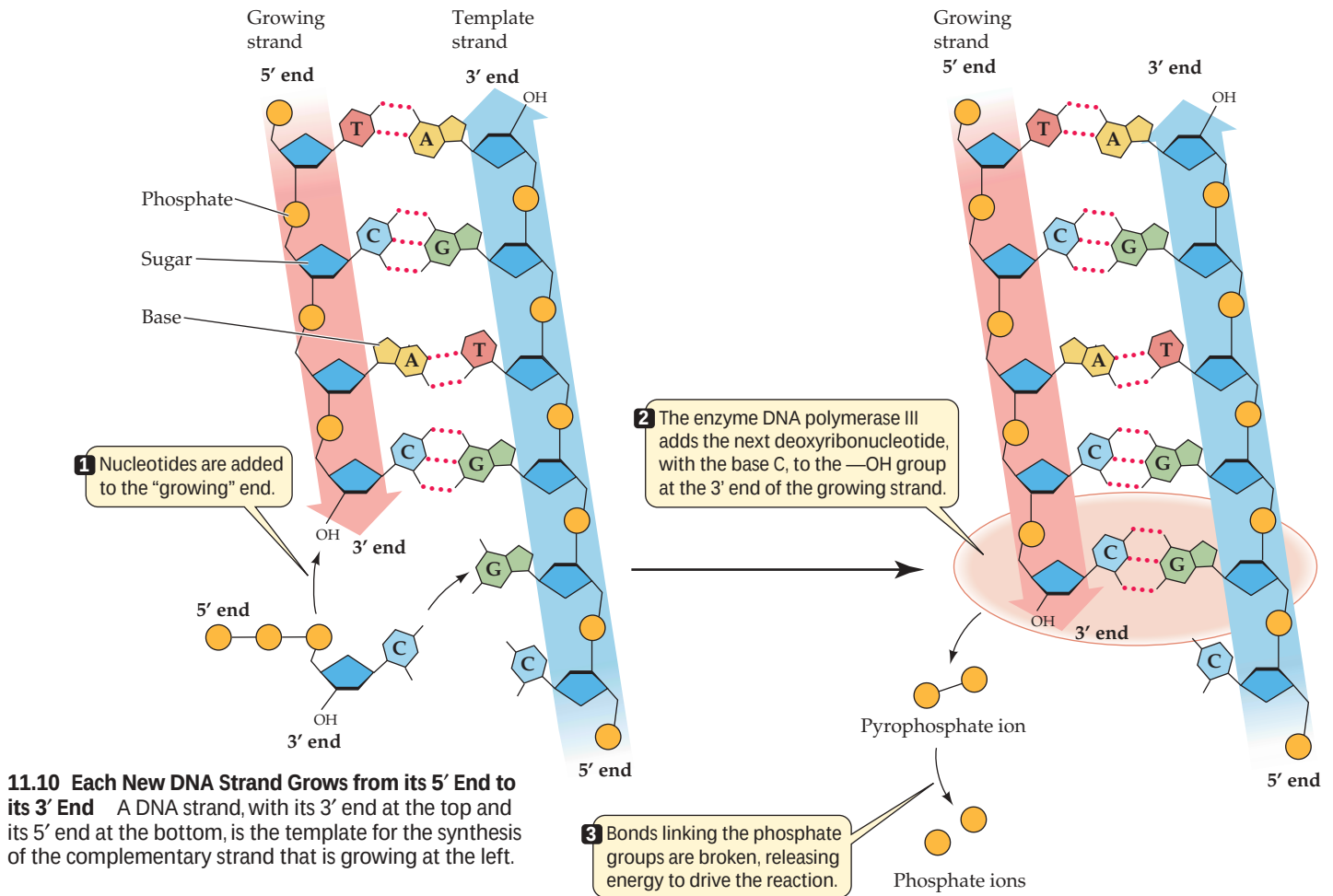
11.7 Base Pairing in DNA Is Complementary

The purines (A and G) pair with the pyrimidines (T and C, respectively) to form equal-sized base pairs resembling "rungs" on a "ladder" (the sugar-phosphate backbones). The ladder is twisted into a double helical structure.

Because the AT and GC pairs are of equal length and fit identically into the double helix (like rungs on a ladder), the diameter of the helix is uniform. The base pairs are flat (see Figure 11.6), and their stacking in the center of the molecule is stabilized by hydrophobic interactions (see Chapter 2), contributing to the overall stability of the double helix.

What does it mean to say that the two DNA strands are *antiparallel*? The direction of a polynucleotide can be defined by looking at the phosphodiester linkages between adjacent

nucleotides (*-diester* refers to the two bonds formed by —OH groups reacting with acidic *phosphate* groups). In the sugar-phosphate backbone of DNA, the phosphate groups connect to the 3' carbon of one deoxyribose molecule and the 5' carbon of the next, linking successive sugars together (see Figure 11.7). The number followed by a prime (') designates the position of a carbon atom in the five-carbon sugar deoxyribose.

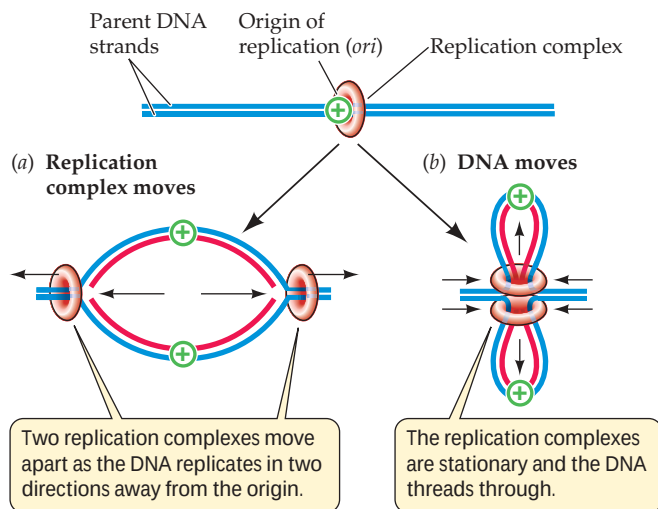


Small, circular DNAs replicate from a single origin

The first event at the origin of replication is the localized unwinding (denaturation) of DNA. There are several forces that hold the two strands together, including hydrogen

bonding and the hydrophobic interactions of bases. An enzyme called **DNA helicase** uses energy from ATP hydrolysis to unwind the DNA, and special proteins called **single-strand binding proteins** bind to the unwound strands to keep them from reassociating into a double helix. This process makes the two template strands available for complementary base pairing.

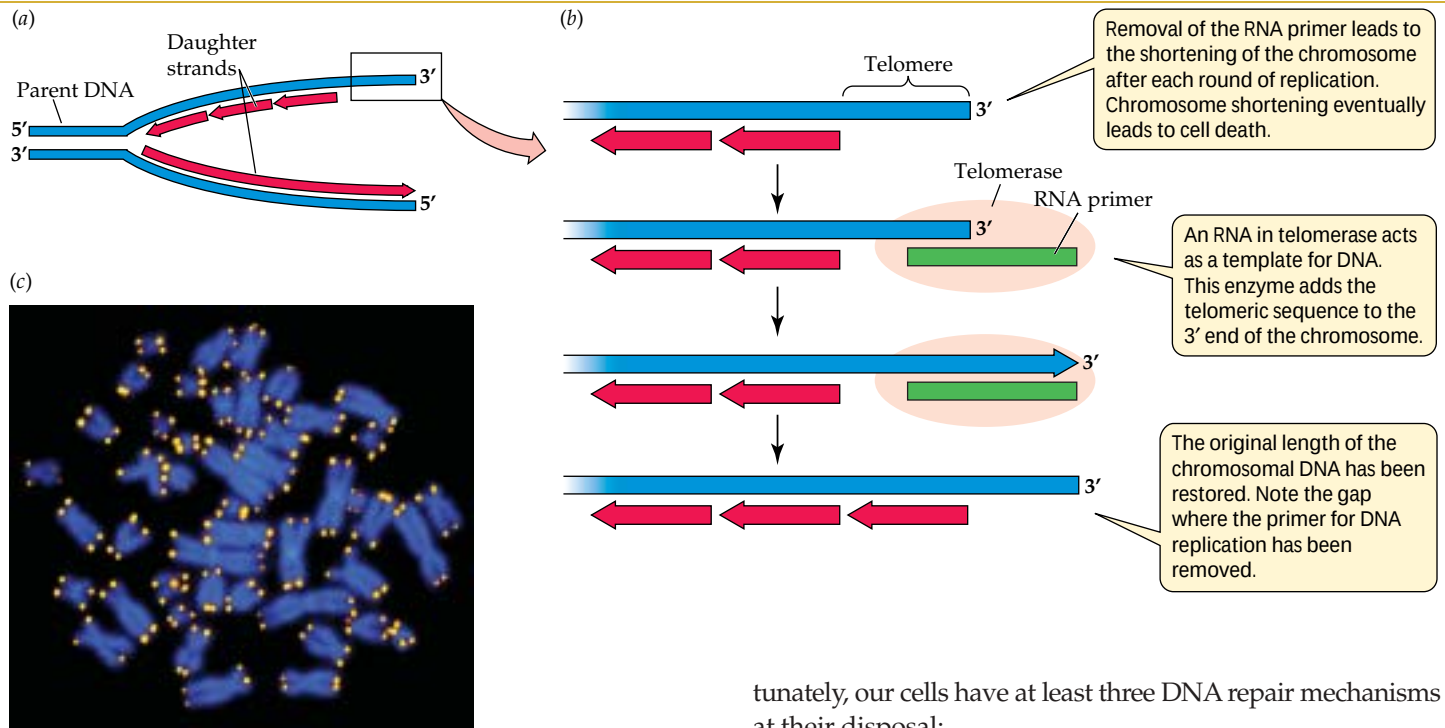
Small circular chromosomes, such as the 3-million-base-pair DNA of bacteria, have a single origin of replication. As the DNA moves through the replication complex, the replication forks grow around the circle (Figure 11.12a). Two interlocking circular DNAs are formed, and they are separated by an enzyme called **DNA topoisomerase**.



11.11 Two Views of DNA Replication (a) It was once thought that the replication complex moved along DNA. (b) Newer evidence suggests that the DNA is threaded through the stationary complex.

Large, linear DNAs have many origins

In large linear chromosomes, such as a human chromosome with 80 million base pairs, there are hundreds of origins of replication. Origins of replication that are adjacent to one another along the linear chromosome can be bound by replication complexes at the same time and replicated simultaneously. So there are many replication forks in eukaryotic DNA (Figure 11.12b).



11.18 Telomeres and Telomerase (a) Removal of RNA primer at the 3' end of the lagging strand leaves a region of DNA unreplicated. (b) The enzyme telomerase binds to the 3' end and extends the lagging strand of DNA. An RNA sequence embedded in telomerase provides a template so that, overall, the DNA does not get shorter. (c) Bright fluorescent staining marks the telomeric regions on these blue-stained human chromosomes.

cer cells to divide continuously. Since most normal cells do not have this ability, telomerase is an attractive target for drugs designed to attack tumors specifically.

There is also interest in telomerase and aging. When a gene expressing high levels of telomerase is added to human cells in culture, their telomeres do not shorten. Instead of dying after 20–30 cell generations, the cells become immortal. It remains to be seen how this finding relates to the aging of a large organism.

DNA Proofreading and Repair

DNA is accurately replicated and faithfully maintained. The price of failure can be great: the transmission of genetic information is at stake, as is the functioning and even the life of a cell or multicellular organism. Yet the replication of DNA is not perfectly accurate, and the DNA of nondividing cells is subject to damage by environmental agents. In the face of these threats, how has life gone on so long?

The preservers of life are DNA repair mechanisms. DNA polymerases initially make a significant number of mistakes in assembling polynucleotide strands. The observed error rate of one for every 10^6 bases replicated would result in about 1,000 mutations every time a human cell divided. For-

tunately, our cells have at least three DNA repair mechanisms at their disposal:

- A **proofreading** mechanism corrects errors in replication as DNA polymerase makes them.
- A **mismatch repair** mechanism scans DNA immediately after it has been replicated and corrects any base-pairing mismatches.
- An **excision repair** mechanism removes abnormal bases that have formed because of chemical damage and replaces them with functional bases.

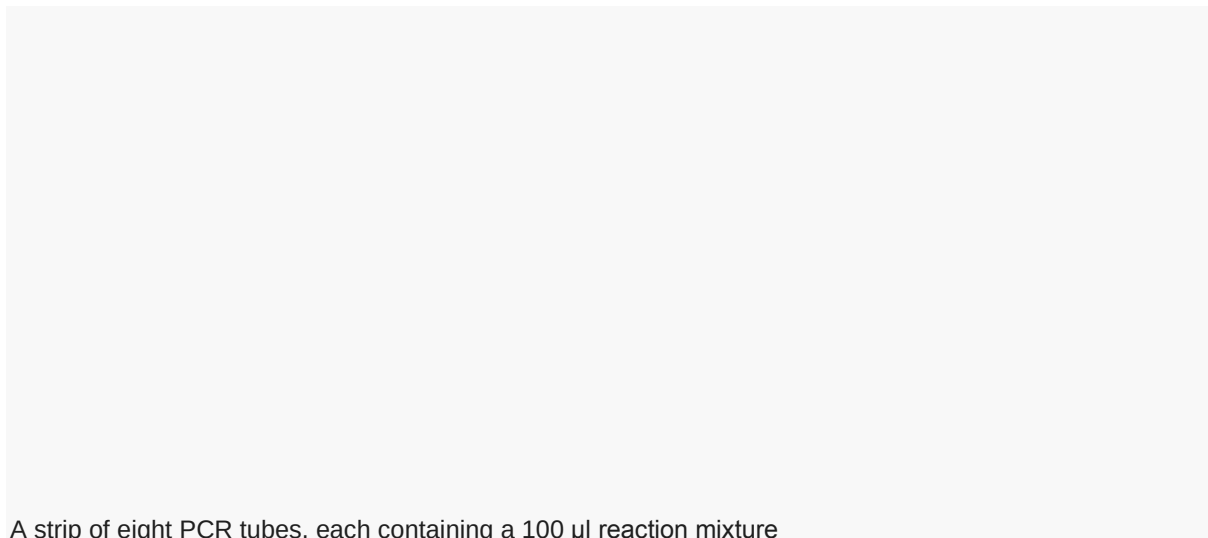
Proofreading mechanisms ensure that DNA replication is accurate

After introducing a new nucleotide into a growing polynucleotide strand, DNA polymerase performs a proofreading function (Figure 11.19a). When a DNA polymerase recognizes a mispairing of bases, it removes the improperly introduced nucleotide and tries again. (Other proteins of the replication complex also play roles in proofreading.) The error rate for this process is only about 1 in 10,000 base pairs, and lowers the overall error rate for replication to about one base in every 10^{10} bases replicated.

Mismatch repair mechanisms correct base-pairing errors

After DNA has been replicated, a second set of proteins surveys the newly replicated molecule and looks for remaining mismatched base pairs (Figure 11.19b). For example, this mismatch repair system might detect an AC base pair instead of an AT pair. But how does the repair system “know” whether the AC pair should be repaired by removing the C and replacing it with T or by removing the A and replacing it with G?

Polymerase chain reaction



A strip of eight PCR tubes, each containing a 100 µl reaction mixture

The **polymerase chain reaction (PCR)** is a technology in [molecular biology](#) used to [amplify](#) a single copy or a few copies of a piece of [DNA](#)



Placing a strip of eight PCR tubes, each containing a 100 μ l reaction mixture, into the [thermal cycler](#)

PCR principles and procedure



Figure 1a: A [thermal cycler](#) for PCR



Figure 1b: An older model three-temperature [thermal cycler](#) for PCR

PCR amplifies a specific region of a DNA strand (the DNA target). Most PCR methods typically amplify DNA fragments of between 0.1 and 10 [kilo base pairs](#) (kbp), although some techniques allow for amplification of fragments up to 40 kbp in size. The amount of amplified product is determined by the available substrates in the reaction, which become limiting as the reaction progresses. A basic PCR set up requires several components and reagents. These components include:

- *DNA template* that contains the DNA region (target) to amplify
- Two [primers](#) that are [complementary](#) to the [3'](#) (three prime) ends of each of the [sense and anti-sense](#) strand of the DNA target

- [Taq polymerase](#) or another [DNA polymerase](#) with a temperature optimum at around 70 °C
- *Deoxynucleoside triphosphates* (dNTPs, sometimes called "deoxynucleotide triphosphates"; [nucleotides](#) containing triphosphate groups), the building-blocks from which the DNA polymerase synthesizes a new DNA strand
- [Buffer solution](#), providing a suitable chemical environment for optimum activity and stability of the DNA polymerase
- [Bivalent cations](#), [magnesium](#) or [manganese](#) ions; generally Mg^{2+} is used, but Mn^{2+} can be used for PCR-mediated DNA [mutagenesis](#), as higher Mn^{2+} concentration increases the error rate during DNA synthesis
- *Monovalent cation* [potassium](#) ions

The PCR is commonly carried out in a reaction volume of 10–200 µl in small reaction tubes (0.2–0.5 ml volumes) in a [thermal cycler](#). The thermal cycler heats and cools the reaction tubes to achieve the temperatures required at each step of the reaction (see below). Many modern thermal cyclers make use of the [Peltier effect](#), which permits both heating and cooling of the block holding the PCR tubes simply by reversing the electric current. Thin-walled reaction tubes permit favorable thermal conductivity to allow for rapid thermal equilibration. Most thermal cyclers have heated lids to prevent condensation at the top of the reaction tube. Older thermocyclers lacking a heated lid require a layer of oil on top of the reaction mixture or a ball of wax inside the tube.

Procedure

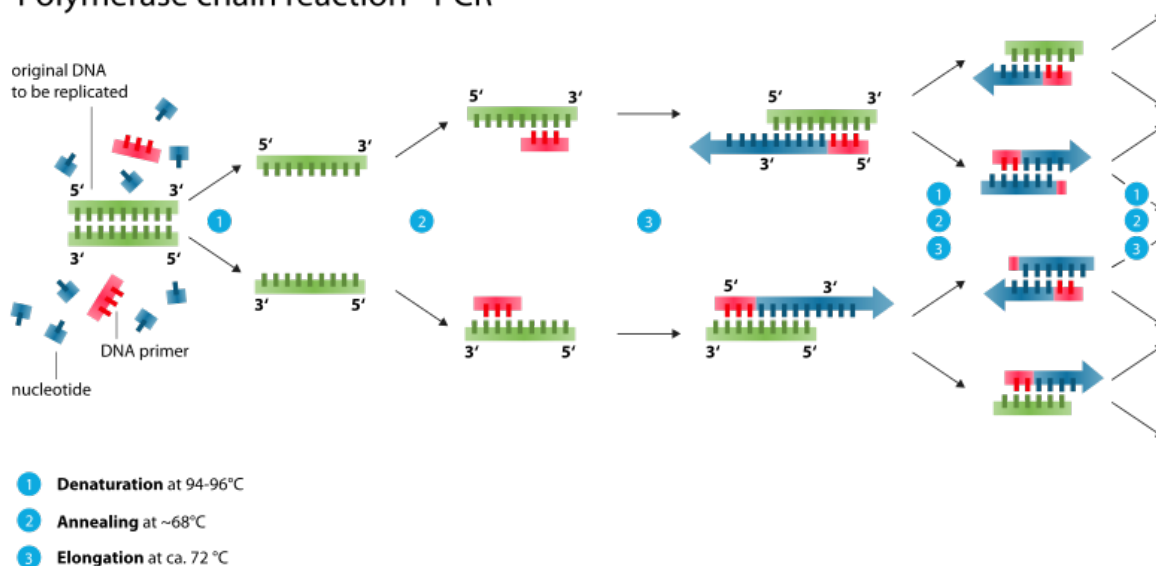
Typically, PCR consists of a series of 20–40 repeated temperature changes, called cycles, with each cycle commonly consisting of 2–3 discrete temperature steps, usually three (Figure below). The cycling is often preceded by a single temperature step at a high temperature (>90 °C), and followed by one hold at the end for final product extension or brief storage. The temperatures used and the length of time they are applied in each cycle depend on a variety of parameters. These include the enzyme used for DNA synthesis, the concentration of divalent ions and dNTPs in the reaction, and the melting temperature ([T_m](#)) of the primers.^[10]

- *Initialization step* (Only required for DNA polymerases that require heat activation by [hot-start PCR](#).^[11]): This step consists of heating the reaction to a temperature of 94–96 °C (or 98 °C if extremely thermostable polymerases are used), which is held for 1–9 minutes.
- *Denaturation step*: This step is the first regular cycling event and consists of heating the reaction to 94–98 °C for 20–30 seconds. It causes [DNA melting](#) of the DNA template by disrupting the hydrogen bonds between complementary bases, yielding single-stranded DNA molecules.
- *Annealing step*: The reaction temperature is lowered to 50–65 °C for 20–40 seconds allowing annealing of the primers to the single-stranded DNA template. This temperature must be low enough to allow for [hybridization](#) of the primer to the strand, but high enough for the hybridization to be specific, i.e., the primer should only bind to a perfectly complementary part of the template. If the temperature is too low, the primer could bind imperfectly. If it is too high, the primer might not bind. Typically the annealing temperature is about 3–5 °C below the T_m of the primers used. Stable DNA–DNA hydrogen bonds are only formed when the

primer sequence very closely matches the template sequence. The polymerase binds to the primer-template hybrid and begins DNA formation.

- *Extension/elongation step:* The temperature at this step depends on the DNA polymerase used; [Taq polymerase](#) has its optimum [activity](#) temperature at 75–80 °C, and commonly a temperature of 72 °C is used with this enzyme. At this step the DNA polymerase synthesizes a new DNA strand complementary to the DNA template strand by adding dNTPs that are complementary to the template in 5' to 3' direction, condensing the 5'-[phosphate group](#) of the dNTPs with the 3'-[hydroxyl group](#) at the end of the nascent (extending) DNA strand. The extension time depends both on the DNA polymerase used and on the length of the DNA fragment to amplify. As a rule-of-thumb, at its optimum temperature, the DNA polymerase polymerizes a thousand bases per minute. Under optimum conditions, i.e., if there are no limitations due to limiting substrates or reagents, at each extension step, the amount of DNA target is doubled, leading to exponential (geometric) amplification of the specific DNA fragment.
- *Final elongation:* This single step is occasionally performed at a temperature of 70–74 °C (this is the temperature needed for optimal activity for most polymerases used in PCR) for 5–15 minutes after the last PCR cycle to ensure that any remaining single-stranded DNA is fully extended.
- *Final hold:* This step at 4–15 °C for an indefinite time may be employed for short-term storage of the reaction.

Polymerase chain reaction - PCR



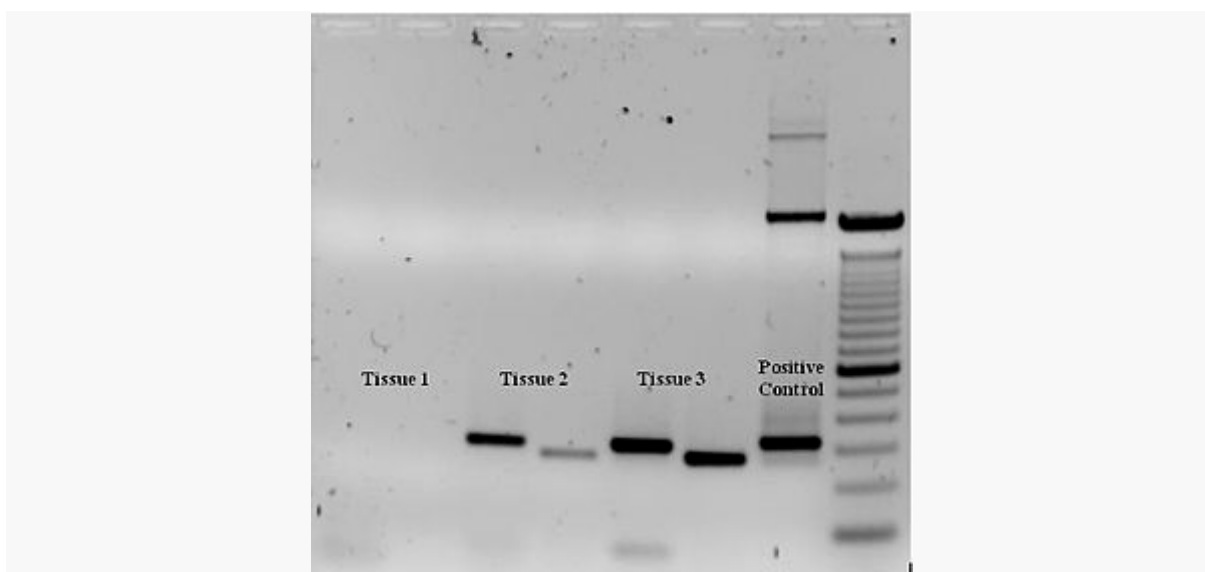


Figure 3: Ethidium bromide-stained PCR products after [gel electrophoresis](#). Two sets of primers were used to amplify a target sequence from three different tissue samples. No amplification is present in sample #1; DNA bands in sample #2 and #3 indicate successful amplification of the target sequence. The gel also shows a positive control, and a DNA ladder containing DNA fragments of defined length for sizing the bands in the experimental PCRs.

To check whether the PCR generated the anticipated DNA fragment (also sometimes referred to as the amplicon or [amplicon](#)), [agarose gel electrophoresis](#) is employed for size separation of the PCR products. The size(s) of PCR products is determined by comparison with a [DNA ladder](#) (a molecular weight marker), which contains DNA fragments of known size, run on the gel alongside the PCR products (see Fig. 3).

PCR stages

The PCR process can be divided into three stages:

Exponential amplification: At every cycle, the amount of product is doubled (assuming 100% reaction efficiency). The reaction is very sensitive: only minute quantities of DNA must be present.^[14]

Leveling off stage: The reaction slows as the DNA polymerase loses activity and as consumption of reagents such as dNTPs and primers causes them to become limiting.

Plateau: No more product accumulates due to exhaustion of reagents and enzyme.

PCR optimization

Main article: [PCR optimization](#)

In practice, PCR can fail for various reasons, in part due to its sensitivity to contamination causing amplification of spurious DNA products. Because of this, a number of techniques and procedures have been developed for optimizing PCR conditions.^{[15][16]} Contamination with extraneous DNA is addressed with lab protocols and procedures that separate pre-PCR mixtures from potential DNA contaminants.^[9] This usually involves spatial separation of PCR-setup areas from areas for analysis or purification of PCR products, use of disposable plasticware, and thoroughly cleaning the work surface between reaction setups. Primer-design techniques are important in improving PCR product yield and in avoiding the formation of spurious products, and the usage of alternate buffer components or polymerase enzymes can help with amplification of long or otherwise problematic regions of DNA. Addition of reagents, such as [formamide](#), in buffer systems may increase the specificity and yield of PCR. Computer simulations of theoretical PCR results ([Electronic PCR](#)) may be performed to assist in primer design.

Application of PCR

Selective DNA isolation

PCR allows isolation of DNA fragments from genomic DNA by selective amplification of a specific region of DNA. This use of PCR augments many methods, such as generating [hybridization probes](#) for [Southern](#) or [northern](#) hybridization and [DNA cloning](#), which require larger amounts of DNA, representing a specific DNA region. PCR supplies these techniques with high amounts of pure DNA, enabling analysis of DNA samples even from very small amounts of starting material.

Other applications of PCR include [DNA sequencing](#) to determine unknown PCR-amplified sequences in which one of the amplification primers may be used in Sanger sequencing, isolation of a DNA sequence to expedite recombinant DNA technologies involving the insertion of a DNA sequence into a [plasmid](#), [phage](#), or [cosmid](#) (depending on size) or the genetic material of another organism. Bacterial colonies (*E. coli*) can be rapidly screened by PCR for correct DNA [vector](#) constructs.^[19] PCR may also be used for [genetic fingerprinting](#); a forensic technique used to identify a person or organism by comparing experimental DNAs through different PCR-based methods.

Some PCR 'fingerprints' methods have high discriminative power and can be used to identify genetic relationships between individuals, such as parent-child or between siblings, and are used in paternity testing (Fig. 4). This technique may also be used to determine evolutionary relationships among organisms when certain molecular clocks are used (i.e., the [16S rRNA](#) and *recA* genes of microorganisms).^[citation needed]

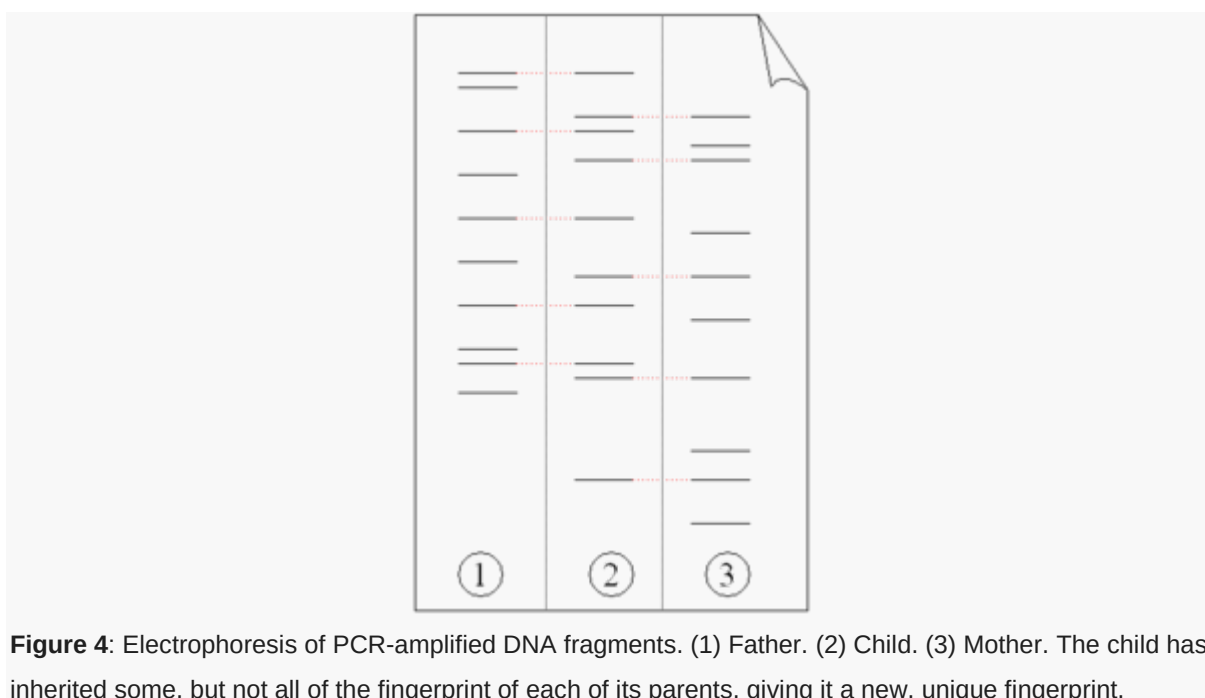


Figure 4: Electrophoresis of PCR-amplified DNA fragments. (1) Father. (2) Child. (3) Mother. The child has inherited some, but not all of the fingerprint of each of its parents, giving it a new, unique fingerprint.

Amplification and quantification of DNA

Because PCR amplifies the regions of DNA that it targets, PCR can be used to analyze extremely small amounts of sample. This is often critical for [forensic analysis](#), when only a trace amount of DNA is available as evidence. PCR may also be used in the analysis of [ADNA](#) that is tens of thousands of years old. These PCR-based techniques have been successfully used on animals, such as a forty-thousand-year-old [mammoth](#), and also on human DNA, in applications ranging from the analysis of Egyptian [mummies](#) to the identification of a [Russian tsar](#) and the body of English king [Richard III](#).

Quantitative PCR methods allow the estimation of the amount of a given sequence present in a sample—a technique often applied to quantitatively determine levels of [gene](#)

[expression](#). [Quantitative PCR](#) is an established tool for DNA quantification that measures the accumulation of DNA product after each round of PCR amplification.

PCR in diagnosis of diseases

PCR permits early diagnosis of [malignant](#) diseases such as [leukemia](#) and [lymphomas](#), which is currently the highest-developed in cancer research and is already being used routinely. PCR assays can be performed directly on genomic DNA samples to detect translocation-specific malignant cells at a sensitivity that is at least 10,000 fold higher than that of other methods.

PCR allows for rapid and highly specific diagnosis of infectious diseases, including those caused by bacteria or viruses. PCR also permits identification of non-cultivable or slow-growing microorganisms such as [mycobacteria](#), [anaerobic bacteria](#), or [viruses](#) from [tissue culture](#) assays and [animal models](#). The basis for PCR diagnostic applications in microbiology is the detection of infectious agents and the discrimination of non-pathogenic from pathogenic strains by virtue of specific genes.

Viral DNA can likewise be detected by PCR. The primers used must be specific to the targeted sequences in the DNA of a virus, and the PCR can be used for diagnostic analyses or DNA sequencing of the viral genome. The high sensitivity of PCR permits virus detection soon after infection and even before the onset of disease. Such early detection may give physicians a significant lead time in treatment. The amount of virus ("[viral load](#)") in a patient can also be quantified by PCR-based DNA quantitation techniques (see below).

Limitations

DNA polymerase is prone to error, which in turn causes mutations in the PCR fragments that are made. Additionally, the specificity of the PCR fragments can mutate to the template DNA, due to nonspecific binding of primers. Furthermore prior information on the sequence is necessary in order to generate the primers. ^[23]

Variations on the basic PCR technique

[Allele-specific PCR](#): a diagnostic or cloning technique based on single-nucleotide variations (SNVs not to be confused with [SNPs](#)) (single-base differences in a patient). It requires prior knowledge of a DNA sequence, including differences between [alleles](#), and uses primers whose 3' ends encompass the SNV (base pair buffer around SNV usually incorporated). PCR amplification under stringent conditions is much less efficient in the presence of a mismatch between template and primer, so successful amplification with an SNP-specific primer signals presence of the specific SNP in a sequence. See [SNP genotyping](#) for more information.

- [Assembly PCR](#) or *Polymerase Cycling Assembly (PCA)*: artificial synthesis of long DNA sequences by performing PCR on a pool of long oligonucleotides with short overlapping segments. The oligonucleotides alternate between sense and antisense directions, and the overlapping segments determine the order of the PCR fragments, thereby selectively producing the final long DNA product.
- [Asymmetric PCR](#): preferentially amplifies one DNA strand in a double-stranded DNA template. It is used in [sequencing](#) and hybridization probing where amplification of only one of the two complementary strands is required. PCR is carried out as usual, but with a great excess of the primer for the strand targeted for amplification. Because of the slow ([arithmetic](#)) amplification later in the reaction after the limiting primer has been used up, extra cycles of PCR are required. A recent modification on this process, known as *Linear-After-The-Exponential-PCR (LATE-PCR)*, uses a limiting primer with a higher melting temperature

(T_m) than the excess primer to maintain reaction efficiency as the limiting primer concentration decreases mid-reaction.

- **Dial-out PCR**: a highly parallel method for retrieving accurate DNA molecules for gene synthesis. A complex library of DNA molecules is modified with unique flanking tags before massively parallel sequencing. Tag-directed primers then enable the retrieval of molecules with desired sequences by PCR.^[28]
- **Digital PCR (dPCR)**: used to measure the quantity of a target DNA sequence in a DNA sample. The DNA sample is highly diluted so that after running many PCRs in parallel, some of them do not receive a single molecule of the target DNA. The target DNA concentration is calculated using the proportion of negative outcomes. Hence the name 'digital PCR'.
- **Helicase-dependent amplification**: similar to traditional PCR, but uses a constant temperature rather than cycling through denaturation and annealing/extension cycles. **DNA helicase**, an enzyme that unwinds DNA, is used in place of thermal denaturation.^[29]
- **Hot start PCR**: a technique that reduces non-specific amplification during the initial set up stages of the PCR. It may be performed manually by heating the reaction components to the denaturation temperature (e.g., 95 °C) before adding the polymerase.^[30] Specialized enzyme systems have been developed that inhibit the polymerase's activity at ambient temperature, either by the binding of an **antibody** or by the presence of covalently bound inhibitors that dissociate only after a high-temperature activation step. Hot-start/cold-finish PCR is achieved with new hybrid polymerases that are inactive at ambient temperature and are instantly activated at elongation temperature.
- **In silico PCR** (digital PCR, virtual PCR, electronic PCR, e-PCR) refers to computational tools used to calculate theoretical polymerase chain reaction results using a given set of **primers (probes)** to amplify **DNA** sequences from a sequenced **genome** or **transcriptome**. In silico PCR was proposed as an educational tool for molecular biology.
- **Intersequence-specific PCR (ISSR)**: a PCR method for DNA fingerprinting that amplifies regions between simple sequence repeats to produce a unique fingerprint of amplified fragment lengths.
- **Inverse PCR**: is commonly used to identify the flanking sequences around **genomic** inserts. It involves a series of **DNA digestions** and **self ligation**, resulting in known sequences at either end of the unknown sequence.
- **Ligation-mediated PCR**: uses small DNA linkers ligated to the DNA of interest and multiple primers annealing to the DNA linkers; it has been used for **DNA sequencing, genome walking**, and **DNA footprinting**.
- **Methylation-specific PCR (MSP)**: developed by Stephen Baylin and Jim Herman at the Johns Hopkins School of Medicine, and is used to detect methylation of CpG islands in genomic DNA. DNA is first treated with sodium bisulfite, which converts unmethylated cytosine bases to uracil, which is recognized by PCR primers as thymine. Two PCRs are then carried out on

the modified DNA, using primer sets identical except at any CpG islands within the primer sequences. At these points, one primer set recognizes DNA with cytosines to amplify methylated DNA, and one set recognizes DNA with uracil or thymine to amplify unmethylated DNA. MSP using qPCR can also be performed to obtain quantitative rather than qualitative information about methylation.

- **Miniprimer PCR**: uses a thermostable polymerase (S-Tbr) that can extend from short primers ("smalligos") as short as 9 or 10 nucleotides. This method permits PCR targeting to smaller primer binding regions, and is used to amplify conserved DNA sequences, such as the 16S (or eukaryotic 18S) rRNA gene.^[37]
- **Multiplex Ligation-dependent Probe Amplification (MLPA)**: permits amplifying multiple targets with a single primer pair, thus avoiding the resolution limitations of multiplex PCR (see below).
- **Multiplex-PCR**: consists of multiple primer sets within a single PCR mixture to produce amplicons of varying sizes that are specific to different DNA sequences. By targeting multiple genes at once, additional information may be gained from a single test-run that otherwise would require several times the reagents and more time to perform. Annealing temperatures for each of the primer sets must be optimized to work correctly within a single reaction, and amplicon sizes. That is, their base pair length should be different enough to form distinct bands when visualized by gel electrophoresis.
- **Nanoparticle-Assisted PCR (nanoPCR)**: In recent years, it has been reported that some nanoparticles (NPs) can enhance the efficiency of PCR (thus being called nanoPCR), and some even perform better than the original PCR enhancers. It was also found that quantum dots (QDs) can improve PCR specificity and efficiency. Single-walled carbon nanotubes (SWCNTs) and multi-walled carbon nanotubes (MWCNTs) are efficient in enhancing the amplification of long PCR. Carbon nanopowder (CNP) was reported be able to improve the efficiency of repeated PCR and long PCR. ZnO, TiO₂, and Ag NPs were also found to increase PCR yield. Importantly, already known data has indicated that non-metallic NPs retained acceptable amplification fidelity. Given that many NPs are capable of enhancing PCR efficiency, it is clear that there is likely to be great potential for nanoPCR technology improvements and product development.
- **Nested PCR**: increases the specificity of DNA amplification, by reducing background due to non-specific amplification of DNA. Two sets of primers are used in two successive PCRs. In the first reaction, one pair of primers is used to generate DNA products, which besides the intended target, may still consist of non-specifically amplified DNA fragments. The product(s) are then used in a second PCR with a set of primers whose binding sites are completely or partially different from and located 3' of each of the primers used in the first reaction. Nested PCR is often more successful in specifically amplifying long DNA fragments than conventional PCR, but it requires more detailed knowledge of the target sequences.

- [Overlap-extension PCR](#) or *Splicing by overlap extension (SOEing)* : a [genetic engineering](#) technique that is used to splice together two or more DNA fragments that contain complementary sequences. It is used to join DNA pieces containing genes, regulatory sequences, or mutations; the technique enables creation of specific and long DNA constructs. It can also introduce deletions, insertions or point mutations into a DNA sequence.
- [PAN-AC](#): uses isothermal conditions for amplification, and may be used in living cells.
- [quantitative PCR](#) (qPCR): used to measure the quantity of a target sequence (commonly in real-time). It quantitatively measures starting amounts of DNA, cDNA, or RNA. [quantitative PCR](#) is commonly used to determine whether a DNA sequence is present in a sample and the number of its copies in the sample. *Quantitative PCR* has a very high degree of precision. Quantitative PCR methods use fluorescent dyes, such as Sybr Green, EvaGreen or [fluorophore](#)-containing DNA probes, such as [TaqMan](#), to measure the amount of amplified product in real time. It is also sometimes abbreviated to [RT-PCR](#) (*real-time PCR*) but this abbreviation should be used only for [reverse transcription PCR](#). qPCR is the appropriate contractions for [quantitative PCR](#) (real-time PCR).
- *Reverse Transcription PCR (RT-PCR)*: for amplifying DNA from RNA. [Reverse transcriptase](#) reverse transcribes [RNA](#) into [cDNA](#), which is then amplified by PCR. RT-PCR is widely used in [expression profiling](#), to determine the expression of a gene or to identify the sequence of an RNA transcript, including transcription start and termination sites. If the genomic DNA sequence of a gene is known, RT-PCR can be used to map the location of [exons](#) and [introns](#) in the gene. The 5' end of a gene (corresponding to the transcription start site) is typically identified by [RACE-PCR](#) (*Rapid Amplification of cDNA Ends*).
- [Solid Phase PCR](#): encompasses multiple meanings, including [Polony Amplification](#) (where PCR colonies are derived in a gel matrix, for example), Bridge PCR^[44] (primers are covalently linked to a solid-support surface), conventional Solid Phase PCR (where Asymmetric PCR is applied in the presence of solid support bearing primer with sequence matching one of the aqueous primers) and Enhanced Solid Phase PCR (where conventional Solid Phase PCR can be improved by employing high T_m and nested solid support primer with optional application of a thermal 'step' to favour solid support priming).
- *Suicide PCR*: typically used in [paleogenetics](#) or other studies where avoiding false positives and ensuring the specificity of the amplified fragment is the highest priority. It was originally described in a study to verify the presence of the microbe [Yersinia pestis](#) in dental samples obtained from 14th Century graves of people supposedly killed by plague during the medieval [Black Death](#) epidemic. The method prescribes the use of any primer combination only once in a PCR (hence the term "suicide"), which should never have been used in any positive control PCR reaction, and the primers should always target a genomic region never

amplified before in the lab using this or any other set of primers. This ensures that no contaminating DNA from previous PCR reactions is present in the lab, which could otherwise generate false positives.

- *Thermal asymmetric interlaced PCR (TAIL-PCR)*: for isolation of an unknown sequence flanking a known sequence. Within the known sequence, TAIL-PCR uses a nested pair of primers with differing annealing temperatures; a degenerate primer is used to amplify in the other direction from the unknown sequence.
- *Touchdown PCR (Step-down PCR)*: a variant of PCR that aims to reduce nonspecific background by gradually lowering the annealing temperature as PCR cycling progresses. The annealing temperature at the initial cycles is usually a few degrees (3-5 °C) above the T_m of the primers used, while at the later cycles, it is a few degrees (3-5 °C) below the primer T_m . The higher temperatures give greater specificity for primer binding, and the lower temperatures permit more efficient amplification from the specific products formed during the initial cycles.
- *Universal Fast Walking*: for genome walking and genetic fingerprinting using a more specific 'two-sided' PCR than conventional 'one-sided' approaches (using only one gene-specific primer and one general primer—which can lead to artefactual 'noise') by virtue of a mechanism involving lariat structure formation. Streamlined derivatives of UFW are LaNe RAGE (lariat-dependent nested PCR for rapid amplification of genomic DNA ends), 5'RACE LaNe and 3'RACE LaNe.

9

From DNA to Protein: Genotype to Phenotype



In 1978, Georgi Markov, a journalist who was living in London because he had written articles critical of the then-Communist government of Bulgaria, was standing at a bus stop near Waterloo Station. A man, possibly a Bulgarian secret agent, brushed up against him and, seemingly by accident, poked him with an umbrella. Markov felt a sharp pain, and within a few hours, he started to feel weak. High temperature, vomiting, and more severe symptoms soon followed. Two days later he was dead.

Police investigators found a tiny perforated pellet embedded in Markov's leg, and in that pellet was a small amount of ricin, a highly toxic molecule isolated from the seeds of the tropical castor bean plant, *Ricinus communis*. These seeds have been used for centuries as a source of castor oil, a natural product that used to be commonly given to children to "clean out" the digestive tract and is now used in the plastics industry. The toxin is a protein that is not present in the oil, and people found out the hard way that it is one of the most poisonous substances made by any organism.

Ricin kills cells by blocking protein synthesis. More specifically, it catalyzes the modification and cleavage of one of the large RNA molecules that make up the eukaryotic ribosome, the "workbench" of protein synthesis. Proteins are the major phenotypic expression of the genotype—the genetic information encoded in a cell's DNA. Ricin inhibits the cell's ability to express the genotype as phenotype through protein synthesis, and therefore ricin-poisoned cells cannot survive.

This chapter deals with the mechanisms by which genes are expressed as proteins. We will begin with evidence for the relationship between genes and proteins, and then fill in some of the details of the processes of transcription—the copying of the gene sequence of DNA into a sequence of RNA—and translation—the use of the sequence of RNA to make a polypeptide with a defined order of amino acids. Finally, we will define mutations and their phenotypes in specific molecular terms.

One Gene, One Polypeptide

There are many steps between genotype and phenotype. Genes cannot, all by themselves, directly produce a phenotypic result, such as a particular eye color, a specific seed shape, or a cleft chin, any more than a compact disk can play a symphony without the help of a CD player.

***Ricinus communis*, the Castor Bean Plant**

This brightly colored plant, grown in the Tropics as an ornamental, produces ricin, a lethal toxin that inhibits protein synthesis at the ribosome.



10

The Genetics of Viruses and Prokaryotes



Robert Stevens felt sicker and sicker, until he finally went to the emergency room of a hospital near his home in Boca Raton, Florida. The medical staff noted his fever, vomiting, and headache and tested his spinal fluid for infectious agents. They saw a few spores of *Bacillus* bacteria, which they might have dismissed as contamination if some of the hospital staff had not just taken a course on identifying possible germ warfare agents at the U.S. Centers for Disease Control and Prevention.

The spores were put in a culture dish with a growth medium, and colonies of bacteria soon appeared. They were identified as *Bacillus anthracis*—the anthrax bacterium. Stevens had respiratory anthrax, a rare form of the disease that he apparently picked up when he inhaled spores deliberately placed in an envelope sent to the newspaper where he worked. The doctors gave him antibiotics to stem the growth of the infection, but it was too late. The rapidly dividing bacteria produced toxins that overwhelmed his body's defenses. Three days later, on October 5, 2001, Robert Stevens died.

When its genome was sequenced, the killer bacterium was found to belong to a strain of *B. anthracis* that had been used in the U.S. government's biological weapons research program until it was disbanded by international agreement in 1969.

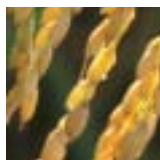
Behind the fears of bioterrorism that surrounded the anthrax infections of Stevens and others lie many aspects of prokaryotic genetics and molecular biology. In this chapter, we will describe some of the science behind the headlines, looking at such aspects as bacterial growth and colony formation, exchanges of genetic material, and genome sequencing. Prokaryotes usually reproduce asexually by cell division, but they can acquire new genes in several ways. These mechanisms range from simple recombination in a sexual process to the transport of genes by infective viruses. We will also describe how the expression of prokaryotic genes is regulated and what DNA sequencing has revealed about the prokaryotic genome.

Viruses are not prokaryotes. In fact, they are not even cells, but intracellular parasites that can reproduce only within living cells. We will begin this chapter with a look at the benefits of studying prokaryotes and viruses. Then we will examine the structure, classification, reproduction, and genetics of viruses.

A Prokaryote Weapon This composite micrograph shows spores of *Bacillus anthracis* (yellow) in human lung tissue. *B. anthracis* is the cause of anthrax, a disease that can be fatal to many mammals, including humans. Because anthrax is transmitted by hardy spores that can survive long periods exposed to the environment, it is a prime weapon for bioterrorism.



11 *The Eukaryotic Genome and Its Expression*



“The most precious things are not jade or pearls, but the five grains.” This ancient Chinese saying refers to rice, wheat, maize (corn), sorghum, and millet. Today the saying remains as true as ever, since these crops provide two-thirds of the human diet worldwide. With the recent publication of the genome sequences of the two major cultivated varieties of rice, agricultural scientists are well on the way to dramatic improvements in the nutritional quality and yield of the grain produced in the paddies of Asia. And the rice genome turns out to be a smaller version of the much larger genomes of the other four grains.

Like other eukaryotes, rice has much more DNA than a typical prokaryote—some 430 million base pairs. But unlike the densely packed prokaryotic genome, the rice genome contains many stretches of DNA that do not code for proteins or RNA. Some of these sequences are “spacers,” which is another way of saying that they either have no function or that no function has yet been found. Others are repetitive sequences, such as the telomeric DNA at the ends of chromosomes (see Figure 11.18).

In addition to the genes for metabolism that they share with prokaryotes, eukaryotes have genes that mark them as complex organisms: genes for addressing, or targeting, proteins to organelles, and genes for cell–cell interaction and cell differentiation. The transcription and later processing of mRNA is more complicated in eukaryotes than in prokaryotes. Elegant molecular machinery allows the precise regulation of gene expression needed for all the cells of these complex organisms to develop and function.

“The Most Precious Things” The genome of rice (*Oryza sativa*), which directly supplies a third of the overall diet of humanity, was recently sequenced.

The Eukaryotic Genome

As biologists began to unravel the intricacies of gene structure and expression in prokaryotes, they tried to generalize their findings by stating, “What’s true for *E. coli* is also true for elephants.” Although much of prokaryotic biochemistry does apply to eukaryotes as well, the old saying has its limitations. Table 14.1 lists some of the differences between prokaryotic and eukaryotic genomes.

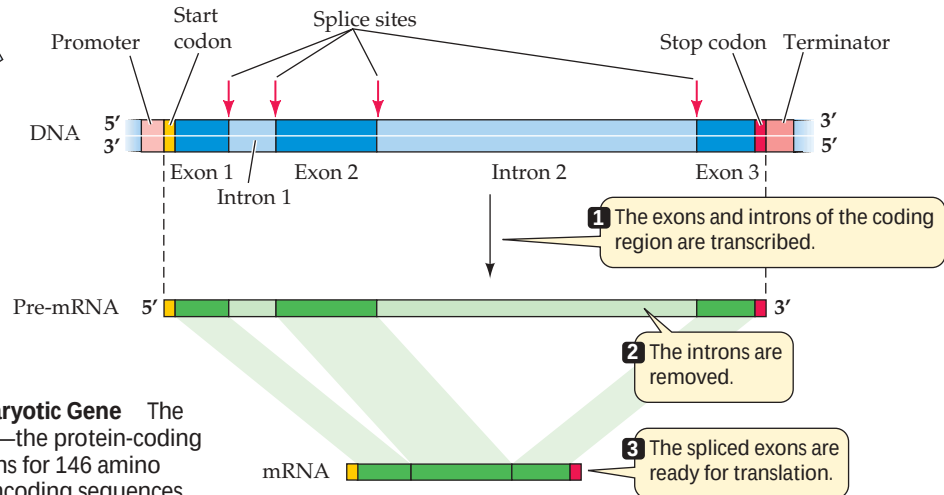
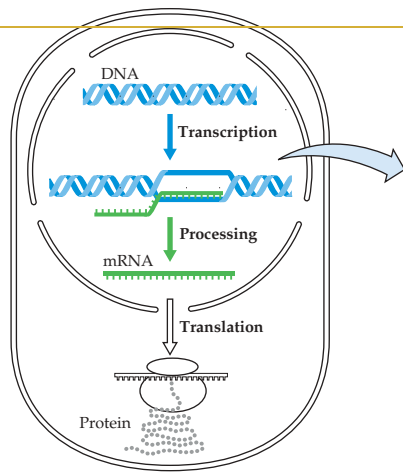
The eukaryotic genome is larger and more complex than the prokaryotic genome

Comparisons of prokaryotic and eukaryotic genomes reveal several features.



All VU books are available in hard copy and delivered all over Pakistan. Order now!

Website: vubookshoppk.com | WhatsApp: 0326-0775533

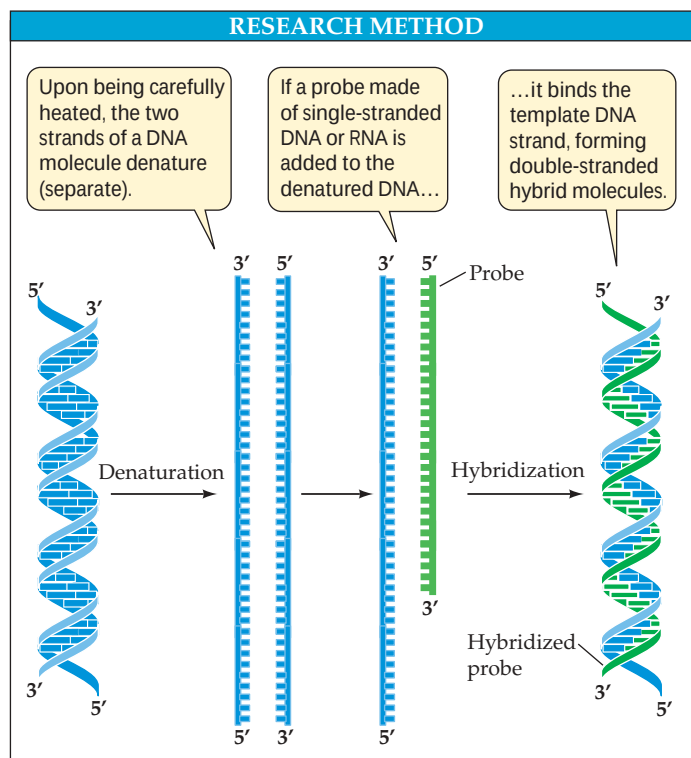


14.4 The Structure and Transcription of a Eukaryotic Gene The β -globin gene is about 1,600 bp long. The exons—the protein-coding sequences—contain 441 base pairs (triplet codons for 146 amino acids plus a triplet stop codon). The introns—noncoding sequences of DNA—between codons 30 and 31 (130 bp long) and 104 and 105 (850 bp long), are initially transcribed, but are spliced out of the initial mRNA transcript.

Where are the introns within a eukaryotic gene? The easiest way to find out is by **nucleic acid hybridization**, the method that originally revealed the existence of introns. This research method, outlined in Figure 14.5, has been crucial for studying the relationship between genes and their transcripts.

Biologists used nucleic acid hybridization to examine the β -globin gene, which encodes one of the globin proteins that make up hemoglobin (Figure 14.6). They first denatured the β -globin DNA by heating it, then added mature β -globin mRNA. As expected, the mRNA bound to the DNA by complementary base pairing. The researchers expected to obtain a linear matchup of the mRNA to the coding DNA. That expectation was met, but only in part: There were indeed stretches of RNA–DNA hybridization, but some looped structures were also visible. These loops were the introns, stretches of DNA that did not have complementary bases on the mature mRNA. Later studies showed that the hybridization of pre-mRNA to DNA was complete, revealing that the introns were indeed transcribed. Somewhere on the path from primary transcript (pre-mRNA) to mature mRNA, the introns had been removed, and the exons had been spliced together. We will examine this splicing process shortly.

Most (but not all) vertebrate genes contain introns, as do many other eukaryotic genes (and even a few prokaryotic ones). Introns interrupt, but do not scramble, the DNA sequence that codes for a polypeptide chain. The base sequence of the exons, taken in order, is complementary to that of the mature mRNA product. The introns, therefore, separate a gene's protein-coding region into distinct parts—the exons. In some cases, the separated exons code for different functional regions, or *domains*, of the protein. For example, the globin proteins that make up hemoglobin have two domains: one for binding to a nonprotein pigment called heme, and another for binding to the other globin subunits. These two domains are encoded by different exons in the globin genes.



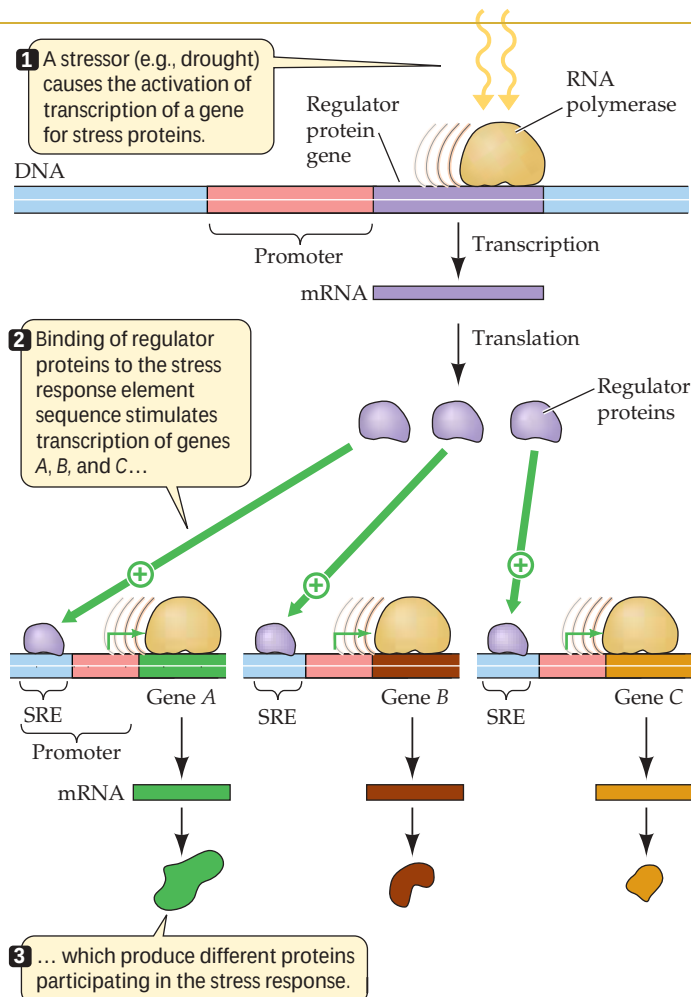
14.5 Nucleic Acid Hybridization Base pairing permits the detection of a sequence complementary to the probe.

Many eukaryotic genes are members of gene families

About half of all eukaryotic protein-coding genes are present in only one copy in the haploid genome. The rest have multiple

All VU books are available in hard copy and delivered all over Pakistan. Order now!

Website: vubookshoppk.com | WhatsApp: 0326-0775533



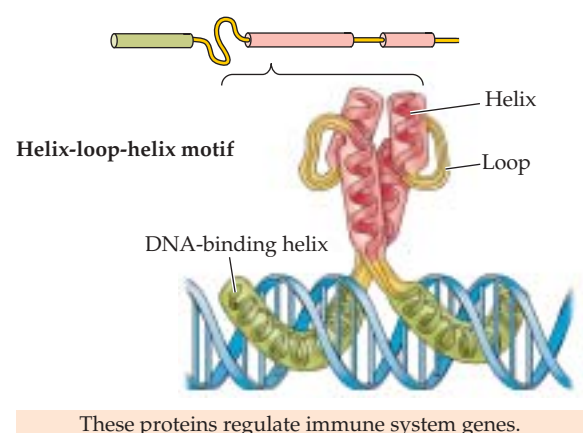
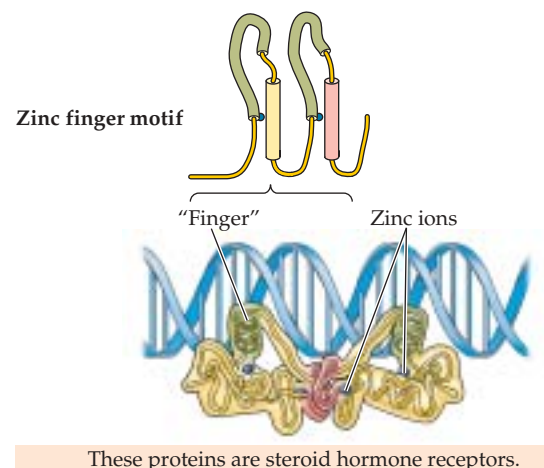
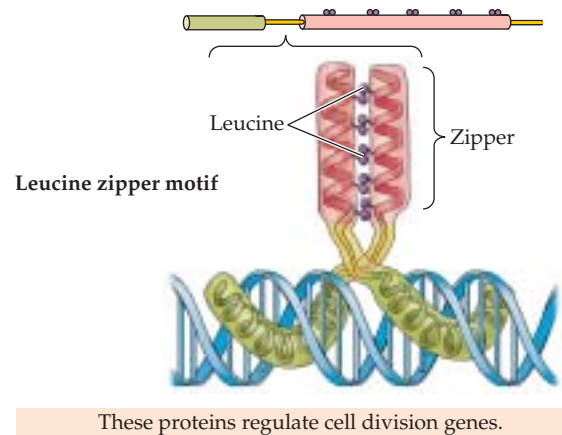
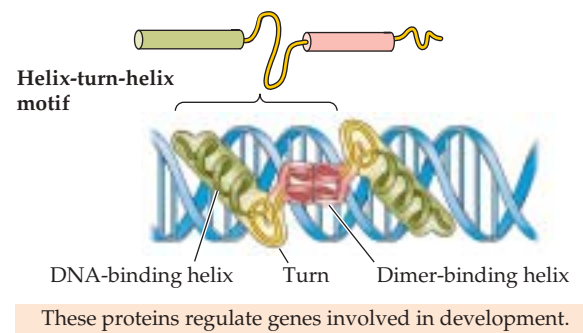
14.14 Coordinating Gene Expression A single signal, such as drought stress, causes the synthesis of a transcriptional regulator for many genes.

of certain types of genes, both during development and in the adult organism.

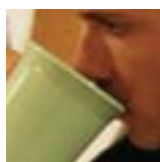
Genes can be inactivated by chromatin structure

Other mechanisms that regulate transcription act on the structure of chromatin and chromosomes. As we saw in Chapter 9, chromatin contains a number of proteins as well as DNA. The packaging of DNA into nucleosomes by these nuclear proteins can make DNA physically inaccessible to RNA polymerase and the rest of the transcription apparatus, much as the binding of a repressor to the operator in the prokaryotic *lac* operon prevents transcription. Chromatin structure at both the local and whole-chromosome levels affects transcription.

14.15 Protein-DNA Interactions The DNA-binding domains of most regulatory proteins have one of these four structural motifs.



12 *Cell Signaling and Communication*



It's probably happened to you: it's late at night, you have a paper due, and you've put it off until the last minute. You're exhausted, but you need to stay awake and alert so that you can get your work done. What do you do? You have a cup (or several cups) of coffee. Many people turn to the caffeine in coffee when they need to wake themselves up and give themselves an energy boost.

To understand how caffeine works, we must understand the pathways by which the body's cells respond to signals in their environment. There are three sequential steps involved in the cell's response to any signal. First, the signal binds to a receptor protein in the cell, often on the outside surface of the plasma membrane. Second, the binding of the signal causes a message to be conveyed to the inside of the cell and amplified. Third, the cell changes its activity in response to the signal.

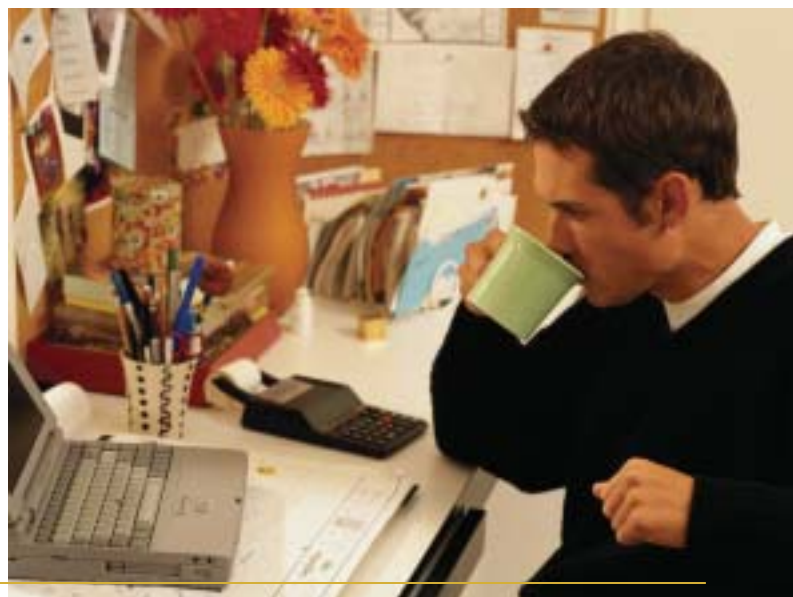
Caffeine acts in different ways in different tissues. A tired person's brain produces adenosine molecules that bind to specific receptor proteins, resulting in decreased brain activity and increased drowsiness. Caffeine's molecular structure is similar to that of adenosine, so it occupies the adenosine receptors without inhibiting brain cell function, and alertness is restored. In heart and liver cells, caffeine indirectly stimulates the same signaling pathway that is normally stimulated by epinephrine, the "fight-or-flight" hormone. In the heart, the result is an increased rate of beating; the liver is stimulated to convert glycogen into glucose and release it into the bloodstream.

We begin this chapter with a discussion of the signals that affect cells. As you will see, these signals include chemicals produced by other cells in the body as well as physical factors in the environment, such as light. Whatever the signal, it affects a cell only if that cell has a receptor protein that binds to that signal. In addition to binding the signal, the receptor must somehow communicate to the rest of the cell that binding has occurred. Finally, this communication process, called signal transduction, must result in a change in the function of the cell. We will describe these three steps of cell signaling—binding, transduction, and the cellular response—in order. We close the chapter with a description of how cells communicate with one another directly via specialized channels in their adjacent plasma membranes.

Signals

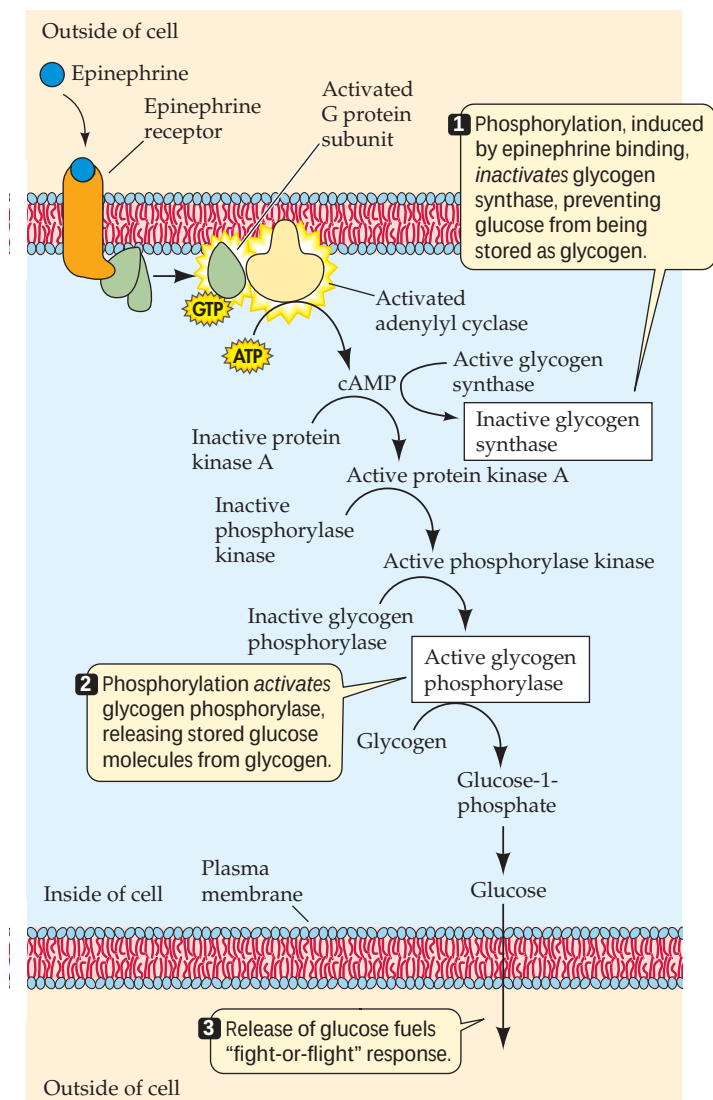
Both prokaryotic and eukaryotic cells process information from their environment. This information can be in the form of a physical stimulus, such as the light reaching your eye as you read this

A Signal to the Body The caffeine in coffee sends signals to cells in the body of this coffee drinker. The effects of these signals help him stay alert.



All VU books are available in hard copy and delivered all over Pakistan. Order now!

Website: vubookshoppk.com | WhatsApp: 0326-0775533

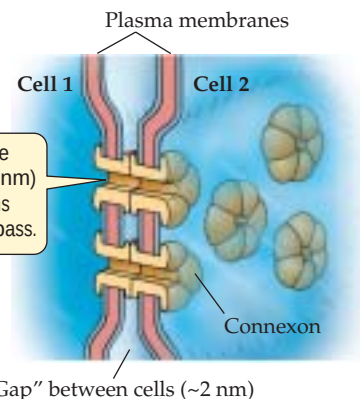


15.15 A Cascade of Reactions Leads to Altered Enzyme Activity

Liver cells respond to epinephrine by activating G proteins, which in turn activate cAMP synthesis. The second messenger initiates a protein kinase cascade. The cascade both inhibits the conversion of glucose to glycogen and stimulates the release of previously stored glucose.

Animal cells communicate by gap junctions

Gap junctions are channels between adjacent cells that occur in many animals, occupying up to 25 percent of the area of the plasma membrane (Figure 15.16). Gap junctions traverse the narrow space between the plasma membranes of two cells (the "gap") by means of thin molecular channels called *connexons*. The walls of these channels are composed of six subunits of an integral membrane protein. In two cells close to each other, two connexons come together, forming a channel that links the two cytoplasms. There may be hundreds of these channels between a cell and its neighbors. The



15.16 Gap Junctions Connect Animal Cells An animal cell may contain hundreds of gap junctions connecting it to neighboring cells. Gap junctions are too small for proteins, but small molecules such as ATP, metabolic intermediates, amino acids, and coenzymes can pass through them.

channels about 1.5 nm in diameter—far too narrow for the passage of large molecules such as proteins. But they are wide enough to allow small signal molecules and ions to pass between the cells. Experiments in which a labeled signal molecule or ion is injected into one cell show that it can readily pass into the adjacent cells if the cells are connected by gap junctions.

Gap junctions permit metabolic cooperation among the linked cells. Such cooperation ensures the sharing of important small molecules such as ATP, metabolic intermediates, amino acids, and coenzymes between cells. It may also ensure that concentrations of ions and small molecules are similar in linked cells, thereby maintaining equivalent regulation of metabolism. It is not clear how important this function is in many tissues, but it is known to be vital in some. In the lens of the mammalian eye, for example, only the cells at the periphery are close enough to the blood supply to allow diffusion of nutrients and wastes. But because lens cells are connected by large numbers of gap junctions, material can diffuse between them rapidly and efficiently.

There is evidence that signal molecules such as hormones and second messengers such as cAMP and PIP2 can move through gap junctions. If this is true, only a few cells would need to have receptors binding a signal in order for the stimulus to spread throughout the tissue. In this way, a tissue could have a coordinated response to the signal.

Plant cells communicate by plasmodesmata

Instead of gap junctions, plants have **plasmodesmata**, which are membrane-lined bridges spanning the thick cell walls that separate plant cells from one another. A typical plant cell has several thousand plasmodesmata.

All VU books are available in hard copy and delivered all over Pakistan. Order now!

Website: vubookshoppk.com | WhatsApp: 0326-0775533

General Principles of Cell Communication

Mechanisms enabling one cell to influence the behavior of another almost certainly existed in the world of unicellular organisms long before multicellular organisms appeared on Earth. Evidence comes from studies of present-day unicellular eucaryotes such as yeasts. Although these cells normally lead independent lives, they can communicate and influence one another's behavior in preparation for sexual mating. In the [budding yeast *Saccharomyces cerevisiae*](#),

Extracellular Signal Molecules Can Act Over Either Short or Long Distances

Many **signal** molecules remain bound to the surface of the **signaling** cell and influence only cells that contact. Such **contact-dependent signaling** is especially important during **development** and in immune responses. In most cases, however, **signal** molecules are secreted. The secreted molecules may be carried far afield to act on distant targets, or they may act as **local mediators**, affecting only cells in the immediate environment of the **signaling** cell. This latter process is called **paracrine signaling**. For paracrine signals to be delivered only to their proper target cells, the secreted molecules must not be allowed to diffuse too far; for this reason they are often rapidly taken up by neighboring target cells, destroyed by extracellular enzymes, or immobilized by the **extracellular matrix**.

For a large, **complex** multicellular organism, short-range **signaling** is not sufficient on its own to coordinate the behavior of its cells. In these organisms, sets of specialized cells have evolved with a specific role in communication between widely separate parts of the body. The most sophisticated of these are nerve cells, or neurons, which typically extend long processes (axons) that enable them to contact target cells far away. When activated by signals from the environment or from other nerve cells, a neuron sends electrical impulses (action potentials) rapidly along its **axon**; when such an impulse reaches the end of the axon, it causes the nerve terminals located there to secrete a chemical **signal** called a **neurotransmitter**. These signals are secreted at specialized cell junctions called *chemical synapses*, which are designed to ensure that the **neurotransmitter** is delivered specifically to the postsynaptic target cell.

A second type of specialized **signaling** cell that controls the behavior of the organism as a whole is an **endocrine cell**. These cells secrete their **signal** molecules, called **hormones**, into the bloodstream, which carries the **signal** to target cells distributed widely throughout the body.

Because endocrine **signaling** relies on **diffusion** and blood flow, it is relatively slow. Synaptic **signaling**, by contrast, can be much faster, as well as more precise. Nerve cells can transmit information over long distances by electrical impulses that travel at rates of up to 100 meters per second; once released from a nerve terminal, a **neurotransmitter** has to diffuse less than 100 **nm** to the target cell, a process that takes less than a millisecond. Another difference between endocrine and **synaptic signaling** is that, whereas hormones are greatly diluted in the bloodstream and interstitial fluid and therefore must be able to act at very low concentrations (typically $< 10^{-8}$ M), neurotransmitters are diluted much less and can achieve high local concentrations. The concentration of **acetylcholine** in the synaptic cleft of an active **neuromuscular junction**, for example, is about 5×10^{-4} M. Correspondingly, neurotransmitter receptors have a relatively low affinity for their **ligand**, which means that the neurotransmitter can dissociate rapidly from the **receptor** to terminate a response. Moreover, after its release from a nerve terminal, a neurotransmitter is quickly removed from the synaptic cleft, either by specific hydrolytic enzymes that destroy it or by specific **membrane transport** proteins that **pump** it back into either the nerve terminal or neighboring glial cells.

Thus, synaptic **signaling** is much more precise than endocrine **signaling**, both in time and in space.

The speed of a response to an extracellular **signal** depends not only on the mechanism of **signal** delivery, but also on the nature of the response in the target cell. Where the response requires only changes in proteins already present in the cell, it can occur in seconds or even milliseconds. When the response involves changes in [gene expression](#) and the synthesis of new proteins, however, it usually requires hours, irrespective of the mode of **signal** delivery.

Autocrine Signaling Can Coordinate Decisions by Groups of Identical Cells

All of the forms of **signaling** discussed so far allow one cell to influence another. Often, the **signaling** cell and target are different cell types. Cells, however, can also send signals to other cells of the same type, as well as to themselves. In such [autocrine signaling](#), a cell secretes **signal** molecules that can bind back to its own receptors. During [development](#), for example, once a cell has been directed along a particular pathway of [differentiation](#), it may begin to secrete autocrine signals to itself that reinforce this developmental decision.

Autocrine **signaling** is most effective when performed simultaneously by neighboring cells of the same type, and it is likely to be used to encourage groups of identical cells to make the same developmental decisions. Thus, [autocrinesignaling](#) is thought to be one possible mechanism underlying the “community effect” that is observed in early [development](#), during which a group of identical cells can respond to a [differentiation](#)-inducing **signal** but a single isolated cell of the same type cannot.

Unfortunately, cancer cells often use [autocrine signaling](#) to overcome the normal controls on cell proliferation and survival that we discuss later. By secreting signals that act back on the cell's own receptors, cancer cells can stimulate their own survival and proliferation and thereby survive and proliferate in places where normal cells of the same type could not. How this dangerous perturbation of normal cell behavior comes about is discussed in Chapter 23.

[Go to:](#)

Gap Junctions Allow Signaling Information to Be Shared by Neighboring Cells

Another way to coordinate the activities of neighboring cells is through [gap junctions](#). These are specialized cell-cell junctions that can form between closely apposed plasma membranes and directly connect the cytoplasm of the joined cells via narrow water-filled channels. The channels allow the exchange of small intracellular **signaling** molecules (*intracellular mediators*), such as Ca^{2+} and cyclic AMP (discussed later), but not of macromolecules, such as proteins or nucleic acids. Thus, cells connected by gap junctions can communicate with each other directly, without having to surmount the barrier presented by the intervening plasma membranes.

As discussed in Chapter 19, the pattern of gap-junction connections in a tissue can be revealed either electrically, with intracellular electrodes, or visually, after the [microinjection](#) of small water-soluble dyes. Studies of this kind indicate that the cells in a developing embryo make and break gap-junction connections in specific and interesting patterns, strongly suggesting that these junctions have an important role in the **signaling** processes that occur between these cells. Mice and humans that are deficient in one particular gap-junction [protein](#) (connexin 43), for example, have severe defects in heart [development](#). Like the [autocrine signaling](#) described above, gap-junction communication helps adjacent cells of a similar type to coordinate their behavior. It is still not known, however, which particular small molecules are important as carriers of signals through gap junctions, and the specific functions of gap-junction communication in animal development remain uncertain.

Each Cell Is Programmed to Respond to Specific Combinations of Extracellular SignalMolecules

A typical cell in a multicellular organism is exposed to hundreds of different signals in its environment. These signals can be soluble, bound to the [extracellular matrix](#), or bound to the surface of a neighboring cell, and they can act in many millions of combinations. The cell must respond to this babel of signals selectively, according to its own specific character, which it has acquired through progressive cell specialization in the course of [development](#). A cell may be programmed to respond to one combination of signals by differentiating, to another combination by multiplying, and to yet another by performing some specialized function such as contraction or secretion.

Most of the cells in a [complex](#) animal are also programmed to depend on a specific combination of signals simply to survive. When deprived of these signals (in a culture dish, for example), a cell activates a suicide program and kills itself—a process called [programmed cell death](#), or [apoptosis](#). Because different types of cells require different combinations of survival signals, each cell type is restricted to different environments in the body. The ability to undergo apoptosis is a fundamental property of animal cells.

In principle, the hundreds of **signal** molecules that animals make can be used to create an almost unlimited number of **signaling** combinations. The use of these combinations to control cell behavior enables an animal to control its cells in highly specific ways by using a limited diversity of **signal** molecules.

Different Cells Can Respond Differently to the Same Extracellular Signal Molecule

The specific way in which a cell reacts to its environment varies. It varies according to the set of [receptor](#) proteins the cell possesses, which determines the particular subset of signals it can respond to, and it varies according to the intracellular machinery by which the cell integrates and interprets the signals it receives. Thus, a single [signal molecule](#) often has different effects on different target cells. The [neurotransmitter acetylcholine](#), for example, stimulates the

contraction of skeletal muscle cells, but it decreases the rate and force of contraction in heart muscle cells. This is because the [acetylcholine receptor](#) proteins on skeletal muscle cells are different from those on heart muscle cells. But receptor differences are not always the explanation for the different effects. In many cases, the same **signal** molecule binds to identical receptor proteins yet produces very different responses in different types of target cells, reflecting differences in the internal machinery to which the receptors are coupled.

The Concentration of a Molecule Can Be Adjusted Quickly Only If the Lifetime of the Molecule Is Short

It is natural to think of **signaling** systems in terms of the changes produced when a **signal** is delivered. But it is just as important to consider what happens when a **signal** is withdrawn. During [development](#), transient signals often produce lasting effects: they can trigger a change in the cell's development that persists indefinitely, through cell memory mechanism. In most cases in adult tissues, however, the response fades when a **signal** ceases. The effect is transitory because the **signal** exerts its effects by altering a set of molecules that are unstable, undergoing continual turnover. Thus, once the **signal** is shut off, the replacement of the old molecules by new ones wipes out all traces of its action. It follows that the speed with which a cell responds to **signal** removal depends on the rate of destruction, or turnover, of the molecules the **signal** affects.

It is also true, although much less obvious, that this turnover rate also determines the promptness of the response when a **signal** is turned on. Consider, for example, two intracellular **signaling** molecules X and Y, both of which are normally maintained at a concentration of 1000 molecules per cell. Molecule Y is synthesized and degraded at a rate of 100 molecules per second, with each [molecule](#) having an average lifetime of 10 seconds. Molecule X has a turnover rate that is 10 times slower than that of Y: it is both synthesized and degraded at a rate of 10 molecules per second, so that each molecule has an average lifetime in the cell of 100 seconds. If a **signal** acting on the cell boosts the rates of synthesis of both X and Y tenfold without any change in the molecular lifetimes, at the end of 1 second the concentration of Y will have increased by nearly 900 molecules per cell ($10 \times 100 - 100$), while the concentration of X will have increased by only 90 molecules per cell. In fact, after a molecule's synthesis rate has been either increased or decreased abruptly, the time required for the molecule to shift halfway from its old to its new [equilibrium](#) concentration is equal to its normal half-life—that is, equal to the time that would be required for its concentration to fall by half if all synthesis were stopped.

The same principles apply to proteins and small molecules, and to molecules in the extracellular space and inside cells. Many intracellular proteins have short half-lives, some surviving for less than 10 minutes. In most cases, these are proteins with key regulatory roles, whose concentrations are rapidly regulated in the cell by changes in their rates of synthesis. Likewise, any covalent modifications of proteins that occur as part of a

rapid **signaling** process—most commonly, the addition of a phosphate group to an [amino acid side chain](#)—must be continuously removed at a rapid rate to make rapid **signaling** possible.

We shall discuss some of these molecular events in detail later for **signaling** pathways that operate via cell-surface receptors. But the principles apply quite generally, as the next example illustrates.

Nitric Oxide Gas Signals by Binding Directly to an Enzyme Inside the Target Cell

Although most extracellular signals are [hydrophilic](#) molecules that bind to receptors on the surface of the target cell, some **signal** molecules are hydrophobic enough and/or small enough to pass readily across the target-cell [plasma membrane](#). Once inside, they directly regulate the activity of a specific intracellular [protein](#). An important and remarkable example is the gas [nitric oxide \(NO\)](#), which acts as a [signal molecule](#) in both animals and plants. In mammals, one of its functions is to regulate smooth muscle contraction. Acetylcholine, for example, is released by autonomic nerves in the walls of a blood vessel, and it causes smooth muscle cells in the vessel wall to relax. The [acetylcholine](#) acts indirectly by inducing the nearby endothelial cells to make and release NO, which then signals the underlying smooth muscle cells to relax. This effect of NO on blood vessels provides an explanation for the mechanism of action of nitroglycerine, which has been used for about 100 years to treat patients with angina (pain resulting from inadequate blood flow to the heart muscle). The nitroglycerine is converted to NO, which relaxes blood vessels. This reduces the workload on the heart and, as a consequence, it reduces the oxygen requirement of the heart muscle.

Many types of nerve cells use [NO](#) gas to **signal** to their neighbors. The NO released by autonomic nerves in the penis, for example, causes the local blood vessel dilation that is responsible for penile erection. NO is also produced as a [local mediator](#) by activated macrophages and neutrophils to help them to kill invading microorganisms. In plants, NO is involved in the defensive responses to injury or infection.

[NO](#) gas is made by the deamination of the [amino acid](#) arginine, catalyzed by the [enzyme](#) *NO synthase*. Because it passes readily across membranes, dissolved NO rapidly diffuses out of the cell where it is produced and into neighboring cells. It acts only locally because it has a short half-life—about 5–10 seconds—in the extracellular space before it is converted to nitrates and nitrites by oxygen and water. In many target cells, including endothelial cells, NO binds to iron in the [active site](#) of the enzyme *guanylyl cyclase*, stimulating this enzyme to produce the [small intracellular mediator](#) *cyclic GMP*, which we discuss later. The effects of NO can occur within seconds, because the normal rate of turnover of cyclic GMP is high: a rapid degradation to GMP by a phosphodiesterase constantly balances the production of cyclic GMP from GTP by guanylyl cyclase. The drug Viagra inhibits this cyclic GMP phosphodiesterase in the penis, thereby increasing the amount of time that cyclic GMP levels

remain elevated after NO production is induced by local nerve terminals. The cyclic GMP, in turn, keeps blood vessels relaxed and the penis erect.

Carbon monoxide (CO) is another gas that is used as an intercellular **signal**. It can act in the same way as [NO](#), by stimulating guanylyl cyclase. These gases are not the only **signal** molecules that can pass directly across the target-cell [plasma membrane](#). A group of small, hydrophobic, nongaseous hormones and local mediators also enter target cells in this way. But instead of binding to enzymes, they bind to intracellular [receptor](#) proteins that directly regulate [gene](#)transcription, as we discuss next.

Nuclear Receptors Are Ligand-activated Gene Regulatory Proteins

A number of small hydrophobic **signal** molecules diffuse directly across the [plasma membrane](#) of target cells and bind to intracellular [receptor](#) proteins. These **signal** molecules include [steroid hormones](#), *thyroid hormones*, *retinoids*, and *vitamin D*. Although they differ greatly from one another in both chemical structure and function, they all act by a similar mechanism. When these **signal** molecules bind to their receptor proteins, they activate the receptors, which bind to [DNA](#) to regulate the transcription of specific genes. The receptors are all structurally related, being part of the [nuclear receptor superfamily](#). This very large superfamily also includes some receptor proteins that are activated by intracellular metabolites rather than by secreted **signal** molecules. Many family members have been identified by [DNA sequencing](#) only, and their [ligand](#) is not yet known; these proteins are therefore referred to as *orphan nuclear receptors*. The importance of such nuclear receptors in some animals is indicated by the fact that 1–2% of the genes in the nematode *C. elegans* code for them, although there are fewer than 50 in humans.

Steroid hormones—which include cortisol, the [steroid](#) sex hormones, vitamin D (in vertebrates), and the moulting [hormone](#) ecdysone (in insects)—are all made from [cholesterol](#). *Cortisol* is produced in the cortex of the adrenal glands and influences the [metabolism](#) of many types of cells. The steroid sex hormones are made in the testes and ovaries, and they are responsible for the secondary sex characteristics that distinguish males from females. Vitamin D is synthesized in the skin in response to sunlight; after it has been converted to its active form in the liver or kidneys, it regulates Ca^{2+} metabolism, promoting Ca^{2+} uptake in the gut and reducing its excretion in the kidneys. The thyroid hormones, which are made from the [amino acid](#) tyrosine, act to increase the metabolic rate in a wide variety of cell types, while the retinoids, such as retinoic acid, are made from vitamin A and have important roles as local mediators in vertebrate [development](#). Although all of these **signal** molecules are relatively insoluble in water, they are made soluble for transport in the bloodstream and other extracellular fluids by binding to specific carrier proteins, from which they dissociate before entering a target cell.

Beside a fundamental difference in the way they **signal** their target cells, most water-insoluble **signal** molecules differ from water-soluble ones in the length of time they persist in the bloodstream or tissue fluids. Most water-soluble hormones are removed and/or broken

down within minutes of entering the blood, and local mediators and neurotransmitters are removed from the extracellular space even faster—within seconds or milliseconds. Steroid hormones, by contrast, persist in the blood for hours and thyroid hormones for days. Consequently, water-soluble **signal** molecules usually mediate responses of short duration, whereas water-insoluble ones tend to mediate responses that are longer lasting.

The intracellular receptors for the **steroid** and thyroid hormones, retinoids, and vitamin D all bind to specific **DNA** sequences adjacent to the genes the **ligand** regulates. Some receptors, such as those for cortisol, are located primarily in the **cytosol** and enter the **nucleus** after ligand binding; others, such as the thyroid and retinoid receptors, are bound to DNA in the nucleus even in the absence of ligand. In either case, the inactive receptors are bound to inhibitory **protein** complexes, and ligand binding alters the **conformation** of the **receptor** protein, causing the inhibitory **complex** to dissociate. The ligand binding also causes the receptor to bind to coactivator proteins that induce **gene** transcription. The transcriptional response usually takes place in successive steps: the direct activation of a small number of specific genes occurs within about 30 minutes and constitutes the *primary response*; the protein products of these genes in turn activate other genes to produce a delayed, *secondary response*; and so on. In this way, a simple hormonal trigger can cause a very complex change in the pattern of gene **expression**.

The responses to **steroid** and thyroid hormones, vitamin D, and retinoids, like responses to extracellular signals in general, are **determined** as much by the nature of the target cell as by the nature of the **signal molecule**. Many types of cells have the identical intracellular **receptor**, but the set of genes that the receptor regulates is different in each cell type. This is because more than one type of **gene regulatory protein** generally must bind to a eucaryotic gene to activate its transcription. An intracellular receptor can therefore activate a gene only if there is the right combination of other gene regulatory proteins, and many of these are cell-type specific. Thus, each of these hormones induces a characteristic set of responses in an animal for two reasons. First, only certain types of cells have receptors for it. Second, each of these cell types contains a different combination of other cell-type-specific gene regulatory proteins that collaborate with the activated receptor to influence the transcription of specific sets of genes.

The Three Largest Classes of Cell-Surface Receptor Proteins Are Ion-Channel-linked, G-Protein-linked, and Enzyme-linked Receptors

As mentioned previously, all water-soluble **signal** molecules (including neurotransmitters and all **signal** proteins) bind to specific **receptor** proteins on the surface of the target cells that they influence. These cell-surface receptor proteins act **assignal transducers**. They convert an extracellular **ligand**-binding event into intracellular signals that alter the behavior of the target cell.

Most cell-surface **receptor** proteins belong to one of three classes, defined by the **transduction** mechanism they use. **Ion-channel-linked receptors**, also known

as *transmitter-gated ion channels* or *ionotropic receptors*, are involved in rapid **synaptic signaling** between electrically excitable cells. This type of **signaling** is mediated by a small number of neurotransmitters that transiently open or close an **ion channel** formed by the **protein** to which they bind, briefly changing the ion permeability of the **plasma membrane** and thereby the excitability of the postsynaptic cell. The ion-channel-linked receptors belong to a large family of **homologous**, multipass transmembrane proteins.

G-protein-linked receptors act indirectly to regulate the activity of a separate plasma-membrane-bound target protein, which can be either an **enzyme** or an **ion channel**. The interaction between the **receptor** and this target protein is mediated by a third protein, called a **trimeric GTP-binding protein** (*G protein*). The activation of the target protein can change the concentration of one or more intracellular mediators (if the target protein is an enzyme), or it can change the ion permeability of the **plasma membrane** (if the target protein is an ion channel). The intracellular mediators affected act in turn to alter the behavior of yet other **signaling** proteins in the cell. All of the G-protein-linked receptors belong to a large family of **homologous**, seven-pass transmembrane proteins.

Enzyme-linked receptors, when activated, either function directly as enzymes or are directly associated with enzymes that they activate. They are formed by single-pass transmembrane proteins that have their **ligand-binding site** outside the cell and their catalytic or **enzyme-binding site** inside. Enzyme-linked receptors are heterogeneous in structure compared with the other two classes. The great majority, however, are **protein kinases**, or are associated with protein kinases, and ligand binding to them causes the **phosphorylation** of specific sets of proteins in the target cell.

There are some cell-surface receptors that do not fit into any of the above classes. Some of these depend on intracellular proteolytic events to **signal** the cell, and we discuss them only after we explain in detail how **G-protein-linked receptors** and **enzyme-linked receptors** operate. We start with some general principles of **signaling** via cell-surface receptors.

Most Activated Cell-Surface Receptors Relay Signals Via Small Molecules and a Network of Intracellular Signaling Proteins

Signals received at the surface of a cell by either **G-protein-linked** or **enzyme-linked** receptors are relayed into the cell interior by a combination of small and large **intracellular signaling molecules**. The resulting chain of intracellular **signaling** events ultimately alters *target proteins*, and these altered target proteins are responsible for modifying the behavior of the cell.

The small intracellular **signaling** molecules are called **small intracellular mediators**, or **second messengers** (the “first messengers” being the extracellular signals). They are generated in large numbers in response to **receptor** activation and rapidly diffuse away from their source, broadcasting the **signal** to other parts of the cell. Some, such as *cyclic AMP* and Ca^{2+} , are water-soluble and diffuse in the **cytosol**, while others, such as *diacylglycerol*, are **lipid-soluble** and diffuse in the plane of the **plasma membrane**. In either

case, they pass the **signal** on by binding to and altering the behavior of selected **signaling** proteins or target proteins.

The large intracellular **signaling** molecules are [intracellular signaling proteins](#). Many of these relay the **signal** into the cell by either activating the next **signaling protein** in the chain or generating small intracellular mediators. These proteins can be classified according to their particular function, although many fall into more than one category:

1.

Relay proteins simply pass the message to the next **signaling** component in the chain.

2.

Messenger proteins carry the **signal** from one part of the cell to another, such as from the [cytosol](#) to the [nucleus](#).

3.

Adaptor proteins link one **signaling protein** to another, without themselves conveying a **signal**.

4.

Amplifier proteins, which are usually either enzymes or [ion](#) channels, greatly increase the **signal** they receive, either by producing large amounts of small intracellular mediators or by activating large numbers of downstream intracellular **signaling** proteins. When there are multiple amplification steps in a relay chain, the chain is often referred to as a **signaling cascade**.

5.

Transducer proteins convert the **signal** into a different form. The [enzyme](#) that makes cyclic AMP is an example: it both converts the **signal** and amplifies it, thus acting as both a transducer and an amplifier.

6.

Bifurcation proteins spread the **signal** from one **signaling** pathway to another.

7.

Integrator proteins receive signals from two or more **signaling** pathways and integrate them before relaying a **signal** onward.

8.

Latent [gene](#) regulatory proteins are activated at the cell surface by activated receptors and then migrate to the [nucleus](#) to stimulate gene transcription.

Other types of intracellular proteins also have important roles in intracellular **signaling**. *Modulator proteins* modify the activity of intracellular **signaling** proteins and thereby regulate the strength of **signaling** along the pathway. *Anchoring proteins* maintain specific **signaling** proteins at a precise location in the cell by tethering them to a [membrane](#) or the [cytoskeleton](#). *Scaffold proteins* are adaptor and/or anchoring proteins that bind multiple **signaling** proteins together in a functional [complex](#) and often hold them at a specific location.

Some Intracellular Signaling Proteins Act as Molecular Switches

Many intracellular **signaling** proteins behave like **molecular switches**: on receipt of a **signal** they switch from an inactive to an active state, until another process switches them off. As we discussed earlier, the switching off is just as important as the switching on. If a **signaling** pathway is to recover after transmitting a **signal** so that it can be ready to transmit another, every activated [molecule](#) in the pathway must be returned to its original inactivated state.

The molecular switches fall into two main classes that operate in different ways, although in both cases it is the gain or loss of phosphate groups that determines whether the [protein](#) is active or inactive. The largest class consists of proteins that are activated or inactivated by [phosphorylation](#) (discussed in Chapter 3). For these proteins, the switch is thrown in one direction by a [protein kinase](#), which adds one or more phosphate groups to the **signaling** protein, and in the other direction by a [protein phosphatase](#), which removes the phosphate groups from the protein. It is estimated that one-third of the proteins in a eucaryotic cell are phosphorylated at any given time.

Many of the **signaling** proteins controlled by [phosphorylation](#) are themselves [protein](#) kinases, and these are often organized into **phosphorylation cascades**. One [protein kinase](#), activated by phosphorylation, phosphorylates the next protein kinase in the sequence, and so on, relaying the **signal** onward and, in the process, amplifying it and sometimes spreading it to other **signaling** pathways. Two main types of protein kinases operate as intracellular **signaling** proteins. The great majority are *serine/threonine kinases*, which phosphorylate proteins on serines and (less often) threonines. Others are *tyrosine kinases*, which phosphorylate proteins on tyrosines. An occasional kinase can do both. Genome sequencing reveals that about 2% of our genes encode protein kinases, and it is thought that hundreds of distinct types of protein kinases are present in a typical mammalian cell.

The other main class of molecular switches involved in **signaling** are **GTP-binding proteins** (discussed in Chapter 3). These switch between an active state when GTP is bound and an inactive state when GDP is bound. Once activated, they have intrinsic [GTPase](#) activity and shut themselves off by hydrolyzing their bound GTP to GDP. There are two major types of GTP-binding proteins—large *trimeric GTP-binding proteins* (also called [G proteins](#)), which relay the signals from [G-protein](#)-linked receptors, and

small *monomeric GTPases* (also called *monomeric GTP-binding proteins*). The latter also help to relay intracellular signals, but in addition they are involved in regulating vesicular traffic and many other processes in eucaryotic cells.

As discussed earlier, [complex](#) cell behaviors, such as cell survival and cell proliferation, are generally stimulated by specific combinations of extracellular signals rather than by a single **signal** acting alone. The cell therefore has to integrate the information coming from separate signals so as to make an appropriate response—to live or die, to divide or not, and so on. This integration usually depends on integrator proteins, which are equivalent to the microprocessors in a computer: they require multiple **signal** inputs to produce an output that causes the desired biological effect. Two examples that show how such integrator proteins can operate are illustrated in.

Intracellular Signaling Complexes Enhance the Speed, Efficiency, and Specificity of the Response

Even a single type of extracellular **signal** acting through a single type of [G-protein](#)-linked or [enzyme-linked receptor](#) usually activates multiple parallel **signaling** pathways and can thereby influence multiple aspects of cell behavior—such as shape, movement, [metabolism](#), and [gene expression](#). Indeed, these two main classes of cell-surface receptors often activate some of the same **signaling** pathways, and there is usually no obvious reason why a particular extracellular **signal** utilizes one class of receptors rather than the other.

The complexity of these **signal**-response systems, with multiple interacting relay chains of **signaling** proteins, is daunting. It is not clear how an individual cell manages to display specific responses to so many different extracellular signals, many of which bind to the same class of [receptor](#) and activate many of the same **signaling** pathways. One strategy that the cell uses to achieve specificity involves [scaffold proteins](#), which organize groups of interacting **signaling** proteins into **signaling complexes**. Because the scaffold guides the interactions between the successive components in such a [complex](#), the **signal** can be relayed with precision, speed, and efficiency; moreover, unwanted cross-talk between **signaling** pathways is avoided. In order to amplify a **signal**, however, and spread it to other parts of the cell, at least some of the components in most **signaling** pathways are likely to be freely diffusible.

In other cases, **signaling** complexes form only transiently, as when **signaling** proteins assemble around a [receptor](#) after an extracellular [signal molecule](#) has activated it. In some of these cases, the cytoplasmic tail of the activated receptor is phosphorylated during the activation process, and the phosphorylated amino acids then serve as docking sites for the assembly of other **signaling** proteins ([Figure 15-19B](#)). In yet other cases, receptor activation leads to the production of modified [phospholipid](#) molecules in the adjacent [plasma membrane](#), and these lipids then recruit specific intracellular **signaling** proteins to this region

of membrane. All such **signaling** complexes form only transiently and rapidly disassemble after the extracellular **ligand** dissociates from the receptor.

Interactions Between Intracellular Signaling Proteins Are Mediated by Modular Binding Domains

The assembly of both stable and transient **signaling** complexes depends on a variety of highly conserved, small **binding domains** that are found in many intracellular **signaling** proteins. Each of these compact **protein** modules binds to a particular structural **motif** in the protein (or **lipid**) with which the **signaling** protein interacts. Because of these modular domains, **signaling** proteins bind to one another in multiple combinations, like Lego bricks, with the proteins often forming a three-dimensional network of interactions that determines the route followed by the **signaling** pathway. By joining existing domains together in novel combinations, the use of such modular binding domains has presumably facilitated the rapid evolution of new **signaling** pathways.

Src homology 2 (SH2) domains and phosphotyrosine-binding (PTB) domains, for example, bind to phosphorylated tyrosines in a particular peptide sequence on activated receptors or intracellular **signaling** proteins. *Src* homology 3 (SH3) domains bind to a short proline-rich **amino acid** sequence. *Pleckstrin* homology (PH) domains (first described in the *Pleckstrin* **protein** in blood platelets) bind to the charged headgroups of specific phosphorylated inositol phospholipids that are produced in the **plasma membrane** in response to an extracellular **signal**; they thereby enable the protein they are part of to dock on the membrane and interact with other recruited **signaling** proteins. Some **signaling** proteins function only as adaptors to link two other proteins together in a **signaling** pathway, and they consist solely of two or more binding domains.

Scaffold proteins often contain multiple *PDZ* domains (originally found in a region of a **synapse** called the postsynaptic density), each of which binds to a specific **motif** on a **receptor** or **signaling protein**. The *InaD* **scaffold protein** in *Drosophila* **photoreceptor** cells is a striking example. It contains five PDZ domains, one of which binds a light-activated **ion channel**, while the others each bind to a different **signaling** protein involved in the response of the cell to light. If any of these PDZ domains are missing, the corresponding **signaling** protein fails to assemble in the **complex**, and the fly's vision is defective.

Some cell-surface receptors and intracellular **signaling** proteins are thought to cluster together transiently in specific microdomains in the **lipid bilayer** of the **plasma membrane** that are enriched in **cholesterol** and glycolipids. Some of the proteins are directed to these **lipid rafts** by covalently attached lipid molecules. Like scaffold proteins, these lipid scaffolds may promote speed and efficiency in the **signaling** process by serving as sites where **signaling** molecules can assemble and interact.

Cells Can Respond Abruptly to a Gradually Increasing Concentration of an Extracellular Signal

Some cellular responses to extracellular **signal** molecules are smoothly graded in simple proportion to the concentration of the molecule. The primary responses to steroid hormones (see [Figure 15-14](#)) often follow this pattern, presumably because the nuclear hormone receptor protein binds a single molecule of hormone and each specific DNA recognition sequence in a steroid-hormone-responsive gene acts independently. As the concentration of hormone increases, the concentration of activated receptor-hormone complexes increases proportionally, as does the number of complexes bound to specific recognition sequences in the responsive genes; the response of the cell is therefore a gradual and linear one.

Many responses to extracellular **signal** molecules, however, begin more abruptly as the concentration of the molecule increases. Some may even occur in a nearly all-or-none manner, being undetectable below a threshold concentration of the molecule and then reaching a maximum as soon as this concentration is exceeded. What might be the molecular basis for such steep or even switchlike responses to graded signals?

One mechanism for sharpening the response is to require that more than one intracellular effector molecule or complex bind to some target macromolecule to induce a response. In some steroid-hormone-induced responses, for example, it seems that more than one activated receptor-hormone complex must bind simultaneously to specific regulatory sequences in the DNA to activate a particular gene. As a result, as the hormone concentration rises, gene activation begins more abruptly than it would if only one bound complex were sufficient for activation. A similar cooperative mechanism often operates in the **signaling** cascades activated by cell-surface receptors. As we discuss later, four molecules of the small intracellular mediator cyclic AMP, for example, must bind to each molecule of cyclic-AMP-dependent protein kinase to activate the kinase. Such responses become sharper as the number of cooperating molecules increases, and if the number is large enough, responses approaching the all-or-none type can be achieved.

Responses are also sharpened when an intracellular **signaling molecule** activates one enzyme and, at the same time, inhibits another enzyme that catalyzes the opposite reaction. A well-studied example of this common type of regulation is the stimulation of glycogen breakdown in skeletal muscle cells induced by the hormone *adrenaline* (epinephrine). Adrenaline's binding to a G-protein-linked cell-surface receptor leads to an increase in intracellular cyclic AMP concentration, which both activates an enzyme that promotes glycogen breakdown and inhibits an enzyme that promotes glycogen synthesis.

All of these mechanisms can produce responses that are very steep but, nevertheless, always smoothly graded according to the concentration of the extracellular **signal molecule**. Another mechanism, however, can produce true all-or-none responses, such that raising the **signal** above a critical threshold level trips a sudden switch in the responding cell. All-or-none threshold responses of this type generally depend on *positive feedback*; by this mechanism, nerve and muscle cells generate all-or-none *action potentials* in response to

neurotransmitters (discussed in Chapter 11). The activation of **ion**-channel-linked **acetylcholine** receptors at a **neuromuscular junction**, for example, results in a net influx of Na^+ that locally depolarizes the muscle **plasma membrane**. This causes voltage-gated Na^+ channels to open in the same membrane region, producing a further influx of Na^+ , which further depolarizes the membrane and thereby opens more Na^+ channels. If the initial depolarization exceeds a certain threshold value, this positive feedback has an explosive “runaway” effect, producing an **action potential** that propagates to involve the entire muscle membrane.

An accelerating positive feedback mechanism can also operate through **signaling** proteins that are enzymes rather than **ion** channels. Suppose, for example, that a particular intracellular **signaling ligand** activates an **enzyme** located downstream in a **signaling** pathway and that two or more molecules of the product of the enzymatic **reaction** bind back to the same enzyme to activate it further. The consequence is a very low rate of synthesis of the product in the absence of the ligand. The rate increases slowly with the concentration of ligand until, at some threshold level of ligand, enough of the product has been synthesized to activate the enzyme in a self-accelerating, runaway fashion. The concentration of the product then suddenly increases to a much higher level. Through these and a number of other mechanisms not discussed here, the cell will often translate a gradual change in the concentration of a **signaling** ligand into a switchlike change, creating an all-or-none response by the cell.

A Cell Can Remember The Effect of Some Signals

The effect of an extracellular **signal** on a target cell can, in some cases, persist well after the **signal** has disappeared. The enzymatic accelerating positive feedback system just described represents one type of mechanism that displays this kind of persistence. If such a system has been switched on by raising the concentration of intracellular activating **ligand** above threshold, it will generally remain switched on even when the extracellular **signal** disappears; instead of faithfully reflecting the current level of **signal**, the response system displays a memory. We shall encounter a specific example of this later, when we discuss a **protein kinase** that is activated by Ca^{2+} to phosphorylate itself and other proteins; the autophosphorylation keeps the kinase active long after Ca^{2+} levels return to normal, providing a memory trace of the initial **signal**.

Transient extracellular signals often induce much longer-term changes in cells during the **development** of a multicellular organism. Some of these changes can persist for the lifetime of the organism. They usually depend on self-activating memory mechanisms that operate further downstream in a **signaling** pathway, at the level of **gene** transcription. The signals that trigger muscle cell determination, for example, turn on a series of muscle-specific gene regulatory proteins that stimulate the transcription of their own genes, as well as genes producing many other muscle cell proteins. In this way, the decision to become a muscle cell is made permanent.

Cells Can Adjust Their Sensitivity to a Signal

In responding to many types of stimuli, cells and organisms are able to detect the same percentage of change in a **signal** over a very wide range of stimulus intensities. This requires that the target cells undergo a reversible process of [adaptation](#), or [desensitization](#), whereby a prolonged exposure to a stimulus decreases the cells' response to that level of exposure. In chemical **signaling**, [adaptation](#) enables cells to respond to *changes* in the concentration of a **signaling ligand** (rather than to the absolute concentration of the ligand) over a very wide range of ligand concentrations. The general principle is one of a negative feedback that operates with a delay. A strong response modifies the machinery for making that response, such that the machinery resets itself to an off position. Owing to the delay, however, a sudden change in the stimulus is able to make itself felt strongly for a short period before the negative feedback has time to kick in.

Desensitization to a [signal molecule](#) can occur in various ways. Ligand binding to cell-surface receptors, for example, may induce their [endocytosis](#) and temporary sequestration in endosomes. Such [ligand](#)-induced [receptor](#) endocytosis can lead to the destruction of the receptors in lysosomes, a process referred to as *receptor down-regulation*. In other cases, [desensitization](#) results from a rapid inactivation of the receptors—for example, as a result of a receptor [phosphorylation](#) that follows its activation, with a delay. Desensitization can also be caused by a change in a [protein](#) involved in transducing the **signal** or by the production of an inhibitor that blocks the **transduction** process.

Having discussed some of the general principles of cell **signaling**, we now turn to the [G-protein](#)-linked receptors. These are by far the largest class of cell-surface receptors, and they mediate the responses to the great majority of extracellular signals. This superfamily of [receptor](#) proteins not only mediates intercellular communication; it is also central to vision, smell, and taste perception.

Summary

Each cell in a multicellular animal has been programmed during [development](#) to respond to a specific set of extracellular signals produced by other cells. These signals act in various combinations to regulate the behavior of the cell. Most of the signals mediate a form of **signaling** in which local mediators are secreted, but then are rapidly taken up, destroyed, or immobilized, so that they act only on neighboring cells. Other signals remain bound to the outer surface of the **signaling** cell and mediate [contact-dependent signaling](#). Centralized control is exerted both by endocrine **signaling**, in which hormones secreted by endocrine cells are carried in the blood to target cells throughout the body, and by [synaptic signaling](#), in which neurotransmitters secreted by [nerve cell](#) axons act locally on the postsynaptic cells that the axons contact.

Cell **signaling** requires not only extracellular **signal** molecules, but also a [complementary](#) set of [receptor](#) proteins in each cell that enable it to bind and respond to the **signal** molecules in a characteristic way. Some small hydrophobic **signal** molecules, including [steroid](#) and thyroid

hormones, diffuse across the [plasma membrane](#) of the target cell and activate intracellular receptor proteins that directly regulate the transcription of specific genes. The dissolved gases [nitric oxide](#) and carbon monoxide act as local mediators by diffusing across the plasma membrane of the target cell and activating an intracellular [enzyme](#)—usually guanylyl cyclase, which produces [cyclic GMP](#) in the target cell. But most extracellular **signal** molecules are [hydrophilic](#) and can activate receptor proteins only on the surface of the target cell; these receptors act as **signal** transducers, converting the extracellular binding event into intracellular signals that alter the behavior of the target cell.

There are three main families of cell-surface receptors, each of which transduces extracellular signals in a different way. Ion-channel-linked receptors are transmitter-gated [ion](#) channels that open or close briefly in response to the binding of an [neurotransmitter](#). [G-protein](#)-linked receptors indirectly activate or inactivate plasma-[membrane](#)-bound enzymes or ion channels via trimeric GTP-binding proteins (G proteins). Enzyme-linked receptors either act directly as enzymes or are associated with enzymes; these enzymes are usually protein kinases that phosphorylate specific proteins in the target cell.

Once activated, [enzyme](#)- and [G-protein](#)-linked receptors relay a **signal** into the cell interior by activating chains of intracellular **signaling** proteins; some transduce, amplify, or spread the **signal** as they relay it, while others integrate signals from different **signaling** pathways. Many of these **signaling** proteins function as switches that are transiently activated by [phosphorylation](#) or GTP binding. Functional **signaling** complexes are often formed by means of modular binding domains in the **signaling** proteins; these domains allow complicated protein assemblies to function in **signaling** networks.

Target cells can use a variety of intracellular mechanisms to respond abruptly to a gradually increasing concentration of an extracellular **signal** or to convert a short-lasting **signal** into a long-lasting response. In addition, through [adaptation](#), they can often reversibly adjust their sensitivity to a **signal** to allow the cells to respond to changes in the concentration of a particular [signal molecule](#) over a large range of concentrations.

Signaling through G-Protein-Linked Cell-Surface Receptors

[G-protein-linked receptors](#) form the largest family of cell-surface receptors and are found in all eucaryotes. About 5% of the genes in the nematode *C. elegans*, for example, encode such receptors, and thousands have already been defined in mammals; in mice, there are about 1000 concerned with the sense of smell alone. [G-protein](#)-linked receptors mediate the responses to an enormous diversity of signal molecules, including hormones, neurotransmitters, and local mediators. These signal molecules that activate them are as varied in structure as they are in function: the list includes proteins and small peptides, as well as derivatives of amino acids and fatty acids. The same [ligand](#) can activate many different [receptor](#) family members; at least 9 distinct G-protein-linked receptors are activated by adrenaline, for example, another 5 or more by [acetylcholine](#), and at least 15 by the [neurotransmitter](#) serotonin.

Despite the chemical and functional diversity of the signal molecules that bind to them, all [G-protein](#)-linked receptors have a similar structure. They consist of a single [polypeptide](#) chain that threads back and forth across the [lipid bilayer](#) seven times and are therefore sometimes called *serpentine receptors*. In addition to their characteristic orientation in the [plasma membrane](#), they have the same functional relationship to the G proteins they use to signal the cell interior that an extracellular [ligand](#) is present.

As we discuss later, this superfamily of seven-pass transmembrane proteins includes [rhodopsin](#), the light-activated [protein](#) in the vertebrate eye, as well as the large number of olfactory receptors in the vertebrate nose. Other family members are found in unicellular organisms: the receptors in yeasts that recognize secreted mating factors are an example. In fact, it is thought that the [G-protein](#)-linked receptors that mediate cell-cell signaling in multicellular organisms evolved from sensory receptors that were possessed by their unicellular eucaryotic ancestors.

It is remarkable that about half of all known drugs work through [G-protein](#)-linked receptors. Genome sequencing projects are revealing vast numbers of new family members, many of which are likely targets for new drugs that remain to be discovered.

Trimeric G Proteins Disassemble to Relay Signals from G-Protein-linked Receptors

When extracellular signaling molecules bind to serpentine receptors, the receptors undergo a conformational change that enables them to activate **trimeric GTP-binding proteins (G proteins)**. These G proteins are attached to the cytoplasmic face of the [plasma membrane](#), where they serve as relay molecules, functionally coupling the receptors to enzymes or [ion](#) channels in this membrane. There are various types of G proteins, each specific for a particular set of serpentine receptors and for a particular set of downstream target proteins in the plasma membrane. All have a similar structure, however, and they operate in a similar way.

[G](#) proteins are composed of three [protein](#) subunits— α , β , and γ . In the unstimulated state, the α [subunit](#) has GDP bound and the G protein is inactive. When stimulated by an activated [receptor](#), the α subunit releases its bound GDP, allowing GTP to bind in its place. This exchange causes the trimer to dissociate into two activated components—an α [subunit](#) and a $\beta\gamma$ [complex](#).

The dissociation of the trimeric [G protein](#) activates its two components in different ways. GTP binding causes a conformational change that affects the surface of the α [subunit](#) that associates with the $\beta\gamma$ [complex](#) in the trimer. This change causes the release of the $\beta\gamma$ complex, but it also causes and the α subunit to adopt a new shape that allows it to interact with its target proteins. The $\beta\gamma$ complex does not change its [conformation](#), but the surface previously masked by the α subunit is now available to interact with a second set of target proteins. The targets of the dissociated components of the G protein are either enzymes or [ion](#) channels in the [plasma membrane](#), and they relay the signal onward.

The α [subunit](#) is a [GTPase](#), and once it hydrolyzes its bound GTP to GDP, it reassociates with a $\beta\gamma$ [complex](#) to re-form an inactive [G protein](#), reversing the activation process. The time during which the α subunit and $\beta\gamma$ complex remain apart and active is usually short, and it depends on how quickly the α subunit hydrolyzes its bound GTP. An isolated α subunit is an inefficient GTPase, and, left to its own devices, the subunit would inactivate only after several minutes. Its activation is usually reversed much faster than this, however, because the GTPase activity of the α subunit is greatly enhanced by the binding of a second protein, which can be either its target protein or a specific modulator known as a *regulator of G protein signaling (RGS)*. **RGS proteins** act as α -subunit-specific *GTPase activating proteins (GAPs)*, and they are thought to have a crucial role in shutting off [G-protein](#)-mediated responses in all eucaryotes. There are about 25 RGS proteins encoded in the human [genome](#), each of which is thought to interact with a particular set of G proteins.

The importance of the [GTPase](#) activity in shutting off the response can be easily demonstrated in a test tube. If cells are broken open and exposed to an analogue of GTP (GTP γ S) in which the terminal phosphate cannot be hydrolyzed, the activated α subunits remain active for a very long time.

Some G Proteins Signal By Regulating the Production of Cyclic AMP

Cyclic AMP (cAMP) was first identified as a [small intracellular mediator](#) in the 1950s. It has since been found to act in this role in all procaryotic and animal cells that have been studied. The normal concentration of cyclic AMP inside the cell is about 10^{-7} M, but an extracellular signal can cause cyclic AMP levels to change by more than twentyfold in seconds ([Figure 15-30](#)). As explained earlier (see [Figure 15-10](#)), such a rapid response requires that a rapid synthesis of the [molecule](#) be balanced by its rapid breakdown or removal. In fact, cyclic AMP is synthesized from ATP by a plasma-[membrane](#)-bound [enzyme adenylyl cyclase](#), and it is rapidly and continuously destroyed by one or more **cyclic AMP phosphodiesterases** that hydrolyze cyclic AMP to adenosine 5'-monophosphate (5'-AMP).

Many extracellular signal molecules work by increasing cyclic AMP content, and they do so by increasing the activity of adenylyl cyclase rather than decreasing the activity of phosphodiesterase. Adenylyl cyclase is a large [multipass transmembrane protein](#) with its catalytic [domain](#) on the cytosolic side of the [plasma membrane](#). There are at least eight isoforms in mammals, most of which are regulated by both [G](#) proteins and Ca^{2+} . All receptors that act via cyclic AMP are coupled to a **stimulatory G protein (G_s)**, which activates adenylyl cyclase and thereby increases cyclic AMP concentration. Another G protein, called **inhibitory G protein (G_i)**, inhibits adenylyl cyclase, but it mainly acts by directly regulating [ion](#) channels (as we discuss later) rather than by decreasing cyclic AMP content. Although it is usually the α [subunit](#) that regulates the cyclase, the $\beta\gamma$ [complex](#) sometimes does so as well, either increasing or decreasing the [enzyme](#)'s activity, depending on the particular $\beta\gamma$ complex and the isoform of the cyclase.

Both G_s and G_i are targets for some medically important bacterial toxins. *Cholera toxin*, which is produced by the bacterium that causes cholera, is an [enzyme](#) that catalyzes the transfer of ADP ribose from intracellular NAD^+ to the α [subunit](#) of G_s . This ADP ribosylation alters the α subunit so that it can no longer hydrolyze its bound GTP, causing it to remain in an active state that stimulates adenylyl cyclase indefinitely. The resulting prolonged elevation in cyclic AMP levels within intestinal epithelial cells causes a large efflux of Cl^- and water into the gut, thereby causing the severe diarrhea that characterizes cholera. *Pertussis toxin*, which is made by the bacterium that causes pertussis (whooping cough), catalyzes the ADP ribosylation of the α subunit of G_i , preventing the subunit from interacting with receptors; as a result, this α subunit retains its bound GDP and is unable to regulate its target proteins. These two toxins are widely used as tools to determine whether a cell's response to a signal is mediated by G_s or by G_i .

Individuals who are genetically deficient in a particular G_s α [subunit](#) show decreased responses to certain hormones. As a consequence, they display metabolic abnormalities, have abnormal bone [development](#), and are mentally retarded.

Cyclic-AMP-dependent Protein Kinase (PKA) Mediates Most of the Effects of Cyclic AMP

Although cyclic AMP can directly activate certain types of [ion](#) channels in the [plasma membrane](#) of some highly specialized cells, in most animal cells it exerts its effects mainly by activating **cyclic-AMP-dependent protein kinase(PKA)**. This [enzyme](#) catalyzes the transfer of the terminal phosphate group from ATP to specific serines or threonines of selected target proteins, thereby regulating their activity.

[PKA](#) is found in all animal cells and is thought to account for the effects of cyclic AMP in most of these cells. The substrates for PKA differ in different cell types, which explains why the effects of cyclic AMP vary so markedly depending on the cell type.

In the inactive state, [PKA](#) consists of a [complex](#) of two catalytic subunits and two regulatory subunits. The binding of cyclic AMP to the regulatory subunits alters their [conformation](#), causing them to dissociate from the complex. The released catalytic subunits are thereby activated to phosphorylate specific [substrate protein](#) molecules. The regulatory subunits of PKA also are important for localizing the kinase inside the cell: special *PKA anchoring proteins* bind both to the regulatory subunits and to a [membrane](#) or a component of the [cytoskeleton](#), thereby tethering the [enzyme](#) complex to a particular subcellular [compartment](#). Some of these anchoring proteins also bind other kinases and some phosphatases, creating a signaling complex.

Some responses mediated by cyclic AMP are rapid while others are slow. In skeletal muscle cells, for example, activated [PKA](#) phosphorylates enzymes involved in [glycogen metabolism](#), which simultaneously triggers the breakdown of glycogen to [glucose](#) and inhibits glycogen synthesis, thereby increasing the amount of glucose available to the muscle cell within seconds (see also [Figure 15-30](#)). At the other extreme are responses that take hours to develop fully and involve changes in the transcription of specific genes. In cells that secrete the peptide [hormone](#) *somatostatin*, for example, cyclic AMP activates the [gene](#) that encodes this hormone. The regulatory region of the somatostatin gene contains a short [DNA](#) sequence, called the *cyclic AMP response element (CRE)*, that is also found in the regulatory region of many other genes activated by cyclic AMP. A specific [gene regulatory protein](#) called **CRE-binding (CREB) protein** recognizes this sequence. When CREB is phosphorylated by PKA on a single serine, it recruits a transcriptional coactivator called *CREB-binding protein (CBP)*, which stimulates the transcription of these genes. If this serine is mutated, CREB cannot recruit CBP, and it no longer stimulates gene transcription in response to a rise in cyclic AMP levels.

Protein Phosphatases Make the Effects of PKA and Other Protein Kinases Transitory

Since the effects of cyclic AMP are usually transient, cells must be able to dephosphorylate the proteins that have been phosphorylated by [PKA](#). Indeed, the activity of any [protein](#) regulated by [phosphorylation](#) depends on the balance at any instant between the activities of the kinases that phosphorylate it and the phosphatases that are constantly dephosphorylating it. In general, the dephosphorylation of phosphorylated serines and

threonines is catalyzed by four types of **serine/threonine phosphoprotein phosphatases**—protein phosphatases I, IIA, IIB, and IIC. Except for [protein phosphatase-IIC](#) (which is a minor phosphatase, unrelated to the others), all of these phosphatases are composed of a [homologous](#) catalytic [subunit](#) complexed with one or more of a large set of regulatory subunits; the regulatory subunits help to control the phosphatase activity and enable the [enzyme](#) to select specific targets. *Protein phosphatase I* is responsible for dephosphorylating many of the proteins phosphorylated by PKA. It inactivates CREB, for example, by removing its activating phosphate, thereby turning off the transcriptional response caused by a rise in cyclic AMP concentration. *Protein phosphatase IIA* has a broad specificity and seems to be the main phosphatase responsible for reversing many of the phosphorylations catalyzed by serine/threonine kinases. *Protein phosphatase IIB*, also called *calcineurin*, is activated by Ca^{2+} and is especially abundant in the brain.

Having discussed how trimeric [G](#) proteins link activated receptors to adenylyl cyclase, we now consider how they couple activated receptors to another crucial [enzyme](#), *phospholipase C*. The activation of this enzyme leads to an increase in the concentration of Ca^{2+} in the [cytosol](#), which helps to relay the signal onward. Ca^{2+} is even more widely used as an intracellular mediator than is cyclic AMP.

Some G Proteins Activate the Inositol Phospholipid Signaling Pathway by Activating Phospholipase C- β

Many [G-protein](#)-linked receptors exert their effects mainly via G proteins that activate the plasma-[membrane](#)-bound [enzyme](#) **phospholipase C- β** . The phospholipase acts on an inositol [phospholipid](#) (a *phosphoinositide*) called [phosphatidylinositol 4,5-bisphosphate](#) [**PI(4,5)P₂**], which is present in small amounts in the inner half of the [plasma membrane lipid bilayer](#). Receptors that operate through this **inositol phospholipid signaling pathway** mainly activate a G protein called [G_q](#), which in turn activates phospholipase C- β , in much the same way that [G_s](#) activates adenylyl cyclase. The activated phospholipase cleaves PI(4,5)P₂ to generate two products: *inositol 1,4,5-trisphosphate* and [diacylglycerol](#). At this step, the signaling pathway splits into two branches.

Inositol 1,4,5-trisphosphate (IP₃) is a small, water-soluble [molecule](#) that leaves the [plasma membrane](#) and diffuses rapidly through the [cytosol](#). When it reaches the [endoplasmic reticulum \(ER\)](#), it binds to and opens IP₃-gated Ca^{2+} -release channels in the ER membrane. Ca^{2+} stored in the ER is released through the open channels, quickly raising the concentration of Ca^{2+} in the cytosol. We discuss later how Ca^{2+} acts to propagate the signal. Several mechanisms operate to terminate the initial Ca^{2+} response: (1) IP₃ is rapidly dephosphorylated by specific phosphatases to form IP₂; (2) IP₃ is phosphorylated to IP₄ (which may function as another intracellular mediator); and (3) Ca^{2+} that enters the cytosol is rapidly pumped out, mainly to the exterior of the cell.

At the same time that the IP₃ produced by the hydrolysis of PI(4,5)P₂ is increasing the concentration of Ca^{2+} in the [cytosol](#), the other [cleavage](#) product of PI(4,5)P₂—

[diacylglycerol](#)—is exerting different effects. Diacylglycerol remains embedded in the [membrane](#), where it has two potential signaling roles. First, it can be further cleaved to release arachidonic [acid](#), which can either act as a messenger in its own right or be used in the synthesis of other small [lipid](#) messengers called *eicosanoids*. Eicosanoids, such as the *prostaglandins*, are made by most vertebrate cell types and have a wide variety of biological activities. They participate in pain and inflammatory responses, for example, and most anti-inflammatory drugs (such as aspirin, ibuprofen, and cortisone) act—in part, at least—by inhibiting their synthesis.

The second, and more important, function of [diacylglycerol](#) is to activate a crucial serine/threonine [protein kinase](#) called [protein kinase C \(PKC\)](#), so named because it is Ca^{2+} -dependent. The initial rise in cytosolic Ca^{2+} induced by IP_3 alters the PKC so that it translocates from the [cytosol](#) to the cytoplasmic face of the [plasma membrane](#). There it is activated by the combination of Ca^{2+} , diacylglycerol, and the negatively charged membrane [phospholipid](#) phosphatidylserine. Once activated, PKC phosphorylates target proteins that vary depending on the cell type. The principles are the same as discussed earlier for [PKA](#), although most of the target proteins are different.

Each of the two branches of the inositol [phospholipid](#) signaling pathway can be mimicked by the addition of specific pharmacological agents to intact cells. The effects of IP_3 can be mimicked by using a Ca^{2+} [ionophore](#), such as [A23187](#) or ionomycin, which allows Ca^{2+} to move into the [cytosol](#) from the extracellular fluid (discussed in Chapter 11). The effects of [diacylglycerol](#) can be mimicked by *phorbol esters*, plant products that bind to [PKC](#) and activate it directly. Using these reagents, it has been shown that the two branches of the pathway often collaborate in producing a full cellular response. Some cell types, such as lymphocytes, for example, can be stimulated to proliferate in culture when treated with both a Ca^{2+} ionophore and a PKC activator, but not when they are treated with either reagent alone.

[Ca²⁺ Functions as a Ubiquitous Intracellular Messenger](#)

Many extracellular signals induce an increase in cytosolic Ca^{2+} level, not just those that work via [G](#) proteins. In [egg](#) cells, for example, a sudden rise in cytosolic Ca^{2+} concentration upon [fertilization](#) by a sperm triggers a Ca^{2+} wave that is responsible for the onset of embryonic [development](#) ([Figure 15-37](#)). In muscle cells, Ca^{2+} triggers contraction, and in many secretory cells, including nerve cells, it triggers secretion. Ca^{2+} can be used as a signal in this way because its concentration in the [cytosol](#) is normally kept very low ($\sim 10^{-7}$ M), whereas its concentration in the extracellular fluid ($\sim 10^{-3}$ M) and in the [ER lumen](#) is high. Thus, there is a large gradient tending to drive Ca^{2+} into the cytosol across both the [plasma membrane](#) and the ER membrane. When a signal transiently opens Ca^{2+} channels in either of these membranes, Ca^{2+} rushes into the cytosol, increasing the local Ca^{2+} concentration by 10–20-fold and triggering Ca^{2+} -responsive proteins in the cell.

Three main types of Ca^{2+} channels can mediate this Ca^{2+} signaling:

- 1.

Voltage-dependent Ca^{2+} channels in the [plasma membrane](#) open in response to membrane depolarization and allow, for example, Ca^{2+} to enter activated nerve terminals and trigger [neurotransmitter](#) secretion.

2.

IP_3 -gated Ca^{2+} -release channels allow Ca^{2+} to escape from the [ER](#) when the inositol [phospholipid](#) signaling pathway is activated, as just discussed (see [Figure 15-36](#)).

3.

Ryanodine receptors (so called because they are sensitive to the plant [alkaloid](#) ryanodine) react to a change in [plasma membrane](#) potential to release Ca^{2+} from the [sarcoplasmic reticulum](#) and thereby stimulate the contraction of muscle cells; they are also present in the [ER](#) of many nonmuscle cells, including neurons, where they can contribute to Ca^{2+} signaling.

The concentration of Ca^{2+} in the [cytosol](#) is kept low in resting cells by several mechanisms ([Figure 15-38](#)). Most notably, all eucaryotic cells have a Ca^{2+} -[pump](#) in their [plasma membrane](#) that uses the energy of ATP hydrolysis to pump Ca^{2+} out of the cytosol. Cells such as muscle and nerve cells, which make extensive use of Ca^{2+} signaling, have an additional Ca^{2+} transport [protein](#) (exchanger) in their plasma membrane that couples the efflux of Ca^{2+} to the influx of Na^+ . A Ca^{2+} pump in the [ER](#) membrane also has an important role in keeping the cytosolic Ca^{2+} concentration low: this Ca^{2+} -pump enables the ER to take up large amounts of Ca^{2+} from the cytosol against a steep concentration gradient, even when Ca^{2+} levels in the cytosol are low. In addition, a low-affinity, high-capacity Ca^{2+} pump in the inner mitochondrial membrane has an important role in returning the Ca^{2+} concentration to normal after a Ca^{2+} signal; it uses the [electrochemical gradient](#) generated across this membrane during the [electron-transfer](#) steps of [oxidative phosphorylation](#) to take up Ca^{2+} from the cytosol.

The Frequency of Ca^{2+} Oscillations Influences a Cell's Response

Ca^{2+} -sensitive fluorescent indicators, such as aequorin or fura-2 (discussed in Chapter 9), are often used to monitor cytosolic Ca^{2+} in individual cells after the inositol [phospholipid](#) signaling pathway has been activated. When viewed in this way, the initial Ca^{2+} signal is often seen to be small and localized to one or more discrete regions of the cell. These signals have been called Ca^{2+} blips, quarks, puffs, or sparks, and they are thought to reflect the local opening of individual (or small groups of) Ca^{2+} -release channels in the [ER](#) and to represent elementary Ca^{2+} signaling units. If the extracellular signal is sufficiently strong and persistent, this localized signal can propagate as a regenerative Ca^{2+} wave through the [cytosol](#), much like an [action potential](#) in an [axon](#). Such a Ca^{2+} “spike” is often followed by a series of further spikes, each usually lasting seconds. These Ca^{2+} oscillations can persist for as long as receptors are activated at the cell surface. Both the

waves and the oscillations are thought to depend, in part at least, on a combination of positive and negative feedback by Ca^{2+} on both the IP_3 -gated Ca^{2+} -release channels and the ryanodine receptors: the released Ca^{2+} initially stimulates more Ca^{2+} release, a process known as Ca^{2+} -induced Ca^{2+} release. But then, as its concentration gets high enough, the Ca^{2+} inhibits further release.

The frequency of the Ca^{2+} oscillations reflects the strength of the extracellular stimulus, and it can be translated into a frequency-dependent cell response. In some cases, the frequency-dependent response itself is also oscillatory. In [hormone](#)-secreting pituitary cells, for example, stimulation by an extracellular signal induces repeated Ca^{2+} spikes, each of which is associated with a burst of hormone secretion. The frequency-dependent response can also be nonoscillatory. In some types of cells, for instance, one frequency of Ca^{2+} spikes activates the transcription of one set of genes, while a higher frequency activates the transcription of a different set. How do cells sense the frequency of Ca^{2+} spikes and change their response accordingly? The mechanism presumably depends on Ca^{2+} -sensitive proteins that change their activity as a function of Ca^{2+} spike frequency. A [protein kinase](#) that acts as a molecular memory device seems to have this remarkable property, as we discuss next.

Ca^{2+} /Calmodulin-dependent Protein Kinases (CaM-Kinases) Mediate Many of the Actions of Ca^{2+} in Animal Cells

Ca^{2+} -binding proteins serve as transducers of the cytosolic Ca^{2+} signal. The first such [protein](#) to be discovered was *troponin C* in skeletal muscle cells; its role in muscle contraction is discussed in Chapter 16. A closely related Ca^{2+} -binding protein, known as [calmodulin](#), is found in all eucaryotic cells, where it can constitute as much as 1% of the total protein mass. Calmodulin functions as a multipurpose intracellular Ca^{2+} [receptor](#), mediating many Ca^{2+} -regulated processes. It consists of a highly conserved, single [polypeptide](#) chain with four high-affinity Ca^{2+} -binding sites. When activated by binding Ca^{2+} , it undergoes a conformational change. Because two or more Ca^{2+} ions must bind before [calmodulin](#) adopts its active [conformation](#), the protein responds in a switchlike manner to increasing concentrations of Ca^{2+} : a tenfold increase in Ca^{2+} concentration, for example, typically causes a fiftyfold increase in calmodulin activation.

The allosteric activation of [calmodulin](#) by Ca^{2+} is analogous to the allosteric activation of [PKA](#) by cyclic AMP, except that Ca^{2+} /calmodulin has no enzymic activity itself but instead acts by binding to other proteins. In some cases, calmodulin serves as a permanent regulatory [subunit](#) of an [enzyme complex](#), but mostly the binding of Ca^{2+} enables calmodulin to bind to various target proteins in the cell to alter their activity.

When an activated [molecule](#) of Ca^{2+} /calmodulin binds to its target [protein](#), it undergoes a marked change in [conformation](#) ([Figure 15-40B](#)). Among the targets regulated by calmodulin binding are many enzymes and [membrane transport](#) proteins. As one example, Ca^{2+} /calmodulin binds to and activates the [plasma membrane](#) Ca^{2+} -[pump](#) that pumps Ca^{2+} out

of cells. Thus, whenever the concentration of Ca^{2+} in the [cytosol](#) rises, the pump is activated, which helps to return the cytosolic Ca^{2+} level to normal.

Many effects of Ca^{2+} , however, are more indirect and are mediated by [protein](#) phosphorylations catalyzed by a family of Ca^{2+} /[calmodulin](#)-dependent **protein kinases (CaM-kinases)**. These kinases, just like [PKA](#) and [PKC](#), phosphorylate serines or threonines in proteins, and, as with PKA and PKC, the response of a target cell depends on which CaM-kinase-regulated target proteins are present in the cell. The first CaM-kinases to be discovered—*myosin light-chain kinase*, which activates smooth muscle contraction, and *phosphorylase kinase*, which activates [glycogen](#) breakdown—have narrow [substrate](#) specificities. A number of CaM-kinases, however, have much broader specificities, and these seem to be responsible for mediating many of the actions of Ca^{2+} in animal cells. Some phosphorylate [gene](#) regulatory proteins, such as the CREB protein discussed earlier, and in this way activate or inhibit the transcription of specific genes.

The best-studied example of such a *multifunctional CaM-kinase* is [CaM-kinase II](#), which is found in all animal cells but is especially enriched in the nervous system. It constitutes up to 2% of the total [protein](#) mass in some regions of the brain, and it is highly concentrated in synapses. [CaM-kinase II](#) has at least two remarkable properties that are related. First, it can function as a molecular memory device, switching to an active state when exposed to Ca^{2+} /[calmodulin](#) and then remaining active even after the Ca^{2+} signal has decayed. This is because the kinase phosphorylates itself (a process called *autophosphorylation*) as well as other cell proteins when it is activated by Ca^{2+} /calmodulin. In its autophosphorylated state, the [enzyme](#) remains active even in the absence of Ca^{2+} , thereby prolonging the duration of the kinase activity beyond that of the initial activating Ca^{2+} signal. This activity is maintained until phosphatases overwhelm the autophosphorylating activity of the enzyme and shut it off ([Figure 15-41](#)). CaM-kinase II activation can thereby serve as a memory trace of a prior Ca^{2+} pulse, and it seems to have an important role in some types of memory and learning in the vertebrate nervous system. A [point mutation](#) in CaM-kinase II that removes its autophosphorylation site, but otherwise leaves the kinase activity intact, produces the same learning defect, revealing that the autophosphorylation is critical in these animals.

The second remarkable property of [CaM-kinase II](#) is that it can use its memory mechanism to act as a frequency decoder of Ca^{2+} oscillations. This property is thought to be especially important at a [nerve cell synapse](#), where changes in intracellular Ca^{2+} levels in an activated postsynaptic cell can lead to long-term changes in the subsequent effectiveness of that synapse (discussed in Chapter 11). When CaM-kinase II is immobilized on a solid surface and exposed to both a [protein phosphatase](#) and repetitive pulses of Ca^{2+} /[calmodulin](#) at different frequencies that mimic those observed in stimulated cells, the [enzyme](#)'s activity increases steeply as a function of pulse frequency. Moreover, the frequency response of this multisubunit enzyme depends on its exact [subunit](#) composition, so that a cell can tailor its response to Ca^{2+} oscillations to particular needs by adjusting the composition of the CaM-kinase II enzyme that it makes.

Some G Proteins Directly Regulate Ion Channels

G proteins do not act exclusively by regulating the activity of [membrane](#)-bound enzymes that alter the concentration of cyclic AMP or Ca^{2+} in the [cytosol](#). The α [subunit](#) of one type of G [protein](#) (called G_{12}), for example, activates a protein that converts a monomeric [GTPase](#) of the Rho family (discussed in Chapter 16) into its active form, which then alters the [actin cytoskeleton](#). In some other cases, G proteins directly activate or inactivate [ion](#) channels in the [plasma membrane](#) of the target cell, thereby altering the ion permeability—and hence the excitability of the membrane. Acetylcholine released by the vagus nerve, for example, reduces both the rate and strength of heart muscle cell contraction (see [Figure 15-9A](#)). A special class of [acetylcholine](#) receptors that activate the G_i protein discussed earlier mediates this effect. Once activated, the α subunit of G_i inhibits adenylyl cyclase (as described previously), while the $\beta\gamma$ [complex](#) binds to K^+ channels in the heart muscle cell plasma membrane to open them. The opening of these K^+ channels makes it harder to depolarize the cell, which contributes to the inhibitory effect of acetylcholine on the heart. (These acetylcholine receptors, which can be activated by the fungal [alkaloid](#) muscarine, are called *muscarinic acetylcholine receptors* to distinguish them from the very different *nicotinic acetylcholine receptors*, which are ion-channel-linked receptors on skeletal muscle and nerve cells that can be activated by the binding of nicotine, as well as by acetylcholine.)

Other trimeric G proteins regulate the activity of [ion](#) channels less directly, either by stimulating channel [phosphorylation](#) (by [PKA](#), [PKC](#), or CaM-kinase, for example) or by causing the production or destruction of cyclic nucleotides that directly activate or inactivate ion channels. The *cyclic-nucleotide-gated ion channels* have a crucial role in both smell (olfaction) and vision, as we now discuss.

Smell and Vision Depend on G-Protein-linked Receptors That Regulate Cyclic-Nucleotide-gated Ion Channels

Humans can distinguish more than 10,000 distinct smells, which are detected by specialized olfactory [receptor](#) neurons in the lining of the nose. These cells recognize odors by means of specific [G-protein](#)-linked **olfactory receptors**, which are displayed on the surface of the modified cilia that extend from each cell ([Figure 15-43](#)). The receptors act through cyclic AMP. When stimulated by odorant binding, they activate an olfactory-specific G protein (known as G_{olf}), which in turn activates adenylyl cyclase. The resulting increase in cyclic AMP opens *cyclic-AMP-gated cation channels*, thereby allowing an influx of Na^+ , which depolarizes the olfactory receptor neuron and initiates a nerve impulse that travels along its [axon](#) to the brain.

There are about 1000 different olfactory receptors in a mouse, each encoded by a different [gene](#) and each recognizing a different set of odorants. All of these receptors belong to the [G-protein-linked receptor](#) superfamily. Each olfactory receptor neuron produces only one of these 1000 receptors, and the neuron responds to a specific set of odorants by means of

the specific receptor it displays. The same receptor also has a crucial role in directing the elongating [axon](#) of each developing olfactory neuron to the specific target neurons that it will connect to in the brain. A different set of more than 100 G-protein-linked receptors acts in a similar way to mediate a mouse's responses to *pheromones*, chemical signals detected in a different part of the nose that are used in communication between members of the same species.

Vertebrate vision involves a similarly elaborate, highly sensitive, signal-detection process. Cyclic-[nucleotide](#)-gated [ion](#) channels are also involved, but the crucial cyclic nucleotide is [cyclic GMP](#) rather than cyclic AMP. As with cyclic AMP, a continuous rapid synthesis (by *guanylyl cyclase*) and rapid degradation (by [cyclic GMP phosphodiesterase](#)) controls the concentration of cyclic GMP in cells.

In visual transduction responses, which are the fastest [G-protein](#)-mediated responses known in vertebrates, the [receptor](#) activation caused by light leads to a fall rather than a rise in the level of the cyclic [nucleotide](#). The pathway has been especially well studied in **rod photoreceptors (rods)** in the vertebrate retina. Rods are responsible for noncolor vision in dim light, whereas *cone photoreceptors (cones)* are responsible for color vision in bright light. A rod [photoreceptor](#) is a highly specialized cell with outer and inner segments, a [cell body](#), and a synaptic region where the rod passes a chemical signal to a retinal [nerve cell](#); this nerve cell in turn relays the signal along the visual pathway. The phototransduction apparatus is in the outer segment, which contains a stack of *discs*, each formed by a closed sac of [membrane](#) in which many photosensitive [rhodopsin](#) molecules are embedded. The [plasma membrane](#) surrounding the outer segment contains *cyclic-GMP-gated Na⁺ channels*. These channels are kept open in the dark by [cyclic GMP](#) that has bound to them. Paradoxically, light causes a hyperpolarization (which inhibits [synaptic signaling](#)) rather than a depolarization of the plasma membrane (which could stimulate synaptic signaling). Hyperpolarization (an increase in the [membrane potential](#)—discussed in Chapter 11) results because the activation by light of [rhodopsin](#) molecules in the disc membrane leads to a fall in cyclic GMP concentration and the *closure* of the special Na⁺ channels in the surrounding plasma membrane.

Rhodopsin is a seven-pass transmembrane [molecule](#) [homologous](#) to other members of the [G-protein-linked receptor](#) family, and, like its cousins, it acts through a trimeric G protein. The activating extracellular signal, however, is not a molecule but a [photon](#) of light.

Each [rhodopsin](#) molecule contains a covalently attached chromophore, 11-*cis* retinal, which isomerizes almost instantaneously to all-*trans* retinal when it absorbs a single photon. The isomerization alters the shape of the retinal, forcing a conformational change in the protein (opsin). The activated rhodopsin molecule then alters the G-protein *transducin* (G_t), causing its α [subunit](#) to dissociate and activate [cyclic GMP phosphodiesterase](#). The phosphodiesterase then hydrolyzes cyclic GMP, so that cyclic GMP levels in the [cytosol](#) fall.

This drop in cyclic GMP concentration leads to a decrease in the amount of cyclic GMP bound to the [plasma membrane](#) Na^+ channels, allowing more of these highly cyclic-GMP-sensitive channels to close. In this way, the signal quickly passes from the disc membrane to the plasma membrane, and a light signal is converted into an electrical one.

A number of mechanisms operate in rods to allow the cells to revert quickly to a resting, dark state in the aftermath of a flash of light—a requirement for perceiving the shortness of the flash. A [rhodopsin-specific kinase \(RK\)](#) phosphorylates the cytosolic tail of activated rhodopsin on multiple serines, partially inhibiting the ability of the rhodopsin to activate transducin. An inhibitory [protein](#) called *arrestin* then binds to the phosphorylated rhodopsin, further inhibiting rhodopsin's activity. If the [gene](#) encoding RK is inactivated by [mutation](#) in mice or humans, the light response of rods is greatly prolonged, and the rods eventually die.

At the same time as [rhodopsin](#) is being shut off, an RGS [protein](#) binds to activated transducin, stimulating the transducin to hydrolyze its bound GTP to GDP, which returns transducin to its inactive state. In addition, the Na^+ channels that close in response to light are also permeable to Ca^{2+} , so that when they close, the normal influx of Ca^{2+} is inhibited, causing the Ca^{2+} concentration in the [cytosol](#) to fall. The decrease in Ca^{2+} concentration stimulates guanylyl cyclase to replenish the [cyclic GMP](#), rapidly returning its level to where it was before the light was switched on. A specific Ca^{2+} -sensitive protein mediates the activation of guanylyl cyclase in response to a fall in Ca^{2+} levels. In contrast to [calmodulin](#), this protein is inactive when Ca^{2+} is bound to it and active when it is Ca^{2+} -free. It therefore stimulates the cyclase when Ca^{2+} levels fall following a light response.

These shut-off mechanisms do more than just return the rod to its resting state after a light flash; they also help to enable the [photoreceptor](#) to *adapt*, stepping down the response when it is exposed to light continuously. Adaptation, as we discussed earlier, allows the [receptor](#) cell to function as a sensitive detector of *changes* in stimulus intensity over an enormously wide range of baseline levels of stimulation.

Extracellular Signals Are Greatly Amplified by the Use of Small Intracellular Mediators and Enzymatic Cascades

Despite the differences in molecular details, the signaling systems that are triggered by [G-protein](#)-linked receptors share certain features and are governed by similar general principles. They depend on relay chains of intracellular signaling proteins and small intracellular mediators. In contrast to the more direct signaling pathways used by nuclear receptors discussed earlier, and by [ion](#)-channel-linked receptors discussed in Chapter 11, these relay chains provide numerous opportunities for amplifying the responses to extracellular signals. In the visual transduction cascade just described, for example, a single activated [rhodopsin molecule](#) catalyzes the activation of hundreds of molecules of transducin at a rate of about 1000 transducin molecules per second. Each activated transducin molecule activates a molecule of [cyclic GMP](#) phosphodiesterase, each of which hydrolyzes about 4000 molecules of cyclic GMP per second. This catalytic cascade lasts for about 1 second and

results in the hydrolysis of more than 10^5 cyclic GMP molecules for a single quantum of light absorbed, and the resulting drop in the concentration of cyclic GMP in turn transiently closes hundreds of Na^{2+} channels in the [plasma membrane](#) ([Figure 15-47](#)). As a result, a rod cell can respond to a single [photon](#) of light, in a way that is highly reproducible in its timing and magnitude.

Likewise, when an extracellular [signal molecule](#) binds to a [receptor](#) that indirectly activates adenylyl cyclase via G_s , each receptor [protein](#) may activate many molecules of G_s protein, each of which can activate a cyclase molecule. Each cyclase molecule, in turn, can catalyze the conversion of a large number of ATP molecules to cyclic AMP molecules. A similar amplification operates in the inositol-[phospholipid](#) pathway. A nanomolar (10^{-9} M) change in the concentration of an extracellular signal can thereby induce micromolar (10^{-6} M) changes in the concentration of a [small intracellular mediator](#) such as cyclic AMP or Ca^{2+} . Because these mediators function as allosteric effectors to activate specific enzymes or [ion](#) channels, a single extracellular signal molecule can cause many thousands of protein molecules to be altered within the target cell.

Any such amplifying cascade of stimulatory signals requires that there be counterbalancing mechanisms at every step of the cascade to restore the system to its resting state when stimulation ceases. Cells therefore have efficient mechanisms for rapidly degrading (and resynthesizing) cyclic nucleotides and for buffering and removing cytosolic Ca^{2+} , as well as for inactivating the responding enzymes and [ion](#) channels once they have been activated. This is not only essential for turning a response off, it is also important for defining the resting state from which a response begins. As we saw earlier, in general, the response to stimulation can be rapid only if the inactivating mechanisms are also rapid. Each [protein](#) in the relay chain of signals can be a separate target for regulation, including the [receptor](#), as we discuss next.

G-Protein-linked Receptor Desensitization Depends on Receptor Phosphorylation

As discussed earlier, target cells use a variety of mechanisms to *desensitize*, or *adapt*, when they are exposed to a high concentration of stimulating [ligand](#) for a prolonged period (see [Figure 15-25](#)). We discuss here only those mechanisms that involve an alteration in [G-protein](#)-linked receptors themselves.

These receptors can desensitize in three general ways:

1.

They can become altered so that they can no longer interact with G proteins ([receptor inactivation](#)).

2.

They can be temporarily moved to the interior of the cell (internalized) so that they no longer have access to their ligand (receptor sequestration).

3.

They can be destroyed in lysosomes after internalization (receptor down-regulation).

In each case, the desensitization process depends on phosphorylation of the receptor, by PKA, PKC, or a member of the family of **G-protein-linked receptor kinases (GRKs)**. (The GRKs include the rhodopsin-specific kinase involved in rod photoreceptor desensitization discussed earlier.) The GRKs phosphorylate multiple serine and threonines on a receptor, but they do so only after the receptor has been activated by ligand binding. As with rhodopsin, once a receptor has been phosphorylated in this way, it binds with high affinity to a member of the **arrestin** family of proteins.

The bound arrestin can contribute to the desensitization process in at least two ways. First, it inactivates the receptor by preventing it from interacting with G proteins, an example of receptor uncoupling. Second, it can serve as an adaptor protein to couple the receptor to clathrin-coated pits (discussed in Chapter 13), inducing receptor-mediated endocytosis. Endocytosis results in either the sequestration or degradation (down-regulation) of the receptor, depending on the specific receptor and cell type, the concentration of the stimulating ligand, and the duration of the ligand's presence.

Summary

G-protein-linked receptors can indirectly activate or inactivate either plasma-membrane-bound enzymes or ion channels via G proteins. When stimulated by an activated receptor, a G protein disassembles into an α subunit and a $\beta\gamma$ complex, both of which can directly regulate the activity of target proteins in the plasma membrane. Some G-protein-linked receptors either activate or inactivate adenylyl cyclase, thereby altering the intracellular concentration of the intracellular mediator cyclic AMP. Others activate a phosphoinositide-specific phospholipase C (phospholipase C- β), which hydrolyzes phosphatidylinositol 4,5-bisphosphate [PI(4,5)P₂] to generate two small intracellular mediators. One is inositol 1,4,5-trisphosphate (IP₃), which releases Ca²⁺ from the ER and thereby increases the concentration of Ca²⁺ in the cytosol. The other is diacylglycerol, which remains in the plasma membrane and activates protein kinase C (PKC). A rise in cyclic AMP or Ca²⁺ levels affects cells mainly by stimulating protein kinase A (PKA) and Ca²⁺/calmodulin-dependent protein kinases (CaM-kinases), respectively.

PKC, PKA, and CaM-kinases phosphorylate specific target proteins on serines or threonines and thereby alter the activity of the proteins. Each type of cell has characteristic sets of target proteins that are regulated in these ways, enabling the cell to make its own distinctive response to the small intracellular mediators. The intracellular signaling cascades activated by G-protein-linked receptors allow the responses to be greatly amplified, so that many target proteins are changed for each molecule of extracellular signaling ligand bound to its receptor.

The responses mediated by [G-protein](#)-linked receptors are rapidly turned off when the extracellular signaling [ligand](#) is removed. Thus, the G-protein α [subunit](#) is induced to inactivate itself by hydrolyzing its bound GTP to GDP, IP_3 is rapidly dephosphorylated by a [phosphatase](#) (or phosphorylated by a kinase), cyclic nucleotides are hydrolyzed by phosphodiesterases, Ca^{2+} is rapidly pumped out of the [cytosol](#), and phosphorylated proteins are dephosphorylated by protein phosphatases. Activated G-protein-linked receptors themselves are phosphorylated by GRKs, thereby triggering arrestin binding, which uncouples the receptors from G proteins and promotes [receptor endocytosis](#).

Signaling through Enzyme-Linked Cell-Surface Receptors

Enzyme-linked receptors are a second major type of cell-surface receptor. They were recognized initially through their role in responses to extracellular signal proteins that promote the growth, proliferation, differentiation, or survival of cells in animal tissues. These signal proteins are often collectively called *growth factors*, and they usually act as local mediators at very low concentrations (about 10^{-9} - 10^{-11} M). The responses to them are typically slow (on the order of hours) and usually require many intracellular signaling steps that eventually lead to changes in gene expression. Enzyme-linked receptors have since been found also to mediate direct, rapid effects on the cytoskeleton, controlling the way a cell moves and changes its shape. The extracellular signals that induce these rapid responses are often not diffusible but are instead attached to surfaces over which the cell is crawling. Disorders of cell proliferation, differentiation, survival, and migration are fundamental events that can give rise to cancer, and abnormalities of signaling through enzyme-linked receptors have major roles in this class of disease.

Like G-protein-linked receptors, enzyme-linked receptors are transmembrane proteins with their ligand-binding domain on the outer surface of the plasma membrane. Instead of having a cytosolic domain that associates with a trimeric G protein, however, their cytosolic domain either has an intrinsic enzyme activity or associates directly with an enzyme. Whereas a G-protein-linked receptor has seven transmembrane segments,

each subunit of an enzyme-linked receptor usually has only one.

Six classes of enzyme-linked receptors have thus far been identified:

1.

Receptor tyrosine kinases phosphorylate specific tyrosines on a small set of intracellular signaling proteins.

2.

Tyrosine-kinase-associated receptors associate with intracellular proteins that have tyrosine kinase activity.

3.

Receptorlike tyrosine phosphatases remove phosphate groups from tyrosines of specific intracellular signaling proteins. (They are called “receptorlike” because the presumptive ligands have not yet been identified, and so their receptor function has not been directly demonstrated.)

4.

Receptor serine/threonine kinases phosphorylate specific serines or threonines on associated latent gene regulatory proteins.

5.

Receptor guanylylcyclases directly catalyze the production of cyclic GMP in the cytosol.

6.

Histidine-kinase-associated receptors activate a “two-component” signaling pathway in which the

kinase phosphorylates itself on histidine and then immediately transfers the phosphate to a second [intracellular signaling protein](#).

We begin our discussion with the [receptor](#) tyrosine kinases, the most numerous of the [enzyme](#)-linked receptors. We then consider the other classes in turn.

Activated Receptor Tyrosine Kinases Phosphorylate Themselves

The extracellular signal proteins that act through [receptor](#) **tyrosine kinases** consist of a large variety of secreted growth factors and hormones. Notable examples discussed elsewhere in this book include *epidermal* [growth factor](#) (EGF), *platelet*-derived growth factor (PDGF), [fibroblast](#) growth factors (FGFs), [hepatocyte](#) growth factor (HGF), [insulin](#), *insulinlike growth factor-1* (IGF-1), *vascular endothelial growth factor* (VEGF), [macrophage-colony-stimulating factor](#) (M-CSF), and all the *neurotrophins*, including *nerve growth factor* (NGF).

Many cell-surface-bound signal proteins also act through these receptors. The largest class of these [membrane](#)-bound ligands is the **ephrins**, which regulate the cell adhesion and repulsion responses that guide the migration of cells and axons along specific pathways during animal [development](#). The receptors for ephrins, called **Eph receptors**, are also the most numerous [receptor](#) tyrosine kinases. The ephrins and Eph receptors are unusual in that they can simultaneously act as both [ligand](#) and receptor: on binding to an Eph receptor, some ephrins not only activate the Eph receptor but also become activated themselves to

transmit signals into the interior of the ephrin-expressing cell. In this way, an interaction between an ephrin [protein](#) on one cell and an Eph protein on another cell can lead to bidirectional reciprocal signaling that changes the behavior of both cells. Such *bidirectional signaling* between ephrins and Eph receptors is required, for example, to keep cells in particular parts of the developing brain from mixing with cells in neighboring parts.

Receptor tyrosine kinases can be classified into more than 16 structural subfamilies, each dedicated to its [complementary](#) family of [protein](#) ligands. In all cases, the binding of a signal protein to the [ligand-binding domain](#) on the outside of the cell activates the intracellular tyrosine kinase domain. Once activated, the kinase domain transfers a phosphate group from ATP to selected tyrosine side chains, both on the [receptor](#) proteins themselves and on intracellular signaling proteins that subsequently bind to the phosphorylated receptors.

How does the binding of an extracellular [ligand](#) activate the kinase [domain](#) on the other side of the [plasma membrane](#)? For a [G-protein-linked receptor](#), ligand binding is thought to change the relative orientation of several of the transmembrane α helices, thereby shifting the position of the cytoplasmic loops relative to each other. It is difficult to imagine, however, how a conformational change could propagate across the [lipid bilayer](#) through a single transmembrane [\$\alpha\$ helix](#). Instead, for the [enzyme-linked receptors](#), two or more receptor chains come together in the membrane, forming a dimer or higher [oligomer](#). In some cases, ligand binding induces the

oligomerization. In other cases, the oligomerization occurs before ligand binding, and the ligand causes a reorientation of the receptor chains in the membrane. In either case, the rearrangement induced in cytosolic tails of the receptors initiates the intracellular signaling process. For receptor tyrosine kinases, the rearrangement enables the neighboring kinase domains of the receptor chains to cross-phosphorylate each other on multiple tyrosines, a process referred to as *autophosphorylation*.

To activate a [receptor](#) tyrosine kinase the [ligand](#) usually has to bind simultaneously to two adjacent receptor chains. PDGF, for example, is a dimer, which cross-links two receptors together. Even some monomeric ligands, such as EGF, bind to two receptors simultaneously and cross-link them directly. By contrast, FGFs, which are also monomers, first form multimers by binding to heparansulfate proteoglycans, either on the target cell surface or in the [extracellular matrix](#). In this way, they are able to cross-link adjacent receptors. In [contact-dependent signaling](#), the ligands form clusters in the [plasma membrane](#) of the signaling cell and can thereby cross-link the receptors on the target cell; thus, whereas membrane-bound ephrins activate Eph receptors, soluble ephrins will do so only if they are aggregated.

Because of the requirement for [receptor](#) oligomerization, it is relatively easy to inactivate a specific receptor tyrosine kinase to determine its importance for a cell response. For this purpose, cells are transfected with [DNA](#) encoding a [mutant](#) form of the receptor that oligomerizes normally but has an inactive kinase [domain](#). When coexpressed at a high level with normal receptors, the mutant receptor acts

in a *dominant-negative* way, disabling the normal receptors by forming inactive dimers with them.

Autophosphorylation of the cytosolic tail of receptor tyrosine kinases contributes to the activation process in two ways. First, phosphorylation of tyrosines within the kinase domain increases the kinase activity of the enzyme. Second, phosphorylation of tyrosines outside the kinase domain creates high-affinity docking sites for the binding of a number of intracellular signaling proteins in the target cell. Each type of signaling protein binds to a different phosphorylated site on the activated receptor because it contains a specific phosphotyrosine-binding domain that recognizes surrounding features of the polypeptide chain in addition to the phosphotyrosine. Once bound to the activated kinase, the signaling protein may itself become phosphorylated on tyrosines and thereby activated; alternatively, the binding alone may be sufficient to activate the docked signaling protein. In summary, autophosphorylation serves as a switch to trigger the transient assembly of a large intracellular signaling complex, which then broadcasts signals along multiple routes to many destinations in the cell. Because different receptor tyrosine kinases bind different combinations of these signaling proteins, they activate different responses.

The receptors for insulin and IGF-1 act in a slightly different way. They are tetramers to start with, and ligand binding is thought to induce a rearrangement of the transmembrane receptor chains, so that the two kinase domains come close together. Most of the phosphotyrosine docking sites generated by ligand binding are not on the

receptor itself, but on a specialized docking [protein](#) called *insulin receptor [substrate](#)-1 (IRS-1)*. The activated receptor first autophosphorylates its kinase domains, which then phosphorylate IRS-1 on multiple tyrosines, thereby creating many more docking sites than could be accommodated on the receptor alone. Other docking proteins are used in a similar way by some other receptor tyrosine kinases to enlarge the size of the signaling [complex](#).

Phosphorylated Tyrosines Serve as Docking Sites For Proteins With SH2 Domains

A whole menagerie of intracellular signaling proteins can bind to the phosphotyrosines on activated [receptor](#) tyrosine kinases (or on special docking proteins such as IRS-1) to help to relay the signal onward. Some docked proteins are enzymes, such as **phospholipase C- γ (PLC- γ)**, which functions in the same way as phospholipase C- β —activating the inositol [phospholipid](#) signaling pathway discussed earlier in connection with [G-protein](#)-linked receptors. Through this pathway, receptor tyrosine kinases can increase cytosolic Ca^{2+} levels. Much more often, these receptors depend more on relay chains of protein-protein interactions. For example, another [enzyme](#) that docks on these receptors is the cytoplasmic tyrosine kinase Src, which phosphorylates other signaling proteins on tyrosines. Yet another is [phosphatidylinositol 3'-kinase \(PI 3-kinase\)](#), which, as we discuss later, generates specific [lipid](#) molecules in the [plasma membrane](#) to attract other signaling proteins there.

Although the intracellular signaling proteins that bind to phosphotyrosines on activated [receptor](#) tyrosine kinases and docking proteins have varied structures and functions, they usually share highly conserved phosphotyrosine-binding domains. These can be either [SH2 domains](#) (for *Src* *homology region*, because it was first found in the *Src* [protein](#)) or, less commonly, *PTB domains* (for *phosphotyrosine-binding*). By recognizing specific phosphorylated tyrosines, these small domains serve as modules that enable the proteins that contain them to bind to activated receptor tyrosine kinases, as well as to many other intracellular signaling proteins that have been transiently phosphorylated on tyrosines. Many signaling proteins also contain other protein modules that allow them to interact specifically with other proteins as part of the signaling process. These include the *SH3* [domain](#) (again, so named because it was first discovered in *Src*), which binds to proline-rich motifs in intracellular proteins.

Not all proteins that bind to activated [receptor](#) tyrosine kinases via SH2 domains help to relay the signal onward. Some act to decrease the signaling process, providing negative feedback. One example is the *c-Cbl* [protein](#), which can dock on some activated receptors and catalyze their conjugation with [ubiquitin](#). This ubiquitylation promotes the internalization and degradation of the receptors—a process called receptor down-regulation.

Some signaling proteins are composed almost entirely of SH2 and SH3 domains and function as adaptors to couple tyrosine-phosphorylated proteins to other proteins that do not have their own SH2 domains. Such adaptor proteins

help to couple activated receptors to the important downstream signaling [protein](#) **Ras**. As we discuss next, **Ras** acts as a transducer and bifurcation signaling protein, changing the nature of the signal and broadcasting it along multiple downstream pathways, including a major signaling pathway that can help stimulate cells to proliferate or differentiate. Mutations that activate this pathway, and thereby stimulate [cell division](#) inappropriately, are a causative factor in many types of cancer.

Ras Is Activated by a Guanine Nucleotide Exchange Factor

The [Ras proteins](#) belong to the large **Ras superfamily of monomeric GTPases**. The family also contains two other subfamilies: the *Rho family*, involved in relaying signals from cell-surface receptors to the [actin cytoskeleton](#) and elsewhere (discussed in Chapter 16), and the *Rab family*, involved in regulating the traffic of intracellular transport vesicles. Like almost all of these monomeric GTPases, the **Ras** proteins contain a covalently attached [lipid](#) group that helps to anchor the [protein](#) to a [membrane](#)—in this case, to the cytoplasmic face of the [plasma membrane](#) where the protein functions. There are multiple **Ras** proteins, and different ones act in different cell types. Because they all seem to work in much the same way, we shall refer to them simply as **Ras**.

Ras helps to broadcast signals from the cell surface to other parts of the cell. It is often required, for example, when [receptor](#) tyrosine kinases signal to the [nucleus](#) to stimulate cell proliferation or [differentiation](#) by

altering [gene expression](#). If **Ras** function is inhibited by the [microinjection](#) of neutralizing anti-**Ras** antibodies or a [dominant-negative mutant](#) form of **Ras**, the cell proliferation or differentiation responses normally induced by the activated receptor tyrosine kinases do not occur. Conversely, if a hyperactive mutant [Ras protein](#) is introduced into some cell lines, the effect on cell proliferation or differentiation is sometimes the same as that induced by the binding of ligands to cell-surface receptors. In fact, **Ras** was first discovered as the hyperactive product of a mutant *ras* gene that promoted the [development](#) of cancer; we now know that about 30% of human tumors have a hyperactive *ras* [mutation](#).

Like other GTP-binding proteins, **Ras** functions as a switch, cycling between two distinct conformational states—active when GTP is bound and inactive when GDP is bound. Two classes of signaling proteins regulate **Ras** activity by influencing its transition between active and inactive states. **Guanine nucleotide exchange factors (GEFs)** promote the exchange of bound nucleotide by stimulating the dissociation of GDP and the subsequent uptake of GTP from the [cytosol](#), thereby activating **Ras**. **GTPase-activating proteins (GAPs)** increase the rate of hydrolysis of bound GTP by **Ras**, thereby inactivating **Ras**.

Hyperactive [mutant](#) forms of **Ras** are resistant to [GAP](#)-mediated GTPase stimulation and are locked permanently in the GTP-bound active state, which is why they promote the [development](#) of cancer.

In principle, [receptor](#) tyrosine kinases could activate **Ras** either by activating a [GEF](#) or by inhibiting

a [GAP](#). Even though some GAPs bind directly (via their SH2 domains) to activated receptor tyrosine kinases, whereas GEFs bind only indirectly, it is the indirect coupling of the receptor to a GEF that is responsible for driving **Ras** into its active state. In fact, the loss of function of a **Ras**-specific GEF has a similar effect to the loss of function of that **Ras**. The activation of the other **Ras**-like proteins, including those of the Rho family, is also thought to occur through the activation of GEFs.

Genetic studies in flies and worms, and biochemical studies in mammalian cells, indicate that adaptor proteins link [receptor](#) tyrosine kinases to **Ras**. The **Grb-2 protein** in mammalian cells, for example, binds through its [SH2 domain](#) to specific phosphotyrosines on activated receptor tyrosine kinases and through its SH3 domains to proline-rich motifs on a [GEF](#) called **Sos**. Some activated receptor tyrosine kinases, however, do not display the specific phosphotyrosines required for Grb-2 docking; these receptors recruit another [adaptor protein](#) called *Shc*, which binds both to the activated receptor and to Grb-2, thereby coupling the receptor to Sos by a more indirect route. The assembly of the [complex](#) of receptor-Grb-2-Sos (or receptor-Shc-Grb-2-Sos) brings Sos into position to activate neighboring **Ras** molecules by stimulating it to exchange its bound GDP for GTP. The importance of Grb-2 is indicated by the finding that Grb-2-deficient mice die early in [embryogenesis](#). Very similar sets of proteins are thought to operate in all animals to activate **Ras**.

This pathway from [receptor](#) tyrosine kinases is not the only means of activating **Ras**. Other **Ras** GEFs are activated independently of Sos. One that is found mainly in the

brain, for example, is activated by Ca^{2+} and [diacylglycerol](#) and can couple [G-protein](#)-linked receptors to **Ras** activation.

Once activated, **Ras** in turn activates various other signaling proteins to relay the signal downstream along several pathways. One of the signaling pathways **Ras** activates is a serine/threonine [phosphorylation](#) cascade that is highly conserved in eucaryotic cells from yeasts to humans. As we discuss next, a crucial component in this cascade is a novel type of [protein kinase](#) called [MAP-kinase](#).

Ras Activates a Downstream Serine/Threonine Phosphorylation Cascade That Includes a MAP-Kinase

Both the tyrosine phosphorylations and the activation of **Ras** triggered by activated [receptor](#) tyrosine kinases are short-lived. Tyrosine-specific [protein](#) phosphatases (discussed later) quickly reverse the phosphorylations, and GAPs induce activated **Ras** to inactivate itself by hydrolyzing its bound GTP to GDP. To stimulate cells to proliferate or differentiate, these short-lived signaling events must be converted into longer-lasting ones that can sustain the signal and relay it downstream to the [nucleus](#) to alter the pattern of [gene expression](#). Activated **Ras** triggers this conversion by initiating a series of downstream serine/threonine phosphorylations, which are much longer-lived than tyrosine phosphorylations. Many serine/threonine kinases participate in this [phosphorylation](#) cascade, but three of them constitute the core [module](#) of the cascade. The last of the three is

called a **mitogen-activated protein kinase (MAP-kinase)**.

An unusual feature of a **MAP**-kinase is that its full activation requires the **phosphorylation** of both a threonine and a tyrosine, which are separated in the **protein** by a single **amino acid**. The **protein kinase** that catalyzes both of these phosphorylations is called a *MAP-kinase-kinase*, which in the mammalian **Ras** signaling pathway is called MEK. The requirement for both a tyrosine and a threonine phosphorylation ensures that the MAP-kinase is kept inactive unless specifically activated by a MAP-kinase-kinase, whose only known **substrate** is a MAP-kinase. MAP-kinase-kinase is itself activated by phosphorylation catalyzed by the first kinase in the three-component **module**, *MAP-kinase-kinase-kinase*, which in the mammalian **Ras** signaling pathway is called **Raf**. The Raf kinase is activated by activated **Ras**.

Once activated, the **MAP**-kinase relays the signal downstream by phosphorylating various proteins in the cell, including **gene** regulatory proteins and other **protein** kinases. It enters the **nucleus**, for example, and phosphorylates one or more components of a gene regulatory **complex**. This activates the transcription of a set of *immediate early genes*, so named because they turn on within minutes of the time that cells are stimulated by an extracellular signal, even if protein synthesis is experimentally blocked with drugs. Some of these genes encode other gene regulatory proteins that turn on other genes, a process that requires both protein synthesis and more time. In this way the **Ras**-MAP-kinase signaling pathway conveys signals from the cell surface to the

nucleus and alters the pattern of gene [expression](#) in significant ways. Among the genes activated by this pathway are those required for cell proliferation, such as the genes encoding [G₁ cyclins](#).

[MAP](#)-kinases are usually activated only transiently in response to extracellular signals, and the period of time they remain active can profoundly influence the nature of the response. When EGF activates its receptors on a neural precursor [cell line](#), for example, MAP-kinase activity peaks at 5 minutes and rapidly declines, and the cells later go on to divide. By contrast, when NGF activates its receptors on the same cells, MAP-kinase activity remains high for many hours, and the cells stop proliferating and differentiate into neurons.

[MAP](#)-kinases are inactivated by dephosphorylation, and the specific removal of phosphate from either the tyrosine or the threonine is enough to inactivate the [enzyme](#). In some cases, stimulation by an extracellular signal induces the [expression](#) of a dual-specificity [phosphatase](#) that removes both phosphates and inactivates the kinase, providing a form of negative feedback. In other cases, stimulation causes the kinase to be switched off more rapidly by phosphatases that are already present.

Three-component [MAP](#)-kinase signaling modules operate in all animal cells, as well as in yeasts, with different ones mediating different responses in the same cell. In [budding yeast](#), for example, one such [module](#) mediates the mating pheromone response via the $\beta\gamma$ [complex](#) of a [G protein](#), another the response to starvation, and yet another the response to osmotic shock. Some of these three-component

MAP-kinase modules use one or more of the same kinases and yet manage to activate different effector proteins and hence different responses.

Mammalian cells also use this strategy to prevent cross talk between MAP-kinase signaling pathways. At least 5 parallel MAP-kinase modules can operate in a mammalian cell. These modules are composed of at least 12 MAP-kinases, 7 MAP-kinase-kinases, and 7 MAP-kinase-kinase-kinases. Several of these modules are activated by different kinds of cell stresses, such as UV irradiation, heat shock, osmotic stress, and stimulation by inflammatory cytokines. The three kinases in at least some of these stress-activated modules are held together by binding to a common scaffold protein, just as in yeast. The scaffold strategy provides precision, helps to create a large change in MAP-kinase activity in response to small changes in signal molecule concentration, and avoids cross-talk. However, it reduces the opportunities for amplification and spreading of the signal to different parts of the cell, which require at least some of the components to be diffusible.

When **Ras** is activated by receptor tyrosine kinases, it usually activates more than just the MAP-kinase signaling pathway. It also usually helps activate *PI3-kinase*, which can signal cells to survive and grow.

PI 3-Kinase Produces Inositol Phospholipid Docking Sites in the Plasma Membrane

Extracellular signal proteins stimulate cells to divide, in part by activating the **Ras-MAP**-kinase pathway just discussed. If cells continually divided without growing, however, they would get progressively smaller and would

eventually disappear. Thus, to proliferate, most cells need to be stimulated to enlarge (grow), as well as to divide. In some cases, one signal protein does both; in others one signal protein (a mitogen) mainly stimulates cell division, while another (a growth factor) mainly stimulates cell growth. One of the major intracellular signaling pathways leading to cell growth involves phosphatidylinositol 3-kinase (PI 3-kinase). This kinase principally phosphorylates inositol phospholipids rather than proteins; it can be activated by receptor tyrosine kinases, as well as by many other types of cell-surface receptors, including some that are G-protein-linked.

Phosphatidylinositol (PI) is unique among membrane lipids because it can undergo reversible phosphorylation at multiple sites to generate a variety of distinct inositol phospholipids. When activated, PI 3-kinase catalyzes the phosphorylation of inositol phospholipids at the 3 position of the inositol ring to generate lipids called $PI(3,4)P_2$ or $PI(3,4,5)P_3$. The $PI(3,4)P_2$ and $PI(3,4,5)P_3$ then serve as docking sites for intracellular signaling proteins, bringing these proteins together into signaling complexes, which relay the signal into the cell from the cytosolic face of the plasma membrane.

It is important to distinguish this use of inositol phospholipids from their use we discussed earlier. We considered earlier how $PI(4,5)P_2$ is cleaved by PLC- β (in the case of G-protein-linked receptors) or PLC- γ (in the case of receptor tyrosine kinases) to generate soluble IP_3 and membrane-bound diacylglycerol. The IP_3 releases Ca^{2+} from the ER, while the diacylglycerol activates PKC.

By contrast, $\text{PI}(3,4)\text{P}_2$ and $\text{PI}(3,4,5)\text{P}_3$ are not cleaved by PLC. They remain in the [plasma membrane](#) until they are dephosphorylated by specific *inositol* [phospholipid](#)phosphatases that remove phosphate from the 3 position of the inositol ring. Mutations that inactivate one such [phosphatase](#) (called PTEN), and thereby prolong signaling by PI 3-kinase, promote the [development](#) of cancer, and they are found in many human cancers. The mutations result in prolonged cell survival, indicating that signaling through PI 3-kinase normally promotes cell survival, as well as cell growth.

There are various types of PI 3-kinases. The one that is activated by [receptor](#) tyrosine kinases consists of a catalytic and regulatory [subunit](#). The regulatory subunit is an [adaptor protein](#) that binds to phosphotyrosines on activated receptor tyrosine kinases through its SH2 domains. Another PI 3-kinase has a different regulatory subunit and is activated by the $\beta\gamma$ [complex](#) of a trimeric [G](#) protein when [G-protein](#)-linked receptors are activated by their extracellular [ligand](#). The catalytic subunit, which is similar in both cases, also has a [binding site](#) for activated **Ras**, which allows **Ras** to directly stimulate PI 3-kinases.

Intracellular signaling proteins bind to the $\text{PI}(3,4)\text{P}_2$ and $\text{PI}(3,4,5)\text{P}_3$ that are produced by activated PI 3-kinase mainly through their **Pleckstrin homology (PH) [domain](#)**, first identified in the [platelet protein](#) Pleckstrin. PH domains are found in about 200 human proteins, including Sos (the [GEF](#) discussed earlier that activates **Ras**), and some atypical PKCs that do not depend on Ca^{2+} for their activation. The importance of these domains is illustrated

dramatically by certain genetic immunodeficiency diseases in both humans and mice, where the [PH domain](#) in a cytoplasmic tyrosine kinase called **BTK** is inactivated by [mutation](#). Normally, when [antigen](#) receptors on B lymphocytes (B cells) activate PI 3-kinase, the resulting inositol [lipid](#) docking sites recruit both BTK and [PLC- \$\gamma\$](#) to the cytoplasmic face of the [plasma membrane](#). There, the two proteins interact: BTK phosphorylates and activates PLC- γ , which then cleaves PI(4,5)P₂ to generate IP₃ and [diacylglycerol](#) to relay the signal onward. Because the [mutant](#) BTK cannot bind to the lipid docking sites produced after [receptor](#) activation, the receptors cannot signal the B cells to proliferate or survive, resulting in a severe deficiency in antibody production.

The PI 3-Kinase/Protein Kinase B Signaling Pathway Can Stimulate Cells to Survive and Grow

One way in which PI 3-kinase signals cells to survive is by indirectly activating [protein kinase B \(PKB\)](#) (also called [Akt](#)). This kinase contains a [PH domain](#), which directs it to the [plasma membrane](#) when PI 3-kinase is activated there by an extracellular survival signal. After binding to PI(3,4,5)P₃ on the cytosolic face of the membrane, the PKB alters its [conformation](#) so that it can now be activated in a process that requires [phosphorylation](#) by a [phosphatidylinositol](#)-dependent protein kinase called **PDK1**, which is recruited to the membrane in the same way. Once activated, the PKB returns to the [cytoplasm](#) and phosphorylates a variety of target proteins. One of these, called *BAD*, is a protein that normally encourages cells to undergo [programmed](#)

[cell death](#), or [apoptosis](#) (mentioned earlier and discussed in detail in Chapter 17). By phosphorylating BAD, PKB inactivates it, thereby promoting cell survival. PKB also promotes cell survival by inhibiting other cell death activators, in some cases by inhibiting the transcription of the genes that encode them.

The pathways by which PI 3-kinase signals cells to grow (and increase their [metabolism](#) generally) are [complex](#) and still poorly understood. One way in which growth factors stimulate cell growth is by increasing the rate of [protein](#) synthesis through enhancing the efficiency with which ribosomes translate certain mRNAs into protein.

A [protein kinase](#) called **S6 kinase** is part of one of the signaling pathways from PI 3-kinase to the [ribosome](#). It phosphorylates and thereby activates the S6 [subunit](#) of ribosomes, which helps to increase the translation of a subset of mRNAs that encode ribosomal proteins and other components of the translational apparatus. The activation of S6 kinase is itself a complex process that depends on PDK1 and the [phosphorylation](#) of many sites on the protein. PDK1 may phosphorylate one of these sites in response to PI 3-kinase activation.

Tyrosine-Kinase-associated Receptors Depend on Cytoplasmic Tyrosine Kinases for Their Activity

Many cell-surface receptors depend on tyrosine [phosphorylation](#) for their activity and yet lack an obvious tyrosine kinase [domain](#). These receptors act through **cytoplasmic tyrosine kinases**, which are associated with the receptors and phosphorylate various

target proteins, often including the receptors themselves, when the receptors bind their [ligand](#). The receptors thus function in much the same way as [receptor](#) tyrosine kinases, except that their kinase domain is encoded by a separate [gene](#) and is noncovalently associated with the receptor [polypeptide](#) chain. As with receptor tyrosine kinases, these receptors must oligomerize to function.

Many of these receptors depend on members of the largest family of mammalian cytoplasmic tyrosine kinases, the [Src family](#) of [protein](#) kinases. This family includes the following members: *Src*, *Yes*, *Fgr*, *Fyn*, *Lck*, *Lyn*, *Hck*, and *Blk*. These protein kinases all contain SH2 and SH3 domains and are located on the cytoplasmic side of the [plasma membrane](#), held there partly by their interaction with transmembrane [receptor](#) proteins and partly by covalently attached [lipid](#) chains. Different family members are associated with different receptors and phosphorylate overlapping but distinct sets of target proteins. Lyn, Fyn, and Lck, for example, are each associated with different sets of receptors in lymphocytes. In each case the kinase is activated when an extracellular [ligand](#) binds to the appropriate receptor protein. Src itself, as well as several other family members, can also bind to activated receptor tyrosine kinases; in these cases, the receptor and cytoplasmic kinases mutually stimulate each other's catalytic activity, thereby strengthening and prolonging the signal.

Another type of cytoplasmic tyrosine kinase associates with *integrins*, the main family of receptors that cells use to bind to the [extracellular matrix](#) (discussed in Chapter 19). The binding of matrix components to integrins can

activate intracellular signaling pathways that influence the behavior of the cell. When integrins cluster at sites of matrix contact, they help trigger the assembly of cell-matrix junctions called *focal adhesions*. Among the many proteins recruited into these junctions is the cytoplasmic tyrosine kinase called **focal adhesion kinase (FAK)**, which binds to the cytosolic tail of one of the **integrin** subunits with the assistance of other cytoskeletal **protein**. The clustered FAK molecules cross-phosphorylate each other, creating phosphotyrosine docking sites where the Src kinase can bind. Src and FAK now phosphorylate each other and other proteins that assemble in the junction, including many of the signaling proteins used by **receptor** tyrosine kinases. In this way, the two kinases signal to the cell that it has adhered to a suitable **substratum**, where the cell can now survive, grow, divide, migrate, and so on. Mice deficient in FAK die early in **development**, and their cells do not migrate normally in a culture dish.

Cytokine receptors are the subfamily of **enzyme**-linked receptors that we discuss next. They constitute the largest and most diverse class of receptors that rely on cytoplasmic kinases to relay signals into the cell. They include receptors for many kinds of local mediators (collectively called *cytokines*), as well as receptors for some hormones, such as growth**hormone** (see **Figure 15-62**) and prolactin. As we discuss next, these receptors are stably associated with a class of cytoplasmic tyrosine kinases called *Jaks*, which activate latent **gene** regulatory proteins called *STATs*. The STAT proteins are normally inactive, being located at the cell surface; **cytokine** or

hormone binding causes them to migrate to the [nucleus](#) and activate gene transcription.

Cytokine Receptors Activate the Jak-STAT Signaling Pathway, Providing a Fast Track to the Nucleus

Many intracellular signaling pathways lead from cell-surface receptors to the [nucleus](#), where they alter [gene](#)transcription. The [Jak-STAT signaling pathway](#), however, provides one of the most direct routes. It was initially discovered in studies on the effects of *interferons*, which are cytokines secreted by cells (especially white blood cells) in response to viral infection. Interferons bind to receptors on noninfectedneighboringcells and induce the cells to produce proteins that increase their resistance to viral infection. When activated, interferon receptors activate a novel class of cytoplasmic tyrosine kinases called **Janus kinases (Jaks)** (after the two-faced Roman god). The Jaks then phosphorylate and activate a set of latent gene regulatory proteins called **STATs** (signal transducers and activators oftranscription), which move into the nucleus and stimulate the transcription of specific genes. More than 30 cytokines and hormones activate the Jak-STAT pathway by binding to [cytokine](#) receptors.

All STATs also have an [SH2 domain](#) that enables them to dock onto specific phosphotyrosines on some activated[receptor](#) tyrosine kinase receptors. These receptors can directly activate the bound STAT, independently of Jaks. In fact, the nematode *C. elegans* uses STATs for signaling but does not make any

Jaks or [cytokine](#) receptors, suggesting that STATs evolved before Jaks and cytokine receptors.

[Cytokine receptors](#) are composed of two or more [polypeptide](#) chains. Some [cytokine receptor](#) chains are specific to a particular cytokine receptor, while others are shared among several such receptors. All cytokine receptors, however, are associated with one or more Jaks. There are four known Jaks—Jak1, Jak2, Jak3, and Tyk2—and each is associated with particular cytokine receptors. The receptors for α -interferon, for example, are associated with Jak1 and Tyk2, whereas the receptors for γ -interferon are associated with Jak1 and Jak2. As expected, mice that lack Jak1 do not respond to either of these interferons. The receptor for the [hormone erythropoietin](#), which stimulates erythrocyte precursor cells to survive, proliferate, and differentiate, is associated with only Jak2. In Jak2-deficient mice, erythrocyte [development](#) fails, and the mice die early in development.

Cytokine binding either induces the [receptor](#) chains to oligomerize or reorients the chains in a preformed [oligomer](#). In either case, the binding brings the associated Jaks close enough together for them to cross-phosphorylate each other, thereby increasing the activity of their tyrosine kinase domains. The Jaks then phosphorylate tyrosines on the [cytokine](#) receptors, creating phosphotyrosine docking sites for STATs and other signaling proteins.

There are seven known STATs, each with an [SH2 domain](#) that performs two functions. First, it mediates the binding of the STAT [protein](#) to a phosphotyrosine docking

site on an activated [cytokine receptor](#) (or receptor tyrosine kinase); once bound, the Jaks phosphorylate the STAT on tyrosines, causing it to dissociate from the receptor. Second, the SH2 domain on the released STAT now mediates its binding to a phosphotyrosine on another STAT [molecule](#), forming either a STAT homodimer or [heterodimer](#). The STAT dimer then moves into the [nucleus](#), where, in combination with other [gene](#) regulatory proteins, it binds to a specific [DNA](#) response element in various genes and stimulates their transcription. In response to the [hormone](#) prolactin, for example, which stimulates breast cells to produce milk, activated STAT5 stimulates the transcription of genes that encode milk proteins.

Cytokine receptors activate the appropriate STAT proteins because the [SH2 domain](#) of these STATs recognizes only the specific phosphotyrosine docking sites on these receptors. Activated receptors for α -interferon, for example, recruit both STAT1 and STAT2, whereas activated receptors for γ -interferon recruit only STAT1. If the SH2 domain of the α -interferon [receptor](#) is replaced with the SH2 domain of the γ -interferon receptor, the activated hybrid receptor recruits both STAT1 and STAT2, just like the α -interferon receptor itself.

The responses mediated by STATs are often regulated by negative feedback. In addition to activating genes that encode proteins mediating the [cytokine](#)-induced response, the STAT dimers may also activate genes that encode inhibitory proteins. In some cases, the inhibitor binds to both the activated cytokine receptors and STAT proteins, which blocks further STAT activation and helps to shut off

the response; in other cases, the inhibitor achieves the same result by blocking Jak function.

Such negative feedback mechanisms, however, are not enough on their own to turn off the response. The activated Jaks and STATs also have to be inactivated by dephosphorylation of their phosphotyrosines. As in all signaling pathways that use tyrosine [phosphorylation](#), the dephosphorylation is performed by [protein tyrosine phosphatases](#), which are as important in the signaling process as the protein tyrosine kinases that add the phosphates.

Some Protein Tyrosine Phosphatases May Act as Cell-Surface Receptors

As discussed earlier, only a small number of serine/threonine [phosphatase](#) catalytic subunits are responsible for removing phosphate groups from phosphorylated serines and threonines on proteins. By contrast, there are about 30 [protein tyrosine phosphatases \(PTPs\)](#) encoded in the human [genome](#). Like tyrosine kinases, they occur in both cytoplasmic and transmembrane forms, none of which are structurally related to serine/threonine protein phosphatases. Individual protein tyrosine phosphatases display exquisite specificity for their substrates, removing phosphate groups from only selected phosphotyrosines on a subset of tyrosine-phosphorylated proteins. Together, these phosphatases ensure that tyrosine phosphorylations are short-lived and that the level of tyrosine [phosphorylation](#) in resting cells is very low. They do not, however, simply continuously reverse the effects of protein tyrosine kinases; they are

regulated to act only at the appropriate time in a signaling response or in the cell-division cycle.

Two cytoplasmic tyrosine phosphatases in vertebrates have SH2 domains and are therefore called **SHP-1** and **SHP-2**. SHP-1 helps to terminate some [cytokine](#) responses in blood cells by dephosphorylating activated Jaks: [mutant erythropoietin](#) receptors that cannot recruit SHP-1, for example, activate Jak2 for much longer than normal. Moreover, SHP-1-deficient mice have abnormalities in almost all blood cell lineages, emphasizing the importance of SHP-1 in blood cell [development](#). Both SHP-1 and SHP-2 also help terminate responses mediated by some [receptor](#) tyrosine kinases.

There are a large number of [transmembrane protein](#) tyrosine phosphatases, but the functions of most of them are unknown. At least some are thought to function as receptors; as this has not been directly demonstrated, however, they are referred to as **receptorlike tyrosine phosphatases**. They all have a single transmembrane segment and usually possess two tyrosine [phosphatase](#) domains on the cytosolic side of the [plasma membrane](#). An important example is the **CD45 protein**, which is found on the surface of all white blood cells and has an essential role in the activation of both T and B lymphocytes by foreign antigens. The [ligand](#) that is presumed to bind to the extracellular [domain](#) of the CD45 protein has not been identified. However, the role of CD45 in [signal transduction](#) has been studied by using [recombinant DNA](#) techniques to construct a hybrid protein with an extracellular EGF-binding domain and

intracellular CD45 tyrosine phosphatase domains. The surprising result is that EGF binding seems to inactivate the phosphatase activity of the hybrid protein rather than activating it.

This finding raises the possibility that some [receptor](#) tyrosine kinases and receptor tyrosine phosphatases may collaborate when they bind their respective cell-surface-bound ligands—with the kinases adding more phosphates and the [phosphatase](#) removing fewer—to maximally stimulate the tyrosine [phosphorylation](#) of selected intracellular signaling proteins. The significance of [ligand](#)-induced inhibition of CD45 phosphatase is still uncertain, however, and it seems unlikely to be the whole story; CD45 requires its phosphatase activity to function in [lymphocyte](#) activation.

Some receptorlike tyrosine phosphatases display features of cell-adhesion proteins and can even mediate homophilic cell-cell binding in cell adhesion assays. In the developing nervous system, for example, they may have an important role in guiding the growing tips of developing [nerve cell](#) axons to their targets. In *Drosophila*, the genes encoding several receptorlike tyrosine phosphatases are expressed exclusively in the nervous system, and when some of them are inactivated by [mutation](#), the axons of certain developing neurons fail to find their way to their normal targets. In some cases at least, the [phosphatase](#) activity of the [protein](#) is required to counteract the action of a cytoplasmic tyrosine kinase for normal [axon](#) guidance.

Transmembrane tyrosine phosphatases can also serve as signaling ligands that activate receptors on a neighboring cell. An example is the protein tyrosine phosphatase ζ/β , which is expressed on the surface of certain glial cells in the mammalian brain. It binds to a receptor protein (called *contactin*) on developing nerve cells, stimulating the cells to extend long processes. It is possible that the phosphatase also conveys a signal to the glial cell in this interaction, but such bidirectional signaling has not been directly demonstrated for transmembrane tyrosine phosphatases.

Having discussed the crucial role of tyrosine phosphorylation and dephosphorylation in the intracellular signaling pathways activated by many enzyme-linked receptors, we now turn to a class of enzyme-linked receptors that rely entirely on serine/threonine phosphorylation. These transmembrane serine/ threonine kinases activate an even more direct signaling pathway to the nucleus than does the Jak-STAT pathway discussed earlier. They directly phosphorylate latent gene regulatory proteins called *Smads*, which then migrate into the nucleus to activate gene transcription.

Signal Proteins of the TGF- β Superfamily Act Through Receptor Serine/Threonine Kinases and Smads

The **transforming growth factor- β (TGF- β) superfamily** consists of a large number of structurally related, secreted, dimeric proteins. They act either as hormones or, more commonly, as local mediators to regulate a wide range of biological functions in all

animals. During [development](#), they regulate pattern formation and influence various cell behaviors, including proliferation, [differentiation](#), [extracellular matrix](#) production, and cell death. In adults, they are involved in tissue repair and in immune regulation, as well as in many other processes. The superfamily includes the *TGF-β*s themselves, the *activins*, and the *bone morphogenetic proteins (BMPs)*. The BMPs constitute the largest family.

All of these proteins act through [enzyme](#)-linked receptors that are single-pass transmembrane proteins with a serine/threonine kinase [domain](#) on the cytosolic side of the [plasma membrane](#). There are two classes of these **[receptor](#)serine/threonine kinases**—*type I* and *type II*—which are structurally similar. Each member of the [TGF-β superfamily](#) binds to a characteristic combination of type-I and type-II receptors, both of which are required for signaling. Typically, the [ligand](#) first binds to and activates a type-II receptor homodimer, which recruits, phosphorylates, and activates a type-I receptor homodimer, forming an active tetrameric receptor [complex](#).

Once activated, the [receptor complex](#) uses a strategy for rapidly relaying the signal to the [nucleus](#) that is very similar to the Jak-STAT strategy used by [cytokine](#) receptors. The route to the nucleus, however, is even more direct. The type-I receptor directly binds and phosphorylates a latent [gene regulatory protein](#) of the **Smad family** (named after the first two identified, Sma in *C. elegans* and Mad in *Drosophila*). Activated TGF-β receptors and activin receptors phosphorylate Smad2 or

Smad3, while activated BMP receptors phosphorylate Smad1, Smad5, or Smad8. Once one of these Smads has been phosphorylated, it dissociates from the receptor and binds to Smad4, which can form a complex with any of the above five *receptor-activated Smads*. The Smad complex then moves into the nucleus, where it associates with other gene regulatory proteins, binds to specific sites in [DNA](#), and activates a particular set of target genes.

Some TGF- β family members serve as graded morphogens during [development](#), inducing different responses in a developing cell depending on their concentration (discussed in Chapter 21). The different responses can be reproduced by experimentally altering the amount of active Smad complexes in the [nucleus](#), suggesting that the level of these complexes may provide a direct readout of the level of [receptor](#) activation. If the [DNA](#)-binding sites in different target genes have different affinities for the complexes, then the particular genes activated would reflect the cell's position in the concentration gradient of the [morphogen](#).

As with the Jak-STAT pathway, the Smad pathway is also often regulated by [feedback inhibition](#). Among the target genes activated by Smad complexes are those that encode *inhibitory Smads*, including Smad6 and Smad7. These Smads act as decoys. They bind to activated type-I receptors and prevent other Smads from binding there. This blocks the formation of active Smad complexes and shuts off the response to the TGF- β family [ligand](#). Other types of extracellular ligands can also stimulate the production of inhibitory Smads to antagonize signaling by a TGF- β ligand; γ -interferon, for example, activates the

Jak-STAT pathway, and the resulting activated STAT dimers induce the production of Smad7, which inhibits signaling by TGF- β .

In addition to these intracellular inhibitors, a number of secreted extracellular inhibitory proteins can also neutralize signaling mediated by TGF- β family members. They directly bind to the signal molecules and prevent them from activating their receptors on target cells. *Noggin* and *chordin*, for example, inhibit BMPs, and *folliculin* inhibits activins. Noggin and chordin help to induce the [development](#) of the vertebrate nervous system by preventing BMPs from inhibiting this development (discussed in Chapter 21). The TGF- β family members, as well as some of their inhibitors, are usually secreted as inactive precursors that are subsequently activated by proteolytic [cleavage](#).

We turn now to [enzyme](#)-linked receptors that are neither kinases nor associated with kinases. We saw earlier that [nitric oxide](#) is widely used as a signaling [molecule](#), diffusing through the [plasma membrane](#) of a target cell and stimulating a cytoplasmic guanylylcyclase to produce the intracellular mediator [cyclic GMP](#). The receptors we now consider are transmembrane proteins with guanylylcyclase activity.

Receptor GuanylylCyclases Generate Cyclic GMP Directly

Receptor guanylylcyclases are single-pass transmembrane proteins with an extracellular [binding site](#) for a [signal molecule](#) and an intracellular guanylylcyclase catalytic [domain](#). The binding of the

signal molecule activates the cyclase domain to produce cyclic GMP, which in turn binds to and activates a *cyclic GMP-dependent* protein kinase (PKG), which phosphorylates specific proteins on serine or threonine. Thus, receptor guanylylcyclases use cyclic GMP as an intracellular mediator in the same way that some G-protein-linked receptors use cyclic AMP, except that the linkage between ligand binding and cyclase activity is a direct one.

Among the signal molecules that use receptor guanylylcyclase receptors are the *natriuretic peptides* (NPs), a family of structurally related secreted signal peptides that regulate salt and water balance and dilate blood vessels. There are several types of NPs, including *atrial natriuretic peptide* (ANP) and *brain natriuretic peptide* (BNP). Muscle cells in the atrium of the heart secrete ANP when blood pressure rises. The ANP stimulates the kidneys to secrete Na^+ and water and induces the smooth muscle cells in blood vessels walls to relax. Both of these effects tend to lower the blood pressure. When gene targeting is used to inactivate the ANP receptor guanylylcyclase in mice, the mice have chronically elevated blood pressure, resulting in progressive heart enlargement.

An increasing number of receptor guanylylcyclases are being discovered, but in most cases they are orphan receptors, where the ligand that normally activates them is unknown. The genome of the nematode *C. elegans*, for example, encodes 26 of these receptors. Most of those that have been studied are expressed in specific subsets of sensory neurons, suggesting that they may be involved in

detecting particular molecules in the worm's environment. Some of the orphan receptors in mammals are found in sensory neurons in the part of the nose involved in detecting pheromones.

All the signaling pathways activated by [G-protein](#)-linked and [enzyme](#)-linked receptors we have discussed so far depend on serine/threonine-specific protein kinases, tyrosine-specific protein kinases, or both. Some enzyme-linked receptors, however, depend on an entirely unrelated type of [protein kinase](#), as we now discuss.

Bacterial Chemotaxis Depends on a Two-Component Signaling Pathway Activated by Histidine-Kinase-associated Receptors

As pointed out earlier, many of the mechanisms involved in chemical signaling between cells in multicellular animals are thought to have evolved from mechanisms used by unicellular organisms to respond to chemical changes in their environment. In fact, some of the same intracellular mediators, such as cyclic nucleotides and Ca^{2+} , are used by both types of organisms. Among the best-studied reactions of unicellular organisms to extracellular signals are their chemotactic responses, in which cell movement is oriented toward or away from a source of some chemical in the environment. We conclude this [section](#) on [enzyme](#)-linked receptors with a brief account of bacterial [chemotaxis](#), which depends on a **two-component signaling pathway**, involving [histidine-kinase-associated receptors](#). The same type of signaling pathway is used by yeasts and plants, although apparently not by animals.

Motile bacteria will swim toward higher concentrations of nutrients (*attractants*), such as sugars, amino acids, and small peptides, and away from higher concentrations of various noxious chemicals (*repellents*). They swim by means of flagella, each of which is attached by a short, flexible hook at its base to a small protein disc embedded in the bacterial membrane. This disc is part of a tiny motor that uses the energy stored in the transmembrane H^+ gradient to rotate rapidly and turn the helical flagellum. Because the flagella on the bacterial surface have an intrinsic “handedness,” different directions of rotation have different effects on movement. Counterclockwise rotation allows all the flagella to draw together into a coherent bundle, so that the bacterium swims uniformly in one direction. Clockwise rotation causes them to fly apart, so that the bacterium tumbles chaotically without moving forward. In the absence of any environmental stimulus, the direction of rotation of the disc reverses every few seconds, producing a characteristic pattern of movement in which smooth swimming in a straight line is interrupted by abrupt, random changes in direction caused by tumbling.

The normal swimming behavior of bacteria is modified by chemotactic attractants or repellents, which bind to specific receptor proteins and affect the frequency of tumbling by increasing or decreasing the time that elapses between successive changes in direction of flagellar rotation. When bacteria are swimming in a favorable direction (toward a higher concentration of an attractant or away from a higher concentration of a repellent), they tumble less frequently than when they are swimming in an

unfavorable direction (or when no gradient is present). Since the periods of smooth swimming are longer when a bacterium is traveling in a favorable direction, it will gradually progress in that direction—toward an attractant or away from a repellent.

These responses are mediated by histidine-kinase-associated **chemotaxis receptors**, which typically are dimeric transmembrane proteins that bind specific attractants and repellents on the outside of the plasma membrane. The cytoplasmic tails of the receptors are stably associated with an adaptor protein *CheW* and a histidine kinase *CheA*, which help to couple the receptors to the flagellar motor. Repellent binding activates the receptors, whereas attractant binding inactivates them; a single receptor can bind either type of molecule, with opposite consequences. The binding of a repellent to the receptor activates *CheA*, which phosphorylates itself on a histidine and almost immediately transfers the phosphate to an aspartic acid on a messenger protein *CheY*. The phosphorylated *CheY* dissociates from the receptor, diffuses through the cytosol, binds to the flagellar motor, and causes the motor to rotate clockwise, so that the bacterium tumbles. *CheY* has intrinsic phosphatase activity and dephosphorylates itself in a process that is greatly accelerated by the *CheZ* protein.

The response to an increase in the concentration of an attractant or repellent is only transient, even if the higher level of ligand is maintained, as the bacteria *desensitize*, or *adapt*, to the increased stimulus. Whereas the initial effect on tumbling occurs in less than a second, adaptation takes minutes. The adaptation is a

crucial part of the response, as it enables the bacteria to respond to *changes* in concentration of ligand rather than to steady-state levels. It is mediated by the covalent methylation (catalyzed by a methyl transferase) and demethylation (catalyzed by a methylase) of the [chemotaxis](#) receptors, which change their responsiveness to ligand binding when methylated.

All of the genes and proteins involved in this highly adaptive behavior have now been identified. It therefore seems likely that bacterial [chemotaxis](#) will be the first signaling system to be completely understood in molecular terms. Even in this relatively simple signaling network, computer-based simulations are required to comprehend how the system works as an integrated network. Cell signaling pathways will provide an especially rich area of investigation for a new generation of computational biologists, as their network properties will not be understandable without powerful computational tools.

There are some cell-surface [receptor](#) proteins that do not fit into the three major classes we have discussed thus far—[ion](#)-channel-linked, [G-protein](#)-linked, and [enzyme](#)-linked. In the next [section](#), we consider cell-surface receptors that activate signaling pathways that depend on [proteolysis](#). These pathways have especially important roles in animal [development](#).

Summary

There are five known classes of [enzyme](#)-linked receptors: (1) [receptor](#) tyrosine kinases, (2) tyrosine-kinase-associated receptors, (3) receptor serine/threonine kinases, (4) transmembraneguananylcyclases, and (5) histidine-

kinase-associated receptors. In addition, some transmembrane tyrosine phosphatases, which remove phosphate from phosphotyrosine side chains of specific proteins, are thought to function as receptors, although for the most part their ligands are unknown. The first two classes of receptors are by far the most numerous.

Ligand binding to [receptor](#) tyrosine kinases induces the receptors to cross-phosphorylate their cytoplasmic domains on multiple tyrosines. The autophosphorylation activates the kinases, as well as producing a set of phosphotyrosines that then serve as docking sites for a set of intracellular signaling proteins, which bind via their SH2 (or PTB) domains. Some of the docked proteins serve as adaptors to couple the receptors to the small [GTPase](#) **Ras**, which, in turn, activates a cascade of serine/threonine phosphorylations that converge on a [MAP](#)-kinase, which relays the signal to the [nucleus](#) by phosphorylating [gene](#) regulatory proteins there. **Ras** can also activate another [protein](#) that docks on activated receptor tyrosine kinases—PI 3-kinase—which generates specific inositol phospholipids that serve as docking sites in the [plasma membrane](#) for signaling proteins with PH domains, including [protein kinase](#) B (PKB).

Tyrosine-kinase-associated receptors depend on various cytoplasmic tyrosine kinases for their action. These kinases include members of the [Src family](#), which associate with many kinds of receptors, and the [focal adhesion kinase](#) (**FAK**), which associates with integrins at focal adhesions. The cytoplasmic tyrosine kinases then phosphorylate a variety of signaling proteins to relay the signal onward. The largest family of receptors in this class

is the [cytokine](#) receptors family. When stimulated by [ligand](#) binding, these receptors activate Jak cytoplasmic tyrosine kinases, which phosphorylate STATs. The STATs then dimerize, migrate to the [nucleus](#), and activate the transcription of specific genes. Receptor serine/threonine kinases, which are activated by signaling proteins of the [TGF- \$\beta\$ superfamily](#), act similarly: they directly phosphorylate and activate Smads, which then oligomerize with another Smad, migrate to the nucleus, and activate [gene](#) transcription.

Bacterial [chemotaxis](#) is mediated by histidine-kinase-associated chemotaxis receptors. When activated by the binding of a repellent, the receptors stimulate their associated [protein kinase](#) to phosphorylate itself on histidine and then transfer that phosphate to a messenger protein, which relays the signal to the flagellar motor to alter the bacterium's swimming behavior. Attractants have the opposite effect on this kinase and therefore on swimming.

Signaling Pathways That Depend on Regulated Proteolysis

The need for intercellular signaling is never greater than during animal [development](#). Each cell in the embryo has to be guided along one developmental pathway or another according to its history, its position, and the character of its neighbors. At each step in the pathway, it must exchange signals with its neighbors to coordinate its behavior with theirs. Most of the signaling pathways already discussed are widely used for these purposes. But there are also others that relay signals in other ways from cell-surface

receptors to the interior of the cell. These additional signaling pathways all depend, in part at least, on regulated [proteolysis](#). Although most of them first came to light through genetic studies in *Drosophila*, they have been highly conserved in evolution and are used over and over again during animal development. As we discuss in Chapter 22, they also have a crucial role in the many developmental processes that continue in adult tissues.

We discuss four of these signaling pathways in this [section](#): the pathway mediated by the [receptor protein Notch](#), the pathway activated by secreted *Wnt* proteins, the pathway activated by secreted *Hedgehog* proteins, and the pathway that depends on activation of the latent [gene regulatory protein NF- \$\kappa\$ B](#). All of these pathways have crucial roles in animal [development](#). If any one of them is inactivated in a mouse, for example, development is seriously disturbed, and the mouse dies as an embryo or at birth.

The Receptor Protein Notch Is Activated by Cleavage

Signaling through the [Notch receptor protein](#) may be the most widely used signaling pathway in animal [development](#). It has a general role in controlling [cell fate](#) choices during development, mainly by amplifying and consolidating molecular differences between adjacent cells. Although [Notch](#) signaling is involved in the development of most tissues, it is best known for its role in [nerve cell](#) production in *Drosophila*. The nerve cells usually arise as isolated single cells within an epithelial sheet of precursor cells.

During the process, each future nerve cell or committed nerve-cell precursor signals to its immediate neighbours not to develop in the same way at the same time, a process known as *lateral inhibition*. In a fly embryo, for example, the inhibited cells around the future nerve-cell precursors develop into epidermal cells. Lateral inhibition depends on a [contact-dependent signaling](#) mechanism that is mediated by a signal protein called **Delta**, displayed on the surface of the future neural cell. By binding to **Notch** on a neighboring cell, Delta signals to the neighbor not to become neural. When this signaling process is defective in flies, the neighbors of neural cells also develop as neural cells, producing a huge excess of neurons at the expense of epidermal cells, which is lethal. Signaling between adjacent cells via **Notch** and Delta (or the Deltalike [ligand Serrate](#)) regulates cell fate choices in a wide variety of tissues and animals, helping to create fine-grained patterns of distinct cell types. The **Notch**-mediated signal can have other effects beside lateral inhibition; in some tissues, for example, it works in the opposite way, causing neighboring cells to behave similarly.

Both [Notch](#) and Delta are single-pass transmembrane proteins, and both require proteolytic processing to function. Although it is still unclear why Delta has to be cleaved, the [cleavage](#) of **Notch** is central to how **Notch** activation alters [gene expression](#) in the [nucleus](#). When activated by the binding of Delta on another cell, an intracellular protease cleaves off the cytoplasmic tail of **Notch**, and the released tail moves into the nucleus to activate the transcription of a set of **Notch**-response genes. The **Notch** tail acts by binding to a [gene regulatory](#)

protein called *CSL* (so named because it is called CBF1 in mammals, Suppressor of Hairless in flies, and Lag-1 in worms); this converts *CSL* from a transcriptional repressor into a transcriptional activator. The products of the main genes directly activated by **Notch** signaling are themselves gene regulatory proteins, but with an inhibitory action: they block the expression of genes required for neural differentiation (in the nervous system), and of various other genes in other tissues.

The Notch receptor undergoes three proteolytic cleavages, but only the last two depend on Delta. As part of its normal biosynthesis, a protease called *furin* acts in the Golgi apparatus to cleave the newly synthesized **Notch** protein in its future extracellular domain.

This cleavage converts **Notch** into a heterodimer, which is then transported to the cell surface as the mature receptor. The binding of Delta to **Notch** induces a second cleavage in the extracellular domain, mediated by a different protease. A final cleavage quickly follows, cutting free the cytoplasmic tail of the activated receptor.

The cleavage of the Notch tail occurs very close to the plasma membrane, just within the transmembrane segment. In this respect it resembles the cleavage of another, more sinister transmembrane protein—the β -*amyloid precursor protein (APP)*, which is expressed in neurons and is implicated in Alzheimer's disease. APP is cleaved within its transmembrane segment, releasing one peptide fragment into the extracellular space of the brain and another into the cytosol of the neuron. In Alzheimer's disease, the extracellular fragments accumulate in

excessive amounts and aggregate into filaments that form amyloid plaques, which are believed to injure nerve cells and contribute to their loss. The most frequent genetic cause of early-onset Alzheimer's disease is a [mutation](#) in the *presenilin-1 (PS-1)* [gene](#), which encodes an 8-pass transmembrane protein that participates in the cleavage of APP. The mutations in PS-1 cause cleavage of APP into amyloid-plaque-forming fragments at an increased rate. Genetic evidence in *C. elegans*, *Drosophila*, and mice indicates that the PS-1 protein is a required component of the **Notch** signaling pathway, helping to perform the final cleavage that activates **Notch**. Indeed, **Notch** signaling and cleavage are greatly impaired in PS-1-deficient cells.

Remarkably, [Notch](#) signaling is regulated by [glycosylation](#). The *Fringe* family of glycosyltransferases adds extra sugars to the O-linked [oligosaccharide](#) (discussed in Chapter 13) on **Notch**, which alters the specificity of **Notch** for its ligands. This has provided the first example of the modulation of [ligand-receptor](#) signaling by differential receptor glycosylation.

Wnt Proteins Bind to Frizzled Receptors and Inhibit the Degradation of β -Catenin

Wnt proteins are secreted signal molecules that act as local mediators to control many aspects of [development](#) in all animals that have been studied. They were discovered independently in flies and in mice: in *Drosophila*, the *wingless (wg)* [gene](#) originally came to light because of its role in wing development, while in mice, the *Int-1* gene was found because it promoted the

development of breast tumors when activated by the integration of a [virus](#) next to it. The cell-surface receptors for the Wnts belong to the **Frizzled** family of seven-pass transmembrane proteins. They resemble [G-protein](#)-linked receptors in structure, and some of them can signal through G proteins and the inositol [phospholipid](#) pathway discussed earlier. They mainly signal, however, through G-protein-independent pathways, which require a cytoplasmic signaling protein called *Dishevelled*.

The best characterized of the Dishevelled-dependent pathways acts by regulating the [proteolysis](#) of a multifunctional [protein](#) called **β -catenin** (or Armadillo in flies), which functions both in cell-cell adhesion and as a latent [gene regulatory protein](#). Wnts activate this pathway by binding to both a Frizzled protein and a co-[receptor](#) protein. The co-receptor protein is related to the low density lipoprotein ([LDL](#)) receptor protein and is therefore called *LDL-receptor-related protein (LRP)*. It is uncertain how Frizzled and LRP activate Dishevelled, which relays the signal onward.

In the absence of Wnt signaling, most of a cell's β -catenin is located at cell-cell adherens junctions, where it is associated with *cadherins*, which are transmembrane adhesion proteins. As discussed in Chapter 19, the β -catenin in these junctions helps link the cadherins to the [actin cytoskeleton](#). Any β -catenin not associated with cadherins is rapidly degraded in the [cytoplasm](#). This degradation depends on a large degradation [complex](#), which recruits β -catenin and contains at least three other proteins.

1.

A serine/threonine kinase called **glycogen synthase kinase-3 β (GSK-3 β)** phosphorylates β -catenin, thereby marking the protein for ubiquitylation and rapid degradation in proteasomes.

2.

The tumor-suppressor protein adenomatous polyposis coli (APC) is so named because the gene encoding it is often mutated in a type of benign tumor (adenoma) of the colon. The tumor projects into the lumen as a polyp, which can eventually become malignant. APC helps promote the degradation of β -catenin by increasing the affinity of the degradation complex for β -catenin, as required for effective phosphorylation of β -catenin by GSK-3 β .

3.

A scaffold protein called *axin* holds the protein complex together.

The binding of a Wnt protein to Frizzled and LRP leads to the inhibition of β -catenin phosphorylation and degradation. The mechanism is not understood in detail, but it requires Dishevelled and several other signaling proteins that bind to Dishevelled, including the serine/threonine kinase called *casein kinase 1*. As a result, unphosphorylated β -catenin accumulates in the cytoplasm and nucleus.

In the nucleus, the target genes for Wnt signaling are normally kept silent by an

inhibitory [complex](#) of [gene](#) regulatory proteins, which includes proteins of the *LEF-1/TCF* family bound to the corepressor [protein](#) *Groucho*. The increase in undegraded β -catenin caused by Wnt signaling allows β -catenin to enter the nucleus and bind to LEF-1/TCF, displacing Groucho. The β -catenin now functions as a coactivator, inducing the transcription of the Wnt target genes.

Among the genes activated by β -catenin is *c-myc*, which encodes a [protein](#) (c-Myc) that is a powerful stimulator of cell growth and proliferation (discussed in Chapter 17). Mutations of the [APC gene](#) occur in 80% of human colon cancers. These mutations inhibit the protein's ability to bind β -catenin, so that β -catenin accumulates in the [nucleus](#) and stimulates the transcription of *c-myc* and other Wnt target genes, even in the absence of Wnt signaling. The resulting uncontrolled cell proliferation promotes the [development](#) of cancer.

Hedgehog Proteins Act Through a Receptor Complex of Patched and Smoothed, Which Oppose Each Other

Like Wnt proteins, the **Hedgehog** proteins are a family of secreted signal molecules that act as local mediators in many developmental processes in both invertebrates and vertebrates. Abnormalities in the Hedgehog pathway during [development](#) can be lethal and in adult cells can also lead to cancer. The Hedgehog proteins were discovered in *Drosophila*, where a [mutation](#) in the only [gene](#) encoding such a [protein](#) produces a larva with spiky processes (denticles) resembling a hedgehog. At least three genes encode Hedgehog proteins in vertebrates—*sonic*,

desert, and *indian hedgehog*. The active form of all Hedgehog proteins is unusual in that it is covalently coupled to [cholesterol](#), which helps to restrict its [diffusion](#) following secretion. The cholesterol is added during a remarkable processing step, in which the protein cleaves itself. The proteins are also modified by the addition of a [fatty acid](#) chain, which, for unknown reasons, can be required for their signaling activity.

Two transmembrane proteins, Patched and Smoothed, mediate the responses to all Hedgehog proteins. **Patched** is predicted to cross the [plasma membrane](#) 12 times, and it is the [receptor](#) that binds the Hedgehog [protein](#). In the absence of a Hedgehog signal, Patched inhibits the activity of **Smoothed**, which is a 7-pass [transmembrane protein](#) with a structure similar to a Frizzled protein. This inhibition is relieved when a Hedgehog protein binds to Patched, allowing Smoothed to relay the signal into the cell. Most of what we know about the downstream signaling pathway activated by Smoothed comes from genetic studies in flies, and it is the fly pathway that we summarize here.

In some respects the Hedgehog signaling pathway in *Drosophila* operates similarly to the Wnt pathway. In the absence of a Hedgehog signal, a [gene regulatory protein](#) called **Cubitus interruptus (Ci)** is proteolytically cleaved in proteasomes. Instead of being completely degraded, however, it is processed to form a smaller protein that accumulates in the [nucleus](#), where it acts as a transcriptional [repressor](#), helping to keep some Hedgehog-responsive genes silent. The proteolytic processing of the Ci protein depends on a large multiprotein [complex](#). The

complex contains a serine/threonine kinase (called Fused) of unknown function, an anchoring protein (called Costal) that binds the complex to microtubules (keeping Ci out of the nucleus), and an [adaptor protein](#) (called Suppressor of Fused) ([Figure 15-73A](#)). When Hedgehog binds to Patched to activate the signaling pathway, Ci processing is suppressed, and the unprocessed Ci protein is released from its complex and enters the nucleus, where it activates the transcription of Hedgehog target genes.

Among the genes activated by Ci is the [gene](#) that encodes the Wnt [protein](#) Wingless, which helps pattern tissues in the fly embryo (discussed in Chapter 21). Another target gene is *patched* itself; the resulting increase in Patched protein on the cell surface inhibits further Hedgehog signaling—a form of negative feedback.

Many gaps in the Hedgehog signaling pathway still remain to be filled in. It is not known, for example, how Patched inhibits Smoothed, how Smoothed activates the pathway, how the [proteolysis](#) of Ci is regulated (although it is known that Ci [phosphorylation](#) by [PKA](#) is required for the processing), or how the release of the [complex](#) from microtubules and unprocessed Ci from the complex is controlled.

Even less is known about the Hedgehog pathway in vertebrate cells. In addition to there being at least three types of vertebrate Hedgehog proteins, there are two forms of Patched and three Ci-like proteins (*Gli1*, *Gli2*, and *Gli3*). Unlike in flies, Hedgehog signaling stimulates the transcription of the *Gli* genes, and it is unclear whether all of the Gli proteins undergo proteolytic

processing, although there is evidence that Gli3 does. Inactivating mutations in one of the human *patched* genes, which leads to excessive Hedgehog signaling, occur frequently in the most common form of skin cancer (*basal cell carcinoma*), suggesting that Patched normally helps to keep skin cell proliferation in check.

[Go to:](#)

Multiple Stressful and Proinflammatory Stimuli Act Through an NF-κB-Dependent Signaling Pathway

The **NF-κB proteins** are latent [gene](#) regulatory proteins that lie at the heart of most inflammatory responses. These responses occur as a [reaction](#) to infection or injury and help protect the animal and its cells from these stresses. When excessive or inappropriate, however, inflammatory responses can also damage tissue and cause severe pain, as happens in joints in rheumatoid arthritis, for example. NF-κB proteins also have an important role in intercellular signaling during normal vertebrate [development](#), although the extracellular signals that activate NF-κB in these circumstances are unknown. In *Drosophila*, however, genetic studies have identified both the extracellular and the intracellular proteins that activate the NF-κB family member *Dorsal*, which has a crucial role in specifying the [dorsal-ventral](#) axis of the developing fly embryo. The same intracellular signaling pathway is also involved in defending the fly from infection, just as in vertebrates.

Two vertebrate cytokines are especially important in inducing inflammatory responses—*tumor necrosis factor α* (TNF-α) and [interleukin-1](#) (IL-1). Both are made by cells

of the innate [immune system](#), such as macrophages, in response to infection or tissue injury. These proinflammatory cytokines bind to cell-surface receptors and activate NF- κ B, which is normally sequestered in an inactive form in the [cytoplasm](#) of almost all of our cells. Once activated, NF- κ B turns on the transcription of more than 60 known genes that participate in inflammatory responses. Although TNF- α receptors and IL-1 receptors are structurally unrelated, they operate in much the same way.

There are five NF- κ B proteins in mammals (RelA, RelB, c-Rel, NF- κ B1, and NF- κ B2), and they form a variety of homodimers and heterodimers, each of which activates its own characteristic set of genes. Inhibitory proteins called **I κ B** bind tightly to the dimers and hold them in an inactive state within large [protein](#) complexes in the [cytoplasm](#). Signals such as TNF- α or IL-1 activate the dimers by triggering a signaling pathway that leads to the [phosphorylation](#), ubiquitylation, and consequent degradation of I κ B. The degradation of I κ B exposes a [nuclear localization signal](#) on the NF- κ B proteins, which now move into the [nucleus](#) and stimulate the transcription of specific genes. The phosphorylation of I κ B is performed by a specific serine/threonine kinase called **I κ B kinase (IKK)**.

The mechanism by which the binding of a proinflammatory [cytokine](#) to its cell-surface receptors activates I κ B kinase. Ligand binding causes the cytosolic tails of the clustered receptors to recruit various adaptor proteins and cytoplasmic serine/threonine kinases. One of the recruited kinases is thought to be an *I κ B kinase kinase*

(IKKK) that directly phosphorylates and activates the I κ B kinase (IKK).

Not all of the signaling proteins recruited to the cytosolic tail of the TNF- α [receptor](#) contribute to NF- κ B activation, however. Some can trigger a [MAP](#)-kinase cascade, while others can activate a proteolytic cascade that leads to [apoptosis](#).

Thus far, we have discussed cell signaling mainly in animals, with a few diversions into yeasts and bacteria. But intercellular signaling is just as important for plants as it is for animals, although the mechanisms and molecules used are mainly different, as we discuss next.

Summary

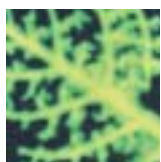
Some signaling pathways that are especially important in animal [development](#) depend on [proteolysis](#) for at least part of their action. [Notch](#) receptors are activated by [cleavage](#) when Delta (or a related [ligand](#)) on another cell binds to them; the cleaved cytosolic tail of **Notch** migrates into the [nucleus](#), where it stimulates [gene](#) transcription. In the Wnt signaling pathway, by contrast, the proteolysis of the latent [gene regulatory protein](#) β -catenin is inhibited when secreted Wnt proteins bind to their receptors; as a result, β -catenin accumulates in the nucleus and activates the transcription of Wnt target genes.

Hedgehog signaling in flies works much like Wnt signaling: in the absence of a signal, a bifunctional, cytoplasmic [gene regulatory protein](#) Ci is proteolytically cleaved to form a transcriptional [repressor](#) that keeps

Hedgehog target genes silenced. The binding of Hedgehog to its [receptor](#) inhibits the proteolytic processing of Ci; as a result, the larger form of Ci accumulates in the [nucleus](#) and activates the transcription of Hedgehog-responsive genes. Signaling through the latent gene regulatory protein NF- κ B also depends on [proteolysis](#). NF- κ B is normally held in an inactive state by the inhibitory protein I κ B within a multiprotein [complex](#) in the [cytoplasm](#). A variety of extracellular stimuli, including proinflammatory cytokines, trigger a [phosphorylation](#) cascade that ultimately phosphorylates I κ B, marking it for degradation; this enables the freed NF- κ B to enter the nucleus and activate the transcription of its target genes.

13

Recombinant DNA and Biotechnology



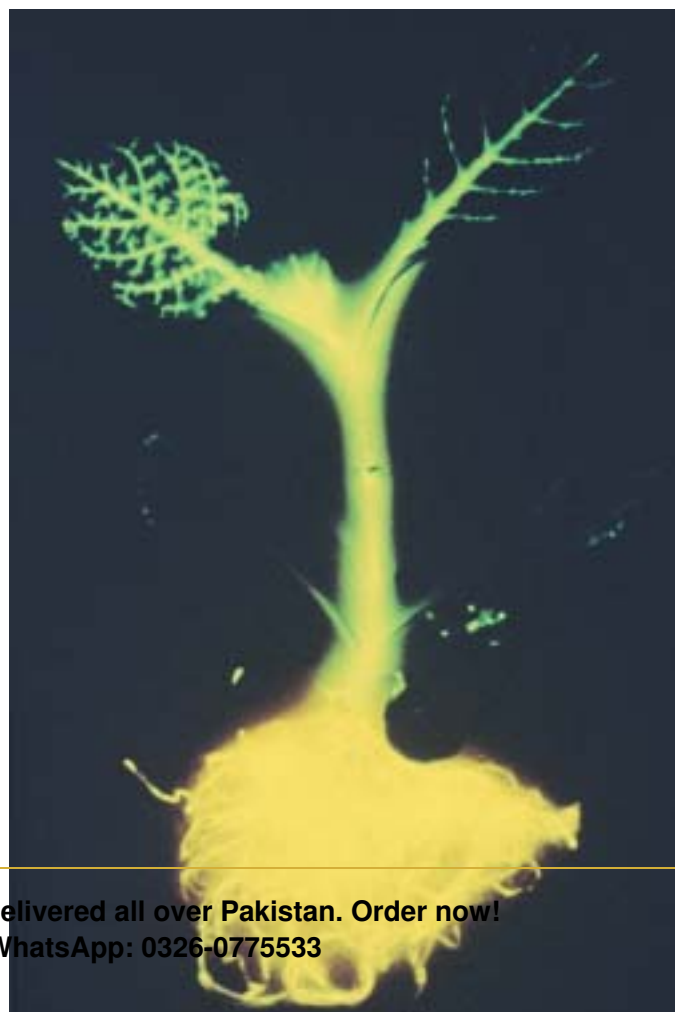
Of the many horrible legacies of the wars of the twentieth century, perhaps none is so lasting as the littering of the countryside with land mines. These inexpensive plastic shells filled with trinitrotoluene (TNT) are built to explode and injure whatever steps on the soil above them. Currently, the most common way of finding land mines is to poke around the soil with a stick—a precarious occupation at best. Since the mines are made of plastic, metal detectors do not work. The agricultural systems of many of the countries with large numbers of land mines, such as Cambodia and Angola, are based on manual labor, so where there are land mines, there cannot be a farm. Clearly, the world needs a sensitive, non-lethal land mine detector.

Enter biotechnology. Neal Stewart, at the University of North Carolina, is developing plants that can detect land mines in a field and show their locations remotely. His method is an excellent example of the application of knowledge of gene transcription and translation as well as of genetic engineering. In Chapter 14, we saw that the control of eukaryotic gene transcription lies at the promoter, a DNA sequence where RNA polymerase and other proteins bind to initiate transcription of the adjacent gene. Certain bacteria have a promoter that is sensitive to TNT, and the binding of a tiny amount of this chemical activates an adjacent gene.

Plants would be ideal biosensors for land mines, as seeds can be spread widely and evenly in a suspect field. But what gene could “announce” the presence of TNT by making a detectable protein? It turns out that certain jellyfish make a protein that fluoresces green when ultraviolet light (“black light”) is shone on it. Stewart has introduced this gene into a plant, placing it alongside the TNT-sensitive promoter. When these plants are grown near a land mine, their roots take up TNT, and their leaves will glow in ultraviolet light. Of course, having people plant these seeds or shine a black light on these plants as they walk through the field would bring us back to the old, dangerous way of detecting land mines. To solve these problems, seeding and remote sensing could be done from airplanes or helicopters flying over the field.

This story—from problem to solution—has been repeated many times in the past two decades. The products of recombinant DNA technology range from life-saving drugs that there is no other way to make in adequate amounts to crop plants with improved agricultural characteristics. Although the basic

A Land Mine Detector Plants can be genetically engineered to express green fluorescent protein from a jellyfish gene. When this “glow-in-the-dark” gene is linked so that it is activated by the presence of TNT in the soil, such plants can act as biosensors to detect the presence of explosive land mines.



14 *Molecular Biology and Medicine*



After his fiftieth birthday, Don's wife urged him to get a long-delayed medical checkup. He felt well, but realized that this was a good time to be screened for the various diseases that affect people as they get older. Don's routine blood count showed a surprise: Normally, people have about 5,000 white cells per milliliter of blood; Don had 40 times as many.

Within a day, he was diagnosed with chronic myeloid leukemia, a serious cancer where white blood cells from the bone marrow proliferate out of control.

An oncologist (a physician who specializes in treating cancer) put Don on an aggressive regimen of chemotherapy. The three drugs Don took were designed to kill dividing cells—hopefully in the tumor, but such drugs also affect normal cells. One drug blocked microtubules from forming, thus preventing the mitotic spindle from assembling; another inserted into the double helix and damaged DNA replication; the third drug inhibited an enzyme involved in nucleotide synthesis.

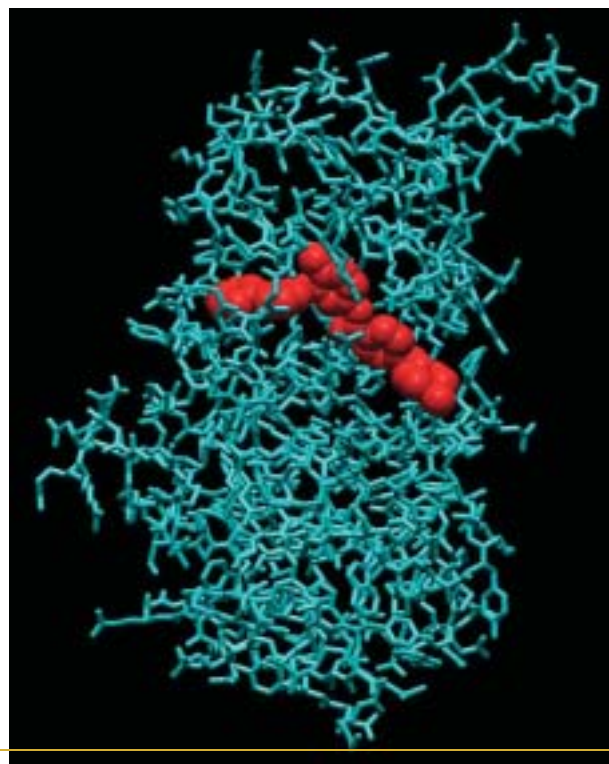
Although the side effects were hard on Don, his white cell count gradually got lower. But after 8 months, the decline stalled at 80,000 cells/ml—still dangerously high. Worse, most of them were not mature white blood cells, but were undifferentiated cells from bone marrow. The chemotherapy drugs had killed Don's normal bone marrow cells. Without these specialized mature white blood cells, Don would die within months.

The timely development in the 1990s of a molecular understanding of this leukemia led to a therapy that saved Don's life. Scientists had known for several decades that chronic myelogenous leukemia cells have a particular chromosome translocation between chromosomes 9 and 22. With the advent of DNA sequencing, it became clear that this translocation fuses together parts of two genes: half of a gene called *bcr* on chromosome 22, and half of a gene called *abl* on chromosome 9. The abnormal protein made from this fused gene has strong protein tyrosine kinase activity.

As described in Chapter 15, protein kinases activate proteins by phosphorylation, binding a phosphate group from ATP. During the 1990s, the leukemia protein kinase was purified, crystallized, and the geometry of its binding site for ATP was described in molecular terms. Armed with this knowledge, organic chemists designed a drug that fit into and obstructed the binding site for ATP on the abnormal kinase, inactivating the enzyme. The drug, called Gleevec, was given to Don, and within a few weeks his white cell count was 5,400—normal! The cancer had been virtually wiped out.

The development of Gleevec is an opening chapter in the molecular medicine of the future. Unlike conventional chemotherapy, which uses

A “Smart Drug” The drug Gleevec (shown here as the red molecule) binds to a protein kinase made by certain leukemia cells, preventing the cancer cells' reproduction. Gleevec is the first example of a rationally designed, specifically targeted cancer-fighting drug produced using the knowledge and techniques of molecular medicine.



All VU books are available in hard copy and delivered all over Pakistan. Order now!

Website: vubookshoppk.com | WhatsApp: 0326-0775533

are always present in the blood and become active only at a wound. People with the genetic disease *hemophilia* lack one of the coagulation proteins. Some people with this disease risk death from even minor cuts, since they cannot stop bleeding.

All of the foregoing protein alterations, and the diseases resulting from them, originate from an altered gene. But some abnormal proteins resulting in diseases may have a very different origin.

Prion diseases are disorders of protein conformation

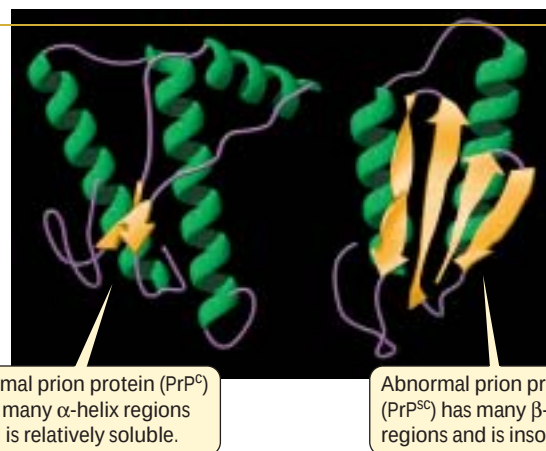
Transmissible spongiform encephalopathies (TSEs) are degenerative brain diseases that occur in many mammals, including humans. In these diseases, the brain gradually develops holes, leaving it looking like a sponge. Scrapie, a TSE that causes affected sheep and goats to rub the wool off their bodies, has been known for 250 years. In the 1980s, a TSE that appeared in cows in Britain was traced to the cows having eaten products from sheep that had scrapie. Then, in the 1990s, some people who had eaten beef from cows with this TSE got a human version of the disease (dubbed “mad cow disease” by the media), again suggesting that the causative agent could cross species lines.

Another instance of humans consuming an infective agent and getting a TSE involved kuru, a disease resulting in dementia that occurred among the Fore tribe of New Guinea. In the 1950s, it was discovered that people with kuru had consumed the brains of people who had died of it. When this ritual cannibalism stopped, so did the epidemic of kuru.

Researchers found that TSEs could be transmitted from one animal to another via brain extracts from a diseased animal. At first, a virus was suspected. But when Tikva Alper at Hammersmith Hospital, London, treated infectious extracts with high doses of ultraviolet light to inactivate nucleic acids, they still caused TSEs. She proposed that the causative agent for TSEs was a protein, not a virus. Later, Stanley Prusiner at the University of California purified the protein responsible and showed it to be free of DNA or RNA. He called it a proteinaceous infective particle, or **prion**.

Normal brain cells contain a membrane protein called PrP^C. A protein with the same amino acid sequence is present in TSE-affected brain tissues, but that protein, called PrP^{Sc}, has an altered shape (Figure 17.4). Thus, TSEs are usually not caused by a mutated gene (the primary structures of the two proteins are the same), but are somehow caused by an alteration in protein conformation. The altered three-dimensional structure of the protein has profound effects on its function in the cell. PrP^{Sc} is insoluble, and it piles up as fibers in brain tissue, causing cell death.

How can the exposure of a normal cell to material containing PrP^{Sc} result in a TSE? The abnormal PrP^{Sc} protein seems to induce a conformational change in the normal PrP^C



17.4 Prion Proteins Normal prion proteins (PrP^C, left) can be converted to the disease-causing form (PrP^{Sc}, right), which has a different three-dimensional structure.

protein so that it too becomes abnormal, just as one rotten apple results in a whole barrel full of rotten apples. Just how the conversion occurs, and how it causes a TSE, are unclear.

Prions appear to represent a highly unusual phenomenon in human disease. The vast majority of infectious and inherited diseases are understood in terms of proteins that are products of functional or dysfunctional genes. But the expression of these genes, like all genes, is influenced by the environment.

Most diseases are caused by both genes and environment

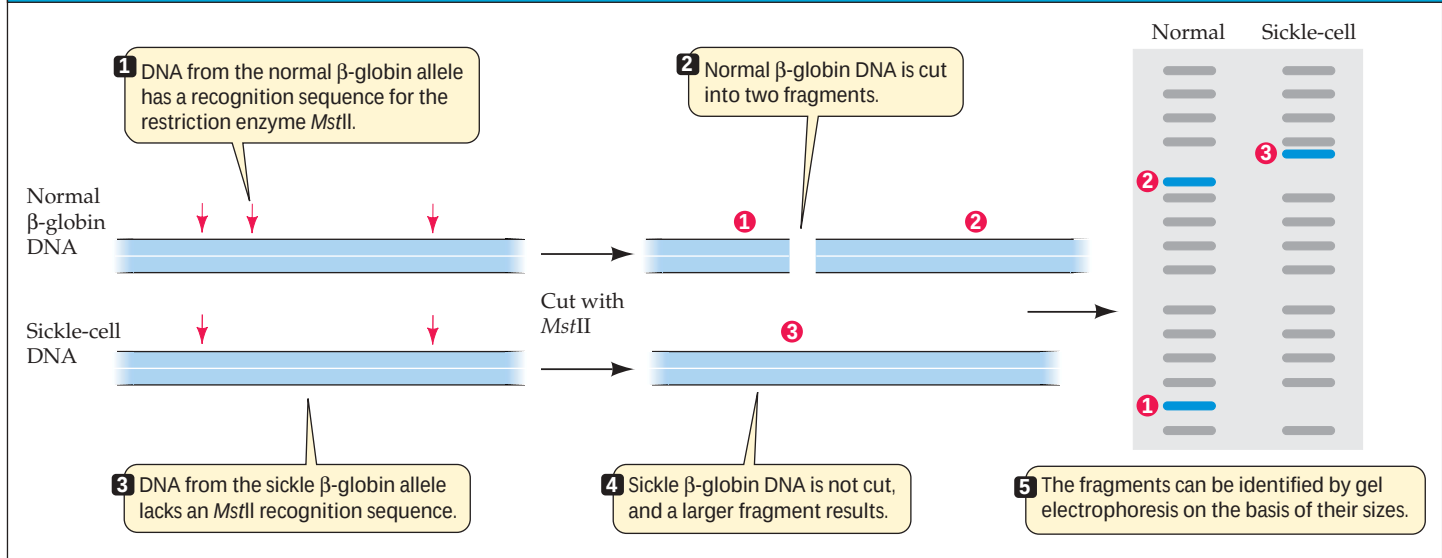
The human diseases for which clinical phenotypes can be traced to a single altered protein and its altered gene may number in the thousands, and in most cases they are dramatic evidence of a one-gene, one-polypeptide relationship. Taken together, these diseases have a frequency of about 1 percent in the total human population.

Far more common, however, are diseases that are **multifactorial**; that is, those that are caused by many genes and proteins interacting with one another and with the environment. Although we tend to call individuals either normal (wild-type) or abnormal (mutant), the sum total of our genes is what determines which of us who eat a high-fat diet will die of a heart attack, or which of us exposed to infectious bacteria will come down with a disease. Estimates suggest that up to 60 percent of all people are affected by diseases that are genetically influenced.

Human genetic diseases have several patterns of inheritance

As in any human genetic system, the alleles that cause genetic diseases are inherited in a dominant or recessive pattern, and are carried on autosomes or on sex chromosomes (see Chap-

RESEARCH METHOD



17.11 DNA Testing by Allele-Specific Cleavage Allele-specific cleavage, a technique similar to RFLP analysis, can be used to detect mutations such as the one that causes sickle-cell anemia.

genetic changes. These changes are mostly alterations in the DNA of somatic cells that are propagated by mitosis.

Cancer cells differ from their normal counterparts

Cancer cells differ from the normal cells from which they originate in two major ways.

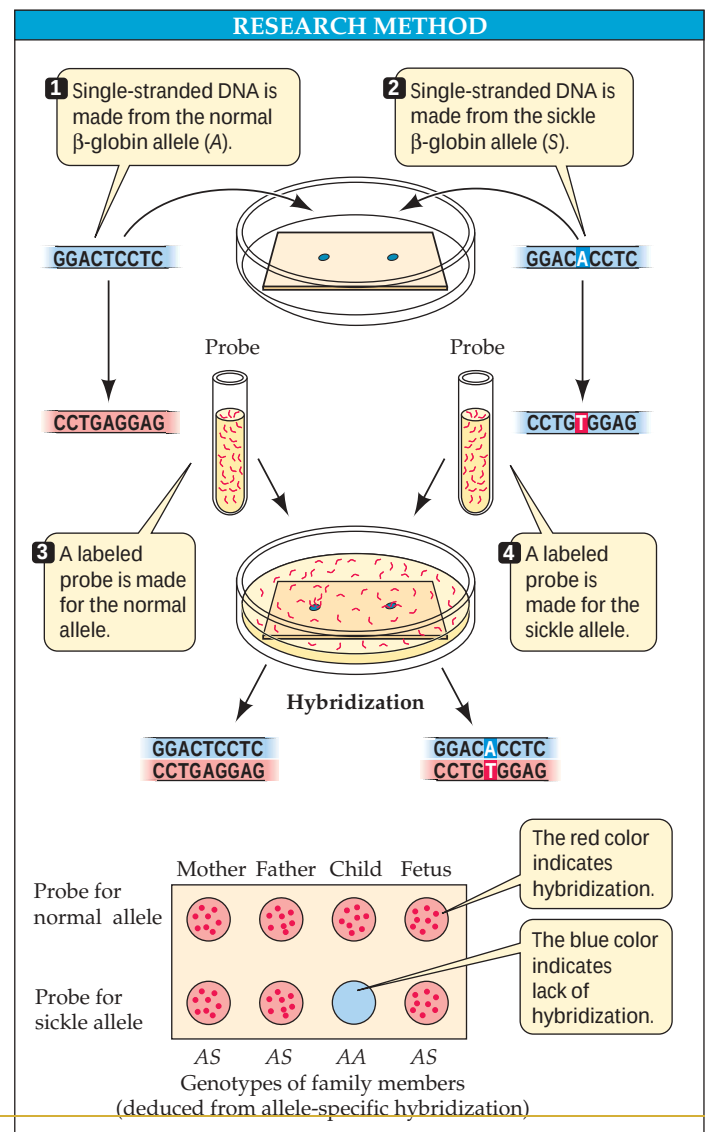
CANCER CELLS LOSE CONTROL OVER CELL DIVISION. Most cells in the body divide only if they are exposed to extracellular influences, such as growth factors or hormones. Cancer cells do not respond to these controls, and instead divide more or less continuously, ultimately forming **tumors** (large masses of cells). By the time a physician can feel a tumor or see one on an X ray or CAT scan, it already contains millions of cells.

Benign tumors resemble the tissue they came from, grow slowly, and remain localized where they develop. A lipoma, for example, is a benign tumor of fat cells that may arise in the armpit and remain there. Benign tumors are not cancers, but they must be removed if they impinge on an important organ, such as the brain.

Malignant tumors, on the other hand, do not look like their parent tissue at all. A flat, specialized lung epithelial cell in the lung wall may turn into a relatively featureless, round,

17.12 DNA Testing by Allele-Specific Oligonucleotide

Hybridization Testing of this family reveals that three of them are heterozygous carriers of the sickle allele. The first child, however, has inherited two normal alleles and is neither affected by the disease nor a carrier.



All VU books are available in hard copy and delivered all over Pakistan. Order now!

Website: vubookshoppk.com | WhatsApp: 0326-0775533

15 *Natural Defenses against Disease*



On January 6, 1777, George Washington, commander of the Revolutionary army of the fledgling United States, made a fateful decision. As he wrote to his chief physician, "Finding smallpox to be spreading much, and fearing that no precaution can prevent it from running through the whole of our army, I have determined that the troops shall be inoculated. Should the disease rage with its usual virulence, we should have more to dread from it than the sword of the enemy."

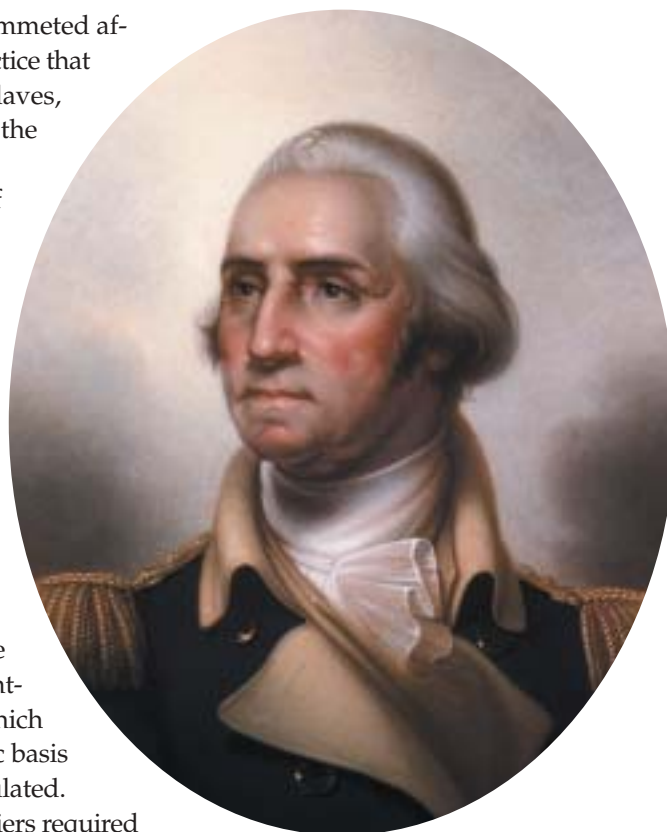
Washington was speaking from experience. He himself had survived the disease in 1751, when he was still a teenager. During 1776 his army lost 1,000 men in battle and 10,000 men to smallpox. This virulent disease, which killed up to 1 of every 4 people exposed to it, had already figured prominently in American history. A century before, it had decimated the native population, making colonization by Europeans easier. Two years previously at Quebec, it had laid waste to an American army that was trying to annex Canada by force.

The death rate due to smallpox in the Revolutionary army plummeted after Washington's order was carried out. How did inoculation, a practice that was learned from the people of the Near East and from African slaves, save the soldiers? And why was Washington himself immune to the disease as it ravaged his army?

The answers to these questions lie in the cells and molecules of the immune system. When Washington caught smallpox in 1751, specialized white blood cells in his body engulfed some of the smallpox viruses by phagocytosis and partly digested them. These cells, called macrophages, displayed fragments of the viruses on their surfaces. Other specialized white blood cells, called T cells, recognized those fragments, which caused them to divide and differentiate. Some descendants of those activated T cells then attacked Washington's virus-infected cells, preventing the lethal spread of the disease. Other descendants of the T cells persisted in his body as "memory cells" and rapidly divided again to defend him when he was exposed to the disease as an adult. Inoculation of Washington's soldiers with powdered scabs from smallpox patients, containing dead smallpox viruses, stimulated the formation of these memory cells in their bodies, once again preventing the virus from spreading following infection. This practice, which had been used for centuries, was finally placed on a more scientific basis by Edward Jenner two decades after Washington's army was inoculated.

These defensive events in the bodies of Washington and his soldiers required the participation of many kinds of cells and proteins. This chapter begins by intro-

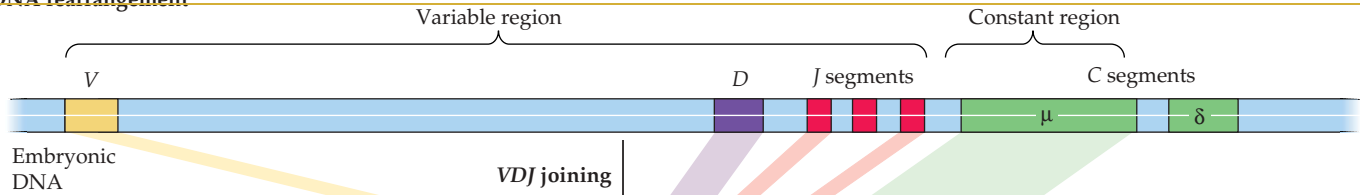
George Washington Washington's decision to immunize his army against smallpox saved many lives and probably helped him win the Revolutionary War.



All VU books are available in hard copy and delivered all over Pakistan. Order now!

Website: vubookshoppk.com | WhatsApp: 0326-0775533

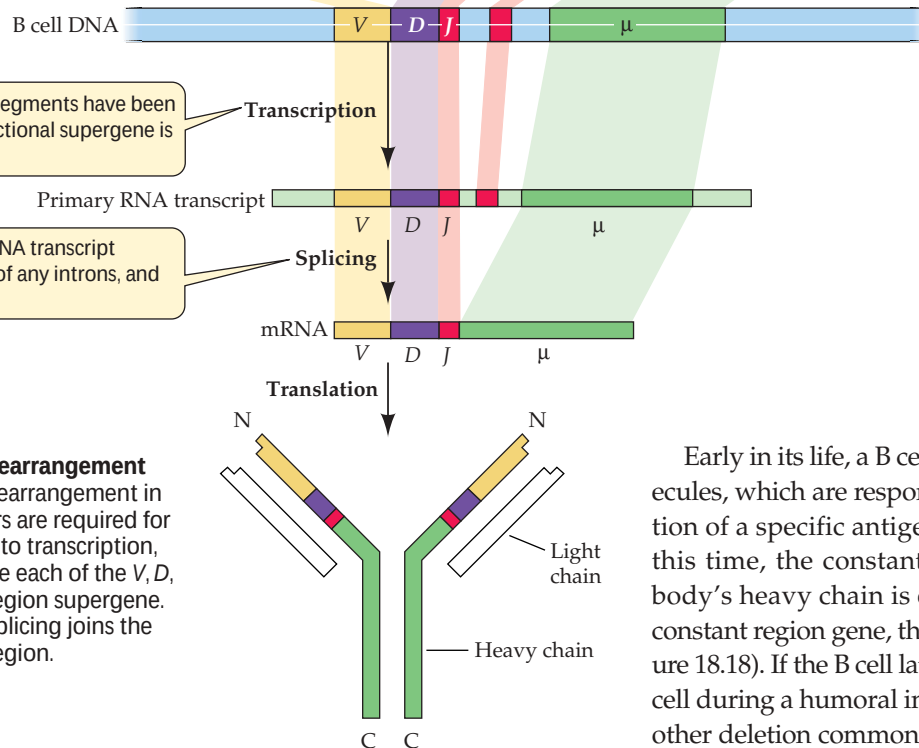
(a) DNA rearrangement



(b) RNA splicing

1 After V, D, J, and C DNA segments have been joined, the resulting functional supergene is transcribed.

2 Splicing of the primary RNA transcript removes the transcripts of any introns, and any extra J genes.



18.19 Heavy-Chain Gene Rearrangement and Splicing Two types of rearrangement in the heavy-chain gene clusters are required for antibody formation. (a) Prior to transcription, DNA is rearranged to join one each of the V, D, and J genes into a variable region supergene. (b) After transcription, RNA splicing joins the VDJ region to the constant region.

When we add these possibilities to the billions of combinations that can be made by random DNA rearrangements, it is not surprising that the immune system can mount a response to almost any natural or artificial substance.

Once this *pretranscriptional* processing is completed, pre-mRNA can be transcribed from each supergene. Posttranscriptional processing removes the remaining introns, so that the mature mRNA contains a continuous coding sequence for an immunoglobulin light chain or heavy chain. Translation then produces the polypeptide chains, which combine to form an active antibody protein.

This genetic system is capable of still other kinds of changes, as seen when a B cell or plasma cell switches the immunoglobulin class it produces, but retains its antibody specificity.

The constant region is involved in class switching

In Table 18.3, we described the different classes of antibodies and their functions. Generally, a B cell makes only one antibody class at a time. But **class switching** can occur, in which a B cell changes which antibody class it synthesizes. For example, a B cell making IgM can switch to making IgG.

Early in its life, a B cell produces IgM molecules, which are responsible for its recognition of a specific antigenic determinant. At this time, the constant region of the antibody's heavy chain is encoded by the first constant region gene, the μ segment (see Figure 18.18). If the B cell later becomes a plasma cell during a humoral immune response, another deletion commonly occurs in the cell's DNA, positioning the heavy-chain variable region gene (consisting of the same V, D, and J segments) next to a constant region gene farther down the original DNA, such as the γ , ϵ , or α genes (Figure 18.20). Such a DNA deletion results in the production of an antibody with a different constant region of the heavy chain, and therefore a different function. However, the antibody produced has *the same variable regions of the light and heavy chains*, and therefore the same antigen specificity, as before. The new antibody falls into one of the four other immunoglobulin classes (IgA, IgD, IgE, or IgG), depending on which of the constant region genes is placed adjacent to the variable region gene.

After switching classes, the plasma cell cannot go back to making the previous immunoglobulin class, because that part of the DNA has been lost. On the other hand, if additional constant region segments are still present, the cell may switch classes again.

What triggers class switching, and what determines the class to which a given B cell will switch? T_H cells direct the course of an immune response and determine the nature of the attack on the antigen. These T cells induce class switching by sending cytokine signals. The cytokines bind to receptors on the target B cells, generating a signal transduction cascade that results in altered transcription of the immunoglobulin genes.

All VU books are available in hard copy and delivered all over Pakistan. Order now!

Website: vubookshoppk.com | WhatsApp: 0326-0775533

Lymphocytes and the Cellular Basis of Adaptive Immunity

Lymphocytes are responsible for the astonishing specificity of adaptive **immune** responses. They occur in large numbers in the blood and [lymph](#) (the colorless fluid in the lymphatic vessels that connect the lymph nodes in the body to each other and to the bloodstream) and in [lymphoid organs](#)

Some cells of the innate [immune system](#) directly present microbial antigens to T cells to initiate an [adaptive immune response](#). The cells that do this most efficiently are called *dendritic cells*, which are present in most vertebrate tissues. They recognize and phagocytose invading microbes or their products at a site of infection and then migrate with their prey to a nearby peripheral [lymphoid organ](#). There they act as [antigen-presenting cells](#), which directly activate T cells to respond to the microbial antigens. Once activated, some of the T cells then migrate to the site of infection, where they help other phagocytic cells, mainly macrophages, destroy the microbes. Other activated T cells remain in the lymphoid organ and help B cells respond to the microbial antigens. The activated B cells secrete antibodies that circulate in the body and coat the microbes, targeting them for efficient [phagocytosis](#).

Thus, innate **immune** responses are activated mainly at sites of infection, whereas adaptive **immune** responses are activated in peripheral lymphoid organs. The two types of responses work together to eliminate invading pathogens.

B Lymphocytes Develop in the Bone Marrow; T Lymphocytes Develop in the Thymus

T cells and B cells derive their names from the organs in which they develop. T cells develop in the *thymus*, and B cells, in mammals, develop in the *bone marrow* in adults or the liver in fetuses.

Despite their different origins, both T and B cells develop from the same *pluripotent hemopoietic stem cells*, which give rise to all of the blood cells, including red blood cells, white blood cells, and platelets. These stem cells (discussed in Chapter 22) are located primarily in *hemopoietic* tissues—mainly the liver in fetuses and the bone marrow in adults. T cells develop in the thymus from precursor cells that migrate there from the hemopoietic tissues via the blood. In most mammals, including humans and mice, B cells develop from stem cells in the hemopoietic tissues themselves. Because they are sites where lymphocytes develop from precursor cells, the thymus and hemopoietic tissues are referred to as **central (primary) lymphoid organs**.

As we discuss later, most lymphocytes die in the central [lymphoid organ](#) soon after they develop, without ever functioning. Others, however, mature and migrate via the blood to the **peripheral (secondary) lymphoid organs**—mainly, the [lymph](#) nodes, spleen, and epithelium-associated lymphoid tissues in the gastrointestinal tract, respiratory tract, and skin (see [Figure 24-3](#)). As mentioned earlier, it is in the peripheral lymphoid organs that T cells and B cells react with foreign antigens.

T and B cells become morphologically distinguishable from each other only after they have been activated by [antigen](#). Nonactivated T and B cells look very similar, even in an [electron microscope](#). Both are small, only marginally bigger than red blood cells, and contain little [cytoplasm](#). Both are activated by antigen to 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](#). In contrast, effector T cells contain very little endoplasmic reticulum and do not secrete antibodies.

There are two main classes of T cells—*cytotoxic T cells* and *helper T cells*. Cytotoxic T cells kill infected cells, whereas helper T cells help activate macrophages, B cells, and cytotoxic T

cells. Effector helper T cells secrete a variety of signal proteins called [cytokines](#), which act as local mediators. They also display a variety of costimulatory proteins on their surface. By means of these cytokines and [membrane](#)-bound costimulatory proteins, they can influence the behavior of the various cell types they help. Effector cytotoxic T cells kill infected target cells also by means of proteins that they either secrete or display on their surface. Thus, whereas B cells can act over long distances by secreting antibodies that are distributed by the bloodstream, T cells can migrate to distant sites, but there they act only locally on neighboring cells.

The Adaptive Immune System Works by Clonal Selection

The most remarkable feature of the adaptive [immune system](#) is that it can respond to millions of different foreign antigens in a highly specific way. B cells, for example, make antibodies that react specifically with the [antigen](#) that induced their production. How do B cells produce such a diversity of specific antibodies? The answer began to emerge in the 1950s with the formulation of the [clonal selection theory](#). According to this theory, an animal first randomly generates a vast diversity of lymphocytes, and then those lymphocytes that can react against the foreign antigens that the animal actually encounters are specifically selected for action. As each [lymphocyte](#) develops in a central [lymphoid organ](#), it becomes committed to react with a particular antigen before ever being exposed to the antigen. It expresses this commitment in the form of cell-surface [receptor](#) proteins that specifically fit the antigen. When a lymphocyte encounters its antigen in a peripheral lymphoid organ, the binding of the antigen to the receptors activates the lymphocyte, causing it both to proliferate and to differentiate into an [effector cell](#). An antigen therefore selectively stimulates those cells that express [complementary](#) antigen-specific receptors and are thus already committed to respond to it. This arrangement is what makes adaptive **immune** responses antigen-specific.

The term “clonal” in [clonal selection theory](#) derives from the postulate that the adaptive [immune system](#) is composed of millions of different families, or clones, of lymphocytes, each consisting of T or B cells descended from a common ancestor. Each ancestral cell was already committed to make one particular [antigen](#)-specific [receptor protein](#), and so all cells in a [clone](#) have the same antigen specificity. According to the clonal selection theory, then, the **immune** system functions on the “ready-made” principle rather than the “made-to-order” one.

There is compelling evidence to support the main tenets of the [clonal selection theory](#). For example, when lymphocytes from an animal that has not been immunized are incubated in a test tube with a number of radioactively labeled antigens, only a very small proportion (less than 0.01%) bind each [antigen](#), suggesting that only a few cells are committed to respond to these antigens. Moreover, when one antigen is made so highly radioactive that it kills any cell that it binds to, the remaining lymphocytes can no longer produce an [immune response](#) to that particular antigen, even though they can still respond normally to other antigens. Thus, the committed lymphocytes must have receptors on their surface that specifically bind that antigen. Although most experiments of this kind have involved B cells and antibody responses, other experiments indicate that T cells, like B cells, operate by clonal selection.

How can the adaptive [immune system](#) produce lymphocytes that collectively display such an enormous diversity of receptors, including ones that recognize synthetic molecules that never occur in nature? We shall see later that the [antigen](#)-specific receptors on both T and B cells are encoded by genes that are assembled from a series of [gene](#) segments by a unique form

of [genetic recombination](#) that occurs early in a [lymphocyte's development](#), before it has encountered antigen. This assembly process generates the enormous diversity of receptors and lymphocytes, thereby enabling the **immune** system to respond to an almost unlimited diversity of antigens.

Most Antigens Activate Many Different Lymphocyte Clones

Most large molecules, including virtually all proteins and many polysaccharides, can serve as antigens. Those parts of an [antigen](#) that combine with the antigen-[binding site](#) on either an antibody [molecule](#) or a [lymphocyte receptor](#) are called [antigenic determinants](#) (or *epitopes*). Most antigens have a variety of antigenic determinants that can stimulate the production of antibodies, specific T cell responses, or both. Some determinants of an antigen produce a greater response than others, so that the [reaction](#) to them may dominate the overall response. Such determinants are said to be *immunodominant*.

The diversity of lymphocytes is such that even a single antigenic determinant is likely to activate many clones, each of which produces an [antigen-binding site](#) with its own characteristic affinity for the determinant. Even a relatively simple structure, like the *dinitrophenyl (DNP)* group in [Figure 24-9](#), can be “looked at” in many ways. When it is coupled to a [protein](#), as shown in the figure, it usually stimulates the production of hundreds of species of anti-DNP antibodies, each made by a different B cell [clone](#). Such responses are said to be *polyclonal*. When only a few clones are activated, the response is said to be *oligoclonal*; and when the response involves only a single B or T cell clone, it is said to be *monoclonal*. Monoclonal antibodies are widely used as tools in biology and medicine, but they have to be produced in a special way, as the responses to most antigens are polyclonal.

Immunological Memory Is Due to Both Clonal Expansion and Lymphocyte Differentiation

The adaptive [immune system](#), like the nervous system, can remember prior experiences. This is why we develop lifelong immunity to many common infectious diseases after our initial exposure to the [pathogen](#), and it is why vaccination works. The same phenomenon can be demonstrated in experimental animals. If an animal is immunized once with [antigen A](#), an [immune response](#) (either antibody or cell-mediated) appears after several days, 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 reinjected with antigen A, it will usually produce a [secondary immune response](#) that is very different from the primary response: the lag period is shorter, and the response is greater. 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 injection of antigen A, the response is typical of a primary, and not a secondary, **immune** response. The secondary response must therefore reflect antigen-specific [immunological memory](#) for antigen A.

The [clonal selection theory](#) provides a useful conceptual framework for understanding the cellular basis of [immunological memory](#). In an adult animal, the peripheral lymphoid organs contain a mixture of cells in at least three stages of maturation: *naïve cells*, *effector cells* and

memory cells. When **naïve cells** encounter [antigen](#) for the first time, some of them are stimulated to proliferate and differentiate into [effector cells](#), which are actively engaged in making a response (effector B cells secrete antibody, while effector T cells kill infected cells or help other cells fight the infection). Instead of becoming effector cells, some naïve cells are stimulated to multiply and differentiate into **memory cells**—cells that are not themselves engaged in a response but are more easily and more quickly induced to become effector cells by a later encounter with the same antigen. Memory cells, like naïve cells, give rise to either effector cells or more memory cells.

Thus, [immunological memory](#) is generated during the primary response in part because the proliferation of [antigen](#)-stimulated naïve cells creates many memory cells—a process known as *clonal expansion*—and in part because memory cells are able to respond more sensitively and rapidly to the same antigen than do naïve cells. And, unlike most effector cells, which die within days or weeks, memory cells can live for the lifetime of the animal, thereby providing lifelong immunological memory.

Acquired Immunological Tolerance Ensures That Self Antigens Are Not Attacked

As discussed in Chapter 25, cells of the innate [immune system](#) recognize molecules on the surface of pathogens that are not found in the host. The adaptive **immune** system has a far more difficult recognition task: it must be able to respond specifically to an almost unlimited number of foreign macromolecules, while avoiding responding to the large number of molecules made by the host organism itself. How does it do it? For one thing, self molecules do not induce the innate **immune** reactions that are required to activate adaptive **immune** responses. But even when an infection triggers innate reactions, self molecules still do not normally induce adaptive **immune** responses. Why not?

One answer is that the adaptive [immune system](#) “learns” not to respond to self antigens. Transplantation experiments provide one line of evidence for this learning process. When tissues are transplanted from one individual to another, as long as the two individuals are not identical twins, the **immune** system of the recipient usually recognizes the donor cells as foreign and destroys them. (For reasons we discuss later, the foreign antigens on the donor cells are so powerful that they can stimulate adaptive **immune** responses in the absence of infection or an adjuvant.) If, however, cells from one strain of mouse are introduced into a neonatal mouse of another strain, some of these cells survive for most of the recipient animal's life, and the recipient will now accept a graft from the original donor, even though it rejects “third-party” grafts. Apparently, nonself antigens can, in some circumstances, induce the **immune** system to become specifically unresponsive to them. This [antigen](#)-specific unresponsiveness to foreign antigens is known as [acquired immunological tolerance](#).

The unresponsiveness of an animal's adaptive [immune system](#) to its own macromolecules (*natural immunological tolerance*) is acquired in the same way. Normal mice, for example, cannot make an [immune response](#) against one of their own [protein](#) components of the [complement system](#) called C5 (discussed in Chapter 25). Mutant mice, however, that lack the [gene](#) encoding C5 (but are otherwise genetically identical to the normal mice) can make a strong **immune** response to this blood protein when immunized with it. Natural

immunological tolerance for a particular self [molecule](#) persists only for as long as the molecule remains present in the body. If a self molecule such as C5 is removed, an animal gains the ability to respond to it after a few weeks or months. Thus, the **immune** system is genetically capable of responding to self molecules but learns not to do so.

The learning process that leads to self-tolerance can involve killing the self-reactive lymphocytes (*clonal [deletion](#)*), functionally inactivating them (*clonal anergy* or *inactivation*), stimulating the cells to produce modified receptors that no longer recognize the self [antigen](#) (*[receptor](#) editing*), or the suppression of self-reactive lymphocytes by a special type of regulatory T cell. The process begins in the central lymphoid organs when newly formed self-reactive lymphocytes first encounter their self antigen. Instead of being activated by binding antigen, the immature lymphocytes are induced to either alter their receptors or die by [apoptosis](#). Lymphocytes that could potentially respond to self antigens that are not present in the central lymphoid organs often die or are either inactivated or suppressed after they have matured and migrated to peripheral lymphoid organs.

Why does the binding of self [antigen](#) lead to tolerance rather than activation? As we discuss later, for a [lymphocyte](#) to be activated in a peripheral [lymphoid organ](#), it must not only bind its antigen but must also receive a *costimulatory signal*. The latter signal is provided by a [helper T cell](#) in the case of a B lymphocyte and by an [antigen-presenting cell](#) in the case of a T lymphocyte. The production of costimulatory signals usually depends on exposure to pathogens, and so a self-reactive lymphocyte normally encounters its antigen in the absence of such signals. Without a costimulatory signal, an antigen tends to kill or inactivate a lymphocyte rather than activate it.

Tolerance to self antigens sometimes breaks down, causing T or B cells (or both) to react against the organism's own tissue antigens. *Myasthenia gravis* is an example of such an [autoimmune disease](#). Affected individuals make antibodies against the [acetylcholine](#) receptors on their own skeletal muscle cells. These antibodies interfere with the normal functioning of the receptors so that the patients become weak and may die because they cannot breathe. The mechanisms responsible for the breakdown of tolerance to self antigens in autoimmune diseases are unknown. It is thought, however, that activation of the innate [immune system](#) by infection may help trigger certain anti-self responses in genetically susceptible individuals.

Lymphocytes Continuously Circulate Through Peripheral Lymphoid Organs

Pathogens generally enter the body through an epithelial surface, usually through the skin, gut, or respiratory tract. How do the microbial antigens travel from these entry points to a peripheral [lymphoid organ](#), such as a [lymph](#) node or the spleen, where lymphocytes are activated (see [Figure 24-6](#))? The route and destination depend on the site of entry. Antigens that enter through the skin or respiratory tract are carried via the lymph to local lymph nodes; those that enter through the gut end up in gut-associated peripheral lymphoid organs such as Peyer's patches; and those that enter the blood are filtered out in the spleen. In most cases, dendritic cells carry the [antigen](#) from the site of infection to the peripheral lymphoid organ, where they become antigen-presenting cells, specialized for activating T cells (as we discuss later).

But the lymphocytes that can recognize a particular microbial [antigen](#) in a peripheral [lymph](#) organ are only a tiny fraction of the total [lymphocyte](#) population. How do these rare cells find an [antigen-presenting cell](#) displaying their antigen? The answer is that they continuously circulate between the lymph and blood until they encounter their antigen. In a lymph node, for example, lymphocytes continually leave the bloodstream by squeezing out between specialized endothelial cells lining small veins called *postcapillary venules*. After percolating through the node, they accumulate in small lymphatic vessels that leave the node and connect with other lymphatic vessels that pass through other lymph nodes downstream. Passing into larger and larger vessels, the lymphocytes eventually enter the main lymphatic vessel (the *thoracic duct*), which carries them back into the blood. This continuous recirculation between the blood and lymph ends only if a lymphocyte encounters its specific antigen (and a costimulatory signal) on the surface of an antigen-presenting cell in a peripheral [lymphoid organ](#). Now the lymphocyte is retained in the peripheral lymphoid organ, where it proliferates and differentiates into effector cells. Some of the effector T cells then leave the organ via the lymph and migrate through the blood to the site of infection.

Lymphocyte recirculation depends on specific interactions between the [lymphocyte](#) cell surface and the surface of the specialized endothelial cells lining the postcapillary venules in the peripheral lymphoid organs. Many cell types in the blood come into contact with these endothelial cells, but only lymphocytes adhere and then migrate out of the bloodstream. The lymphocytes initially adhere to the endothelial cells via *homing receptors* that bind to specific ligands (often called *counterreceptors*) on the [endothelial cell](#) surface. Lymphocyte migration into [lymph](#) nodes, for example, depends on a homing [receptor protein](#) called *L-selectin*, a member of the selectin family of cell-surface lectins discussed in Chapter 19. This protein binds to specific [sugar](#) groups on a counterreceptor that is expressed exclusively on the surface of the specialized endothelial cells in lymph nodes, causing the lymphocytes to adhere weakly to the endothelial cells and to roll slowly along their surface. The rolling continues until another, much stronger adhesion system is called into play by chemo-attractant proteins (called *chemokines*; see below) secreted by endothelial cells. This strong adhesion is mediated by members of the [integrin](#) family of cell adhesion molecules (discussed in Chapter 19), which become activated on the lymphocyte surface. Now the lymphocytes stop rolling and crawl out of the blood vessel into the lymph node.

[Chemokines](#) are small, secreted, positively charged proteins that have a central role in guiding the migrations of various types of white blood cells. They are all structurally related and bind to the surface of endothelial cells, and to negatively charged proteoglycans of the [extracellular matrix](#) in organs. By binding to [G-protein](#)-linked receptors (discussed in Chapter 15) on the surface of specific blood cells, chemokines attract these cells from the bloodstream into an organ, guide them to specific locations within the organ, and then help stop migration. (The AIDS [virus \(HIV\)](#) also binds to [chemokine](#) receptors, which allows the virus to infect white blood cells.) T and B cells initially enter the same region of a [lymph](#) node but are then attracted by different chemokines to separate regions of the node—T cells to the *paracortex* and B cells to *lymphoid follicles*. Unless they encounter their [antigen](#), both types of cells soon leave the lymph node via lymphatic vessels. If they encounter their antigen, however, they remain in the node, proliferate, and differentiate into either effector cells or memory cells. Most of the effector cells leave the node, expressing different chemokine receptors that help guide them to their new destinations—T cells to sites of infection and B cells to the bone marrow.

Summary

Innate **immune** responses are triggered at sites of infection by microbe-specific molecules associated with invading pathogens. In addition to fighting infection directly, these responses help activate adaptive **immune** responses in peripheral lymphoid organs. Unlike innate **immune** responses, adaptive responses provide specific and long-lasting protection against the particular [pathogen](#) that induced them.

The adaptive [immune system](#) is composed of millions of [lymphocyte](#) clones, with the cells in each [clone](#) sharing a unique cell-surface [receptor](#) that enables them to bind a particular [antigen](#). The binding of antigen to these receptors, however, is usually not sufficient to stimulate a lymphocyte to proliferate and differentiate into an [effector cell](#) that can help eliminate the [pathogen](#). Costimulatory signals provided by another specialized cell in a peripheral [lymphoid organ](#) are also required. Helper T cells provide such signals for B cells, while antigen-presenting dendritic cells usually provide them for T cells. Effector B cells secrete antibodies, which can act over long distances to help eliminate extracellular pathogens and their toxins. Effector T cells, by contrast, act locally at sites of infection to either kill infected host cells or help other cells to eliminate pathogens. As part of the [adaptive immune response](#), some lymphocytes proliferate and differentiate into memory cells, which are able to respond faster and more efficiently the next time the same pathogen invades. Lymphocytes that would react against self molecules are either induced to alter their receptors, induced to kill themselves, inactivated, or suppressed, so that the adaptive **immune** system normally reacts only against foreign antigens. Both B and T cells circulate continuously between 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.

Vertebrates inevitably die of infection if they are unable to make antibodies. Antibodies defend us against infection by binding to viruses and microbial toxins, thereby inactivating them. The binding of antibodies to invading pathogens also recruits various types of white blood cells and a system of blood proteins, collectively called *complement* (discussed in Chapter 25). The white blood cells and activated complement components work together to attack the invaders.

Synthesized exclusively by B cells, antibodies are produced in billions of forms, each with a different [amino acid](#) sequence and a different [antigen-binding site](#). Collectively called [immunoglobulins](#) (abbreviated as **Ig**), they are among the most abundant [protein](#) components in the blood, constituting about 20% of the total protein in plasma by weight. Mammals make five classes of antibodies, each of which mediates a characteristic biological response following antigen binding. In this [section](#), we discuss the structure and function of antibodies and how they interact with antigen.

B Cells and Antibodies

Vertebrates inevitably die of infection if they are unable to make antibodies. Antibodies defend us against infection by binding to viruses and microbial toxins, thereby inactivating them. The binding of antibodies to invading pathogens also recruits various types of white blood cells and a system of blood proteins, collectively called *complement* (discussed in

Chapter 25). The white blood cells and activated complement components work together to attack the invaders.

Synthesized exclusively by B cells, antibodies are produced in billions of forms, each with a different [amino acid](#) sequence and a different [antigen-binding site](#). Collectively called [immunoglobulins](#) (abbreviated as [Ig](#)), they are among the most abundant [protein](#) components in the blood, constituting about 20% of the total protein in plasma by weight. Mammals make five classes of antibodies, each of which mediates a characteristic biological response following antigen binding. In this [section](#), we discuss the structure and function of antibodies and how they interact with antigen.

B Cells Make Antibodies as Both Cell-Surface Receptors and Secreted Molecules

As predicted by the [clonal selection theory](#), all antibody molecules made by an individual B cell have the same [antigen-binding site](#). The first antibodies made by a newly formed B cell are not secreted. Instead, they are inserted into the [plasma membrane](#), where they serve as receptors for antigen. Each B cell has approximately 10^5 such receptors in its plasma membrane. As we discuss later, each of these receptors is stably associated with a [complex](#) of transmembrane proteins that activate intracellular signaling pathways when antigen binds to the [receptor](#).

Each B cell produces a single species of antibody, each with a unique [antigen-binding site](#). When a naïve or memory B cell is activated by antigen (with the aid of a [helper T cell](#)), it proliferates and differentiates into an antibody-secreting [effector cell](#). Such cells make and secrete large amounts of soluble (rather than [membrane-bound](#)) antibody, which has the same unique antigen-binding site as the cell-surface antibody that served earlier as the antigen [receptor](#). Effector B cells can begin secreting antibody while they are still small lymphocytes, but the end stage of their maturation pathway is a large *plasma cell* (see [Figure 24-7B](#)), which continuously secretes antibodies at the astonishing rate of about 2000 molecules per second. Plasma cells seem to have committed so much of their [protein-synthesizing](#) machinery to making antibody that they are incapable of further growth and division. Although many die after several days, some survive in the bone marrow for months or years and continue to secrete antibodies into the blood.

A Typical Antibody Has Two Identical Antigen-Binding Sites

The simplest antibodies are Y-shaped molecules with two identical [antigen-binding](#) sites, one at the tip of each arm of the Y. Because of their two antigen-binding sites, they are described as [bivalent](#). As long as an antigen has three or more antigenic determinants, bivalent antibody molecules can cross-link it into a large lattice. This lattice can be rapidly phagocytosed and degraded by macrophages. The efficiency of antigen binding and cross-linking is greatly increased by a flexible *hinge region* in most antibodies, which allows the distance between the two antigen-binding sites to vary.

The protective effect of antibodies is not due simply to their ability to bind [antigen](#). They engage in a variety of activities that are mediated by the tail of the Y-shaped [molecule](#). As we discuss later, antibodies with the same antigen-binding sites can have any one of several different tail regions. Each type of tail region gives the antibody different functional properties, such as the ability to activate the [complement system](#), to bind to phagocytic cells, or to cross the placenta from mother to fetus.

An Antibody Molecule Is Composed of Heavy and Light Chains

The [basic](#) structural unit of an antibody [molecule](#) consists of four [polypeptide](#) chains, two identical **light (L) chains** (each containing about 220 amino acids) and two identical **heavy (H) chains** (each usually containing about 440 amino acids). The four chains are held together by a combination of noncovalent and covalent (disulfide) bonds. The molecule is composed of two identical halves, each with the same [antigen-binding site](#). Both light and heavy chains usually cooperate to form the antigen-binding surface.

There Are Five Classes of Heavy Chains, Each With Different Biological Properties

In mammals, there are five *classes* of antibodies, IgA, IgD, IgE, IgG, and IgM, each with its own class of heavy chain— α , δ , ϵ , γ , and μ , respectively. IgA molecules have α chains, IgG molecules have γ chains, and so on. In addition, there are a number of subclasses of IgG and IgA immunoglobulins; for example, there are four human IgG subclasses (IgG1, IgG2, IgG3, and IgG4), having γ_1 , γ_2 , γ_3 , and γ_4 heavy chains, respectively. The various heavy chains give a distinctive [conformation](#) to the hinge and tail regions of antibodies, so that each class (and subclass) has characteristic properties of its own.

IgM, which has μ heavy chains, is always the first class of antibody made by a developing B cell, although many B cells eventually switch to making other classes of antibody (discussed below). The immediate precursor of a B cell, called a [pre-B cell](#), initially makes μ chains, which associate with so-called *surrogate light chains* (substituting for genuine light chains) and insert into the [plasma membrane](#). The complexes of μ chains and surrogate light chains are required for the cell to progress to the next stage of [development](#), where it makes bona fide light chains. The light chains combine with the μ chains, replacing the surrogate light chains, to form four-chain IgM molecules (each with two μ chains and two light chains). These molecules then insert into the plasma membrane, where they function as receptors for [antigen](#). At this point, the cell is called an *immature naïve B cell*. After leaving the bone marrow, the cell starts to produce cell-surface **IgD** molecules as well, with the same [antigen-binding site](#) as the IgM molecules. It is now called a *mature naïve B cell*. It is this cell that can respond to foreign antigen in peripheral lymphoid organs.

IgM is not only the first class of antibody to appear on the surface of a developing B cell. It is also the major class secreted into the blood in the early stages of a *primary* antibody response, on first exposure to an [antigen](#). (Unlike IgM, IgD molecules are secreted in only small amounts and seem to function mainly as cell-surface receptors for antigen.) In its secreted form, IgM is a pentamer composed of five four-chain units, giving it a total of 10 antigen-binding sites. Each pentamer contains one copy of another [polypeptide](#) chain, called a

J (joining) chain. The J chain is produced by IgM-secreting cells and is covalently inserted between two adjacent tail regions.

The binding of an [antigen](#) to a single secreted pentameric IgM [molecule](#) can activate the [complement system](#). As discussed in Chapter 25, when the antigen is on the surface of an invading [pathogen](#), this activation of complement can either mark the pathogen for [phagocytosis](#) or kill it directly.

The major class of immunoglobulin in the blood is **IgG**, which is a four-chain [monomer](#) produced in large quantities during *secondary immune* responses. Besides activating complement, the tail region of an IgG [molecule](#) binds 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.

IgG molecules are the only antibodies that can pass from mother to fetus via the placenta. Cells of the placenta that are in contact with maternal blood have Fc receptors that bind blood-borne IgG molecules and direct their passage to the fetus. The antibody molecules bound to the receptors are first taken into the placental cells by [receptor-mediated endocytosis](#). They are then transported across the cell in vesicles and released by [exocytosis](#) into the fetal blood (a process called [transcytosis](#), discussed in Chapter 13). Because other classes of antibodies do not bind to these particular Fc receptors, they cannot pass across the placenta. IgG is also secreted into the mother's milk and is taken up from the gut of the neonate into the blood, providing protection for the baby against infection.

IgA is the principal class of antibody in secretions, including saliva, tears, milk, and respiratory and intestinal secretions. Whereas IgA is a four-chain [monomer](#) in the blood, it is an eight-chain dimer in secretions. It is transported through secretory epithelial cells from the extracellular fluid into the secreted fluid by another type of [Fc receptor](#) that is unique to secretory. This Fc receptor can also transport IgM into secretions (but less efficiently), which is probably why individuals with a selective IgA deficiency, the most common form of antibody deficiency, are only mildly affected by the defect.

The tail region of **IgE** molecules, which are four-chain monomers, binds with unusually high affinity ($K_a \sim 10^{10}$ liters/[mole](#)) to yet another class of Fc receptors. These receptors are located on the surface of *mast cells* in tissues and of *basophils* in the blood. The IgE molecules bound to them function as passively acquired receptors for [antigen](#). Antigen binding triggers the mast cell or basophil to secrete a variety of cytokines and biologically active amines, especially *histamine*. These molecules cause blood vessels to dilate and become leaky, which in turn helps white blood cells, antibodies, and complement components to enter sites of infection. The same molecules are also largely responsible for the symptoms of such *allergic* reactions as hay fever, asthma, and hives. In addition, mast cells secrete factors that attract and activate white blood cells called *eosinophils*. These cells also have Fc receptors that bind IgE molecules and can kill various types of parasites, especially if the parasites are coated with IgE antibodies.

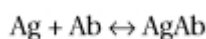
In addition to the five classes of heavy chains found in antibody molecules, higher vertebrates have two types of light chains, κ and λ , which seem to be functionally indistinguishable. Either type of [light chain](#) may be associated with any of the heavy chains. An individual antibody [molecule](#), however, always contains identical light chains and identical heavy

chains: an IgG molecule, for instance, may have either κ or λ light chains, but not one of each. As a result of this symmetry, an antibody's [antigen](#)-binding sites are always identical. Such symmetry is crucial for the cross-linking function of secreted antibodies.

of an Antibody-Antigen Interaction Depends on Both the Number and the Affinity of the Antigen-Binding Sites

The binding of an [antigen](#) to antibody, like the binding of a [substrate](#) to an [enzyme](#), is reversible. It is mediated by the sum of many relatively weak non-covalent forces, including hydrogen bonds and hydrophobic van der Waals forces, and ionic interactions. These weak forces are effective only when the antigen [molecule](#) is close enough to allow some of its atoms to fit into [complementary](#) recesses on the surface of the antibody. The complementary regions of a four-chain antibody unit are its two identical antigen-binding sites; the corresponding region on the antigen is an *antigenic determinant*. Most antigenic macromolecules have many different antigenic determinants and are said to be *multivalent*; if two or more of them are identical (as in a [polymer](#) with a repeating structure), the antigen is said to be *polyvalent*.

The reversible binding [reaction](#) between an [antigen](#) with a single antigenic determinant (denoted Ag) and a single antigen-[binding site](#) (denoted Ab) can be expressed as



The [equilibrium](#) point depends both on the concentrations of Ab and Ag and on the strength of their interaction. Clearly, a larger fraction of Ab will become associated with Ag as the concentration of Ag increases. The strength of the interaction is generally expressed as the **affinity constant** (K_a) (see [Figure 3-44](#)), where

$$K_a = \frac{[\text{AgAb}]}{[\text{Ag}][\text{Ab}]}$$

(the square brackets indicate the concentration of each component at [equilibrium](#)).

The affinity constant, sometimes called the [association constant](#), can be [determined](#) by measuring the concentration of free Ag required to fill half of the [antigen](#)-binding sites on the antibody. When half the sites are filled, $[\text{AgAb}] = [\text{Ab}]$ and $K_a = 1/[\text{Ag}]$. Thus, the reciprocal of the antigen concentration that produces half the maximum binding is equal to the affinity constant of the antibody for the antigen. Common values range from as low as 5×10^4 to as high as 10^{11} liters/[mole](#).

The **affinity** of an antibody for an antigenic determinant describes the strength of binding of a single copy of the antigenic determinant to a single [antigen-binding site](#), and it is independent of the number of sites. When, however, a polyvalent antigen, carrying multiple copies of the same antigenic determinant, combines with a polyvalent antibody, the binding strength is greatly increased because all of the antigen-antibody bonds must be broken simultaneously before the antigen and antibody can dissociate. As a result, a typical IgG [molecule](#) can bind at least 100 times more strongly to a polyvalent antigen if both antigen-binding sites are engaged than if only one site is engaged. The total binding strength of a polyvalent antibody with a polyvalent antigen is referred to as the [avidity](#) of the interaction.

If the affinity of the [antigen](#)-binding sites in an IgG and an IgM [molecule](#) is the same, the IgM molecule (with 10 binding sites) will have a much greater [avidity](#) for a multivalent antigen than an IgG molecule (which has two binding sites). This difference in avidity, often 10^4 -fold or more, is important because antibodies produced early in an [immune response](#) usually have much lower affinities than those produced later. Because of its high total avidity, IgM—the major [Ig](#) class produced early in **immune** responses—can function effectively even when each of its binding sites has only a low affinity.

So far we have considered the general structure and function of antibodies. Next we look at the details of their structure, as revealed by studies of their [amino acid](#) sequence and three-dimensional structure.

Light and Heavy Chains Consist of Constant and Variable Regions

Comparison of the [amino acid](#) sequences of different antibody molecules reveals a striking feature with important genetic implications. Both light and heavy chains have a variable sequence at their N-terminal ends but a constant sequence at their C-terminal ends. Consequently, when the amino acid sequences of many different κ chains are compared, the C-terminal halves are the same or show only minor differences, whereas the N-terminal halves are all very different. Light chains have a **constant region** about 110 amino acids long and a [variable region](#) of the same size. The [variable region](#) of the heavy chains (at their N-terminus) is also about 110 amino acids long, but the heavy-chain constant region is about three or four times longer (330 or 440 amino acids), depending on the class.

It is the N-terminal ends of the light and heavy chains that come together to form the [antigen-binding site](#), and the variability of their [amino acid](#) sequences provides the structural basis for the diversity of antigen-binding sites. The diversity in the variable regions of both light and heavy chains is for the most part restricted to three small [hypervariable regions](#) in each chain; the remaining parts of the [variable region](#), known as *framework regions*, are relatively constant. Only the 5–10 amino acids in each [hypervariable region](#) form the antigen-binding site. As a result, the size of the antigenic determinant that an antibody recognizes is generally comparably small. It can consist of fewer than 25 amino acids on the surface of a [globular protein](#), for example.

The Light and Heavy Chains Are Composed of Repeating Ig Domains

Both light and heavy chains are made up of repeating segments—each about 110 amino acids long and each containing one intrachain disulfide bond. These repeating segments fold independently to form compact functional units called [immunoglobulin \(Ig\) domains](#). As shown in [Figure 24-32](#), a [light chain](#) consists of one variable (V_L) and one constant (C_L) [domain](#). These domains pair with the variable (V_H) and first constant (C_{H1}) domain of the heavy chain to form the [antigen](#)-binding region. The remaining constant domains of the heavy chains form the Fc region, which determines the other biological properties of the antibody. Most heavy chains have three constant domains (C_{H1} , C_{H2} , and C_{H3}), but those of IgM and IgE antibodies have four.

The similarity in their domains suggests that antibody chains arose during evolution by a series of [gene](#) duplications, beginning with a primordial gene coding for a single 110 [amino acid domain](#) of unknown function. This hypothesis is supported by the finding that each domain of the constant region of a heavy chain is encoded by a separate coding sequence ([exon](#)).

From Hypervariable Loops

A number of fragments of antibodies, as well as intact antibody molecules, have been studied by x-ray crystallography. From these examples, we can understand the way in which billions of different [antigen](#)-binding sites are constructed on a common structural theme.

As illustrated, each [Ig domain](#) has a very similar three-dimensional structure based on what is called the *immunoglobulin fold*, which consists of a sandwich of two β sheets held together by a disulfide bond. We shall see later that many other proteins on the surface of lymphocytes and other cells, many of which function as cell-cell adhesion molecules (discussed in Chapter 19), contain similar domains and hence are members of a very large [immunoglobulin \(Ig\) superfamily](#) of proteins.

of a constant domain is shown on the left and of a variable domain on the right.
Both ([more...](#))

The variable domains of antibody molecules are unique in that each has its particular set of three hypervariable regions, which are arranged in three *hypervariable loops*. The hypervariable loops of both the light and heavy variable domains are clustered together to form the [antigen-binding site](#). Because the [variable region](#) of an antibody [molecule](#) consists of a highly conserved rigid framework, with hypervariable loops attached at one end, an enormous diversity of antigen-binding sites can be generated by changing only the lengths and [amino acid](#) sequences of the hypervariable loops. The overall three-dimensional structure necessary for antibody function remains constant.

X-ray analyses of crystals of antibody fragments bound to an antigenic determinant reveal exactly how the hypervariable loops of the light and heavy variable domains cooperate to form an [antigen](#)-binding surface in particular cases. The dimensions and shape of each different site vary depending on the conformations of the [polypeptide](#) chain in the hypervariable loops, which in turn are [determined](#) by the sequences of the [amino acid](#) side chains in the loops. The shapes of binding sites vary greatly—from pockets, to grooves, to undulating flatter surfaces, and even to protrusions—depending on the antibody. Smaller ligands tend to bind to deeper pockets, whereas larger ones tend to bind to flatter surfaces. In addition, the [binding site](#) can alter its shape after antigen binding to better fit the [ligand](#).

Summary

Antibodies defend vertebrates against infection by inactivating viruses and microbial toxins and by recruiting the [complement system](#) and various types of white blood cell to kill the invading pathogens. A typical antibody [molecule](#) is Y-shaped, with two identical [antigen](#)-binding sites at the tips of the Y and binding sites for complement components and/or various cell-surface receptors on the tail of the Y.

Each B cell [clone](#) makes antibody molecules with a unique [antigen-binding site](#). Initially, during B cell [development](#) in the bone marrow, the antibody molecules are inserted into the [plasma membrane](#), where they serve as receptors for antigen. In peripheral lymphoid organs, antigen binding to these receptors, together with costimulatory signals provided by helper T cells, activates the B cells to proliferate and differentiate into either memory cells or antibody-secreting effector cells. The effector cells secrete antibodies with the same unique antigen-binding site as the membrane-bound antibodies.

A typical antibody [molecule](#) is composed of four [polypeptide](#) chains, two identical heavy chains and two identical light chains. Parts of both the heavy and light chains usually combine to form the [antigen](#)-binding sites. There are five classes of antibodies (IgA, IgD, IgE, IgG, and IgM), each with a distinctive heavy chain (α , δ , ϵ , γ , and μ , respectively). The heavy chains also form the tail (Fc region) of the antibody, which determines what other proteins will bind to the antibody and therefore what biological properties the antibody class has. Either type of [light chain](#) (κ or λ) can be associated with any class of heavy chain, but the type of light chain does not seem to influence the properties of the antibody, other than its specificity for antigen.

Each light and heavy chain is composed of a number of [Ig domains](#)— [\$\beta\$ sheet](#) structures containing about 110 amino acids. A [light chain](#) has one variable (V_L) and one constant (C_L) [domain](#), while a heavy chain has one variable (V_H) and three or four constant (C_H) domains. The [amino acid](#) sequence variation in the variable domains of both light and heavy chains is mainly confined to several small hypervariable regions, which protrude as loops at one end of the domains to form the [antigen-binding site](#).

The Generation of Antibody Diversity

Even in the absence of [antigen](#) stimulation, a human can probably make more than 10^{12} different antibody molecules—its *preimmune antibody repertoire*. Moreover, the antigen-binding sites of many antibodies can cross-react with a variety of related but different antigenic determinants, making the antibody defense force even more formidable. The preimmune repertoire is apparently large enough to ensure that there will be an [antigen-binding site](#) to fit almost any potential antigenic determinant, albeit with low affinity. After repeated stimulation by antigen, B cells can make antibodies that bind their antigen with much higher affinity—a process called [affinity maturation](#). Thus, antigen stimulation greatly increases the antibody arsenal.

Antibodies are proteins, and proteins are encoded by genes. Antibody diversity therefore poses a special genetic problem: how can an animal make more antibodies than there are genes in its [genome](#)? (The human genome, for example, contains fewer than 50,000 genes.) This problem is not quite as formidable as it might first appear. Recall that the variable regions of both the light and heavy chains of antibodies usually form the [antigen-binding site](#).

Thus, an animal with 1000 genes encoding light chains and 1000 genes encoding heavy chains could, in principle, combine their products in 1000×1000 different ways to make 10^6 different antigen-binding sites (although, in reality, not every [light chain](#) can combine with every heavy chain to make an antigen-binding site). Nonetheless, the mammalian [immune system](#) has evolved unique genetic mechanisms that enable it to generate an almost unlimited number of different light and heavy chains in a remarkably economical way, by joining separate [gene segments](#) together before they are transcribed. Birds and fish use very different strategies for diversifying antibodies, and even sheep and rabbits use somewhat different strategies from mice and humans. We shall confine our discussion to the mechanisms used by mice and humans.

We begin this [section](#) by discussing the mechanisms that B cells use to produce antibodies with an enormous diversity of [antigen](#)-binding sites. We then consider how a B cell can alter the tail region of the antibody it makes, while keeping the [antigen-binding site](#) unchanged. This ability allows the B cell to switch from making [membrane](#)-bound antibody to making secreted antibody, or from making one class of antibody to making another, all without changing the antigen-specificity of the antibody.

Antibody Genes Are Assembled From Separate Gene Segments During B Cell Development

The first direct evidence that [DNA](#) is rearranged during B cell [development](#) came in the 1970s from experiments in which molecular biologists compared DNA from early mouse embryos, which do not make antibodies, with the DNA of a mouse B cell tumor, which makes a single species of antibody [molecule](#). The specific variable (V)-region and constant (C)-region coding sequences that the tumor cells used were present on the same DNA restriction fragment in the tumor cells but on two different restriction fragments in the embryos. This showed that the DNA sequences encoding an antibody molecule are rearranged at some stage in B cell development.

We now know that each type of antibody chain— κ light chains, λ light chains, and heavy chains—has a separate pool of [gene segments](#) and exons from which a single [polypeptide](#) chain is eventually synthesized. Each pool is on a different [chromosome](#) and contains a large number of gene segments encoding the V region of an antibody chain and, as we saw in [Figure 24-33](#), a smaller number of exons encoding the C region. During the [development](#) of a B cell, a complete coding sequence for each of the two antibody chains to be synthesized is assembled by site-specific [genetic recombination](#) (discussed in Chapter 5). In addition to bringing together the separate gene segments and the C-region exons of the antibody gene, these rearrangements also activate transcription from the gene [promoter](#) through changes in the relative positions of the enhancers and silencers acting on the promoter. Thus, a complete antibody chain can be synthesized only after the [DNA](#) has been rearranged. As we shall see, the process of joining gene segments contributes to the diversity of [antigen](#)-binding sites in several ways.

Each Variable Region Is Encoded by More Than One Gene Segment

When [genomic DNA](#) sequences encoding V and C regions were first analyzed, it was found that a single region of DNA encodes the C region of an antibody chain (see [Figure 24-33](#)), but two or more regions of DNA have to be assembled to encode each V region. Each light-chain V region is encoded by a DNA sequence assembled from two [gene](#) segments—a long **V gene segment** and a short *joining*, or **J gene segment** (not to be confused with the [protein J chain](#) (see [Figure 24-23](#)), which is encoded elsewhere in the [genome](#)). The V-J joining process involved in making a human κ light chain. In the “germ-line” DNA (where the antibody genes are not being expressed and are therefore not rearranged), the cluster of five J gene segments is separated from [\(more...\)](#)

Each heavy-chain V region is encoded by a [DNA](#) sequence assembled from three [gene](#) segments—a V segment, a J segment, and a *diversity segment*, or **D gene segment**

The large number of inherited V, J, and D [gene](#) segments available for encoding antibody chains makes a substantial contribution on its own to antibody diversity, but the combinatorial joining of these segments (called *combinatorial diversification*) greatly increases this contribution. Any of the 40 V segments in the human κ light-chain gene-segment pool, for example, can be joined to any of the 5 J segments, so that at least 200 (40×5) different κ -chain V regions can be encoded by this pool. Similarly, any of the 51 V segments in the human heavy-chain pool can be joined to any of the 6 J segments and any of the 27 D segments to encode at least 8262 ($51 \times 6 \times 27$) different heavy-chain V regions.

The combinatorial diversification resulting from the assembly of different combinations of inherited V, J, and D [gene](#) segments just discussed is an important mechanism for diversifying the [antigen](#)-binding sites of antibodies. By this mechanism alone, a human can produce 287 different V_L regions (200 κ and 116 λ) and 8262 different V_H regions. In principle, these could then be combined to make about 2.6×10^6 (316×8262) different antigen-binding sites. In addition, as we discuss next, the joining mechanism itself greatly increases this number of possibilities (probably more than 10^8 -fold), making it much greater than the total number of B cells (about 10^{12}) in a human.

Imprecise Joining of Gene Segments Greatly Increases the Diversity of V Regions

During B cell [development](#), the V and J [gene](#) segments (for the [light chain](#)) and the V, D, and J gene segments (for the heavy chain) are joined together to form a functional V_L - or V_H -region coding sequence by a process of [site-specific recombination](#) called **V(D)J joining**. Conserved [DNA](#) sequences flank each gene segment and serve as recognition sites for the joining process, ensuring that only appropriate gene segments recombine. Thus, for example, a V segment will always join to a J or D segment but not to another V segment. Joining is mediated by an [enzyme complex](#) called the **V(D)J recombinase**. This complex contains two proteins that are specific to developing lymphocytes, as well as enzymes that help repair damaged DNA in all our cells.

The [lymphocyte](#)-specific proteins of the V(D)J recombinase are encoded by two closely linked genes called *rag-1* and *rag-2* (*rag* = recombination activating genes). The **RAG proteins** introduce double-strand breaks at the flanking [DNA](#) sequences, and this is followed

by a rejoining process that is mediated by both the RAG proteins and the enzymes involved in general DNA double-strand repair (discussed in Chapter 5). Thus, if both *rag* genes are artificially expressed in a [fibroblast](#), the fibroblast is now able to rearrange experimentally introduced antibody [gene](#) segments just as a developing B cell normally does. Moreover, individuals who are deficient in either *rag* gene or in one of the general repair enzymes are highly susceptible to infection because they are unable to carry out *V(D)J* joining and consequently do not have functional B or T cells. (T cells use the same recombinase to assemble the gene segments that encode their [antigen](#)-specific receptors.)

In most cases of [site-specific recombination](#), [DNA](#) joining is precise. But during the joining of antibody (and T cell [receptor](#)) [gene](#) segments, a variable number of nucleotides are often lost from the ends of the recombining gene segments, and one or more randomly chosen nucleotides may also be inserted. This random loss and gain of nucleotides at joining sites is called **junctional diversification**, and it enormously increases the diversity of V-region coding sequences created by recombination, specifically in the third [hypervariable region](#). This increased diversification comes at a price, however. In many cases, it will result in a shift in the [reading frame](#) that produces a nonfunctional gene. Because roughly two in every three rearrangements are “nonproductive” in this way, many developing B cells never make a functional antibody [molecule](#) and consequently die in the bone marrow. B cells making functional antibody molecules that bind strongly to self antigens in the bone marrow are stimulated to re-express the RAG proteins and undergo a second round of *V(D)J* rearrangements, thereby changing the specificity of the cell-surface antibody they make—a process referred to as **receptor editing**. Self-reactive B cells that fail to change their specificity in this way are eliminated through the process of clonal [deletion](#).

Antigen-Driven Somatic Hypermutation Fine-Tunes Antibody Responses

As mentioned earlier, with the passage of time after immunization, there is usually a progressive increase in the affinity of the antibodies produced against the immunizing [antigen](#). This phenomenon, known as [affinity maturation](#), is due to the accumulation of point mutations specifically in both heavy-chain and light-chain V-region coding sequences. The mutations occur long after the coding regions have been assembled, when B cells are stimulated by antigen and helper T cells to generate memory cells in a lymphoid follicle in a peripheral [lymphoid organ](#) (see [Figure 24-16](#)). They occur at the rate of about one per V-region coding sequence per cell generation. Because this is about a million times greater than the spontaneous [mutation rate](#) in other genes, the process is called **somatic hypermutation**. The molecular mechanism is still uncertain, but it is believed to involve some form of error-prone [DNA repair](#) process targeted to the rearranged V-region coding sequence by specific regions of DNA brought together by *V(D)J* joining. Surprisingly, an [enzyme](#) involved in [RNA editing](#) is required, but its function in the hypermutation process is unknown.

Only a small minority of the altered [antigen](#) receptors generated by hypermutation have an increased affinity for the antigen. The few B cells expressing these higher-affinity receptors, however, are preferentially stimulated by the antigen to survive and proliferate, whereas most other B cells die by [apoptosis](#). Thus, as a result of repeated cycles of somatic hypermutation, followed by antigen-driven proliferation of selected clones of memory B cells, antibodies of

increasingly higher affinity become abundant during an [immune response](#), providing progressively better protection against the [pathogen](#).

The Control of V(D)J Joining Ensures That B Cells Are Monospecific

As the [clonal selection theory](#) predicts, B cells are *monospecific*. That is, all the antibodies that any one B cell produces have identical [antigen](#)-binding sites. This property enables antibodies to cross-link antigens into large aggregates, thereby promoting antigen elimination. It also means that an activated B cell secretes antibodies with the same specificity as that of the [membrane](#)-bound antibody on the B cell that was originally stimulated.

The requirement of monospecificity means that each B cell can make only one type of V_L region and one type of V_H region. Since B cells, like most other somatic cells, are [diploid](#), each cell has six [gene](#)-segment pools encoding antibody chains: two heavy-chain pools (one from each parent) and four light-chain pools (one κ and one λ from each parent). If [DNA](#) rearrangements occurred independently in each heavy-chain pool and each light-chain pool, a single cell could make up to eight different antibodies, each with a different [antigen-binding site](#).

In fact, however, each B cell uses only two of the six [gene](#)-segment pools: one of the two heavy-chain pools and one of the four light-chain pools. Thus, each B cell must choose not only between its κ and λ light-chain pools, but also between its maternal and paternal light-chain and heavy-chain pools. This second choice is called [allelic exclusion](#), and it also occurs in the [expression](#) of genes that encode T cell receptors. For most other proteins that are encoded by autosomal genes, both maternal and paternal genes in a cell are expressed about equally.

Allelic exclusion and κ versus λ light-chain choice during B cell [development](#) depend on negative feedback regulation of the $V(D)J$ joining process. A functional rearrangement in one [gene](#)-segment pool suppresses rearrangements in all remaining pools that encode the same type of [polypeptide](#) chain ([Figure 24-40](#)). In B cell clones isolated from transgenic mice expressing a rearranged μ -chain gene, for example, the rearrangement of endogenous heavy-chain genes is usually suppressed. Comparable results have been obtained for light chains. The suppression does not occur if the product of the rearranged gene fails to assemble into a [receptor](#) that inserts into the [plasma membrane](#). It has therefore been proposed that either the receptor assembly process itself or extracellular signals that act on the receptor are involved in the suppression of further gene rearrangements.

Although no biological differences between the constant regions of κ and λ light chains have been discovered, there is an advantage in having two separate pools of [gene](#) segments encoding [light chain](#) variable regions. Having two separate pools increases the chance that a [pre-B cell](#) that has successfully assembled a V_H -region coding sequence will go on to assemble successfully a V_L -region coding sequence to become a B cell. This chance is further increased because, before a developing pre-B cell produces ordinary light chains, it makes surrogate light, which assemble with μ heavy chains. The resulting receptors are displayed on the cell surface and allow the cell to proliferate, producing large numbers of progeny cells, some of which are likely to succeed in producing bona fide light chains.

When Activated by Antigen, a B Cell Switches From Making a Membrane-Bound Antibody to Making a Secreted Form of the Same Antibody

We now turn from the genetic mechanisms that determine the [antigen-binding site](#) of an antibody to those that determine its biological properties—that is, those that determine what form of heavy-chain constant region is synthesized. The choice of the particular [gene](#) segments that encode the antigen-binding site is a commitment for the life of a B cell and its progeny, but the type of C_H region that is made changes during B cell [development](#). The changes are of two types: changes from a [membrane](#)-bound form to a secreted form of the same C_H region and changes in the class of the C_H region made.

All classes of antibody can be made in a [membrane](#)-bound form, as well as in a soluble, secreted form. The membrane-bound form serves as an [antigen receptor](#) on the B cell surface, while the soluble form is made only after the cell is activated by antigen to become an antibody-secreting [effector cell](#). The sole difference between the two forms resides in the C-terminus of the heavy chain. The heavy chains of membrane-bound antibody molecules have a hydrophobic C-terminus, which anchors them in the [lipid bilayer](#) of the B cell's [plasma membrane](#). The heavy chains of secreted antibody molecules, by contrast, have instead a [hydrophilic](#) C-terminus, which allows them to escape from the cell. The switch in the character of the antibody molecules made occurs because the activation of B cells by antigen (and helper T cells) induces a change in the way in which the H-chain [RNA](#) transcripts are made and processed in the [nucleus](#).

B Cells Can Switch the Class of Antibody They Make

During B cell [development](#), many B cells switch from making one class of antibody to making another—a process called [class switching](#). All B cells begin their antibody-synthesizing lives by making IgM molecules and inserting them into the [plasma membrane](#) as receptors for [antigen](#). After the B cells leave the bone marrow, but before they interact with antigen, they switch and make both IgM and IgD molecules as membrane-bound antigen receptors, both with the same antigen-binding sites. On stimulation by antigen and helper T cells, some of these cells are activated to secrete IgM antibodies, which dominate the primary antibody response. Later in the [immune response](#), the combination of antigen and the cytokines that helper T cells secrete induce many B cells to switch to making IgG, IgE, or IgA antibodies. These cells generate both memory cells that express the corresponding classes of antibody molecules on their surface and effector cells that secrete the antibodies. The IgG, IgE, and IgA molecules are collectively referred to as *secondary* classes of antibodies, both because they are produced only after antigen stimulation and because they dominate secondary antibody responses. As we saw earlier, each different class of antibody is specialized to attack microbes in different ways and in different sites.

The constant region of an antibody heavy chain determines the class of the antibody. Thus, the ability of B cells to switch the class of antibody they make without changing the [antigen-binding site](#) implies that the same assembled V_H -region coding sequence (which specifies the antigen-binding part of the heavy chain) can sequentially associate with different C_H -coding sequences. This has important functional implications. It means that, in an individual animal, a particular antigen-binding site that has been selected by environmental antigens can be

distributed among the various classes of antibodies, thereby acquiring the different biological properties of each class.

When a B cell switches from making IgM and IgD to one of the secondary classes of antibody, an irreversible change at the [DNA](#) level occurs—a process called *class switch recombination*. It entails [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. Switch recombination differs from $V(D)J$ joining in several ways: (1) it involves noncoding sequences only and therefore leaves the coding sequence unaffected; (2) it uses different flanking recombination sequences and different enzymes; (3) it happens after [antigen](#) stimulation; and (4) it is dependent on helper T cells.

Summary

Antibodies are produced from three pools of [gene](#) segments and exons. One pool encodes κ light chains, one encodes λ light chains, and one encodes heavy chains. In each pool, separate gene segments that code for different parts of the [variable region](#) of the light or heavy chains are brought together by [site-specific recombination](#) during B cell [development](#). The light-chain pools contain one or more constant- (C -) region exons and sets of variable (V) and joining (J) gene segments. The heavy-chain pool contains sets of C -region exons and sets of V , diversity (D), and J gene segments.

To make an antibody [molecule](#), a V_L [gene](#) segment recombines with a J_L gene segment to produce a [DNA](#) sequence coding for the V region of a [light chain](#), and a V_H gene segment recombines with a D and a J_H gene segment to produce a DNA sequence coding for the V region of a heavy chain. Each of the assembled V -region coding sequences is then cotranscribed with the appropriate C -region sequence to produce an [RNA](#) molecule that codes for the complete [polypeptide](#) chain. Cells making functional heavy and light chains turn off the $V(D)J$ joining process to ensure that each B cell makes only one species of [antigen-binding site](#).

By randomly combining inherited [gene](#) segments that code for V_L and V_H regions, humans can make hundreds of different light chains and thousands of different heavy chains. Because the [antigen-binding site](#) is formed where the hypervariable loops of the V_L and V_H come together in the final antibody, the heavy and light chains can pair to form antibodies with millions of different antigen-binding sites. This number is enormously increased by the loss and gain of nucleotides at the site of gene-segment joining, as well as by somatic mutations that occur with very high frequency in the assembled V -region coding sequences after stimulation by antigen and helper T cells.

All B cells initially make IgM antibodies, and most then make IgD as well. Later many switch and make antibodies of other classes but with the same [antigen-binding site](#) as the original IgM and IgD antibodies. Such [class switching](#) depends on antigen stimulation and helper T cells, and it allows the same antigen-binding sites to be distributed among antibodies with varied biological properties.

T Cells and MHC Proteins

The diverse responses of T cells are collectively called *cell-mediated immune reactions*. This is to distinguish them from antibody responses, which, of course, also depend on cells (B cells). Like antibody responses, T cell 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 antibody responses, require helper T cells for their initiation. Most importantly, unlike B cells, T cells can help eliminate pathogens that reside inside host cells. Much of the rest of this chapter is concerned with how T cells accomplish this feat.

T cell responses differ from B cell responses in at least two crucial ways. First, T cells are activated by foreign [antigen](#) to proliferate and differentiate into effector cells only when the antigen is displayed on the surface of antigen-presenting cells in peripheral lymphoid organs. The T cells respond in this manner because the form of antigen they recognize is different from that recognized by B cells. Whereas B cells recognize intact antigen, T cells recognize fragments of [protein](#) antigens that have been partly degraded inside the [antigen-presenting cell](#). The peptide fragments are then carried to the surface of the presenting cell on special molecules called [MHC proteins](#), which present the fragments to T cells. The second difference is that, once activated, effector T cells act only at short range, either within a secondary [lymphoid organ](#) or after they have migrated into a site of infection. They interact directly with another cell in the body, which they either kill or signal in some way (we shall refer to such cells as *target cells*). Activated B cells, by contrast, secrete antibodies that can act far away.

There are two main classes of T cells—cytotoxic T cells and helper T cells. Effector *cytotoxic T cells* directly kill cells that are infected with a [virus](#) or some other intracellular [pathogen](#). Effector *helper T cells*, by contrast, help stimulate the responses of other cells—mainly macrophages, B cells, and cytotoxic T cells.

In this [section](#), we describe these two classes of T cells and their respective functions. We discuss how they recognize foreign antigens on the surface of [antigen](#)-presenting cells and target cells and consider the crucial part played by [MHC](#) proteins in the recognition process. Finally, we describe how T cells are selected during their [development](#) in the thymus to ensure that only cells with potentially useful receptors survive and mature. We begin by considering the nature of the cell-surface receptors that T cells use to recognize antigen.

T Cell Receptors Are Antibodylike Heterodimers

Because T cell responses depend on direct contact with an [antigen-presenting cell](#) or a target cell, the antigen receptors made by T cells, unlike antibodies made by B cells, exist only in [membrane](#)-bound form and are not secreted. For this reason, T cell receptors were difficult to isolate, and it was not until the 1980s that they were first identified biochemically. On both cytotoxic and helper T cells, the receptors are similar to antibodies. They are composed of two disulfide-linked [polypeptide](#) chains (called α and β), each of which contains two [Ig](#)-like

domains, one variable and one constant. Moreover, the three-dimensional structure of the extracellular part of a T cell [receptor](#) has been [determined](#) by x-ray diffraction, and it looks very much like one arm of a Y-shaped antibody [molecule](#).

The pools of [gene](#) segments that encode the α and β chains are located on different chromosomes. Like antibody heavy-chain pools, the T cell [receptor](#) pools contain separate V, D, and J gene segments, which are brought together by [site-specific recombination](#) during T cell [development](#) in the thymus. With one exception, all the mechanisms used by B cells to generate antibody diversity are also used by T cells to generate T cell receptor diversity. Indeed, the same V(D)J recombinase is used, including the RAG proteins discussed earlier. The mechanism that does not operate in T cell receptor diversification is [antigen](#)-driven somatic hypermutation. Thus, the affinity of the receptors remains low ($K_a \sim 10^5$ - 10^7 liters/[mole](#)), even late in an [immune response](#). We discuss later how various co-receptors and cell-cell adhesion mechanisms greatly strengthen the binding of a T cell to an [antigen-presenting cell](#) or a target cell, helping to compensate for the low affinity of the T cell receptors.

A small minority of T cells, instead of making α and β chains, make a different but related type of [receptor heterodimer](#), composed of γ and δ chains. These cells arise early in [development](#) and are found mainly in epithelia (in the skin and gut, for example). Their functions are uncertain, and we shall not discuss them further.

As with [antigen](#) receptors on B cells, the T cell receptors 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 discuss these proteins in more detail later. First, however, we need to consider how cytotoxic and helper T cells function and the special ways in which they recognize foreign antigen.

Antigen-Presenting Cells Activate T Cells

Before cytotoxic or helper T cells can kill or help their target cells, respectively, they must be activated to proliferate and differentiate into effector cells. This activation occurs in peripheral lymphoid organs on the surface of [antigen-presenting cells](#) that display foreign [antigen](#) complexed with [MHC](#) proteins on their surface.

There are three main types of [antigen](#)-presenting cells in peripheral lymphoid organs that can activate T cells—dendritic cells, macrophages, and B cells. The most potent of these are [dendritic cells](#), whose only known function is to present foreign antigens to T cells. Immature dendritic cells are located in tissues throughout the body, including the skin, gut, and respiratory tract. When they encounter invading microbes at these sites, they endocytose the pathogens or their products and carry them via the [lymph](#) to local lymph nodes or gut-associated lymphoid organs. The encounter with a [pathogen](#) induces the [dendritic cell](#) to mature from an antigen-capturing cell to an [antigen-presenting cell](#) that can activate T cells.

Antigen-presenting cells display three types of [protein](#) molecules on their surface that have a role in activating a T cell to become an [effector cell](#): (1) [MHC proteins](#), which present foreign [antigen](#) to the T cell [receptor](#), (2) [costimulatory proteins](#), which bind

to [complementary](#) receptors on the T cell surface, and (3) *cell-cell adhesion molecules*, which enable a T cell to bind to the [antigen-presenting cell](#) for long enough to become activated.

Effector Cytotoxic T Cells Induce Infected Target Cells to Kill Themselves

[Cytotoxic T cells](#) provide protection against intracellular pathogens such as viruses and some bacteria and parasites that multiply in the host-cell [cytoplasm](#), where they are sheltered from attack by antibodies. They provide this protection by killing the infected cell before the microbes can proliferate and escape from the infected cell to infect neighboring cells.

Once a [cytotoxic T cell](#) has been activated by an infected [antigen-presenting cell](#) to become an [effector cell](#), it can kill any target cell infected with the same [pathogen](#). When the effector T cell recognizes a microbial antigen on the surface of an infected target cell, it focuses its secretory apparatus on the target. We can observe this behavior by studying effector T cells bound to their targets: when labeled with anti-[tubulin](#) antibodies, the T cell [centrosome](#) is seen to be oriented toward the point of contact with the target cell. Moreover, antibody labeling shows that talin and other proteins that help link cell-surface receptors to cortical [actin](#) filaments are concentrated in the cortex of the T cell at the contact site. The aggregation of T cell receptors at the contact site apparently leads to a local alteration in the actin filaments in the [cell cortex](#). A [microtubule](#)-dependent mechanism then moves the centrosome and its associated Golgi apparatus toward the contact site, focusing the killing machinery on the target cell. A similar cytoskeletal polarization is seen when an effector [helper T cell](#) interacts functionally with a target cell.

Once bound to its target cell, a [cytotoxic T cell](#) can employ at least two strategies to kill the target, both of which operate by inducing the target cell to kill itself by undergoing [apoptosis](#). In killing an infected target cell, the cytotoxic T cell usually releases a pore-forming [protein](#) called **perforin**, which is [homologous](#) to the complement component C9 and polymerizes in the target cell [plasma membrane](#) to form transmembrane channels. Perforin is stored in secretory vesicles of the cytotoxic T cell and is released by local [exocytosis](#) at the point of contact with the target cell. The secretory vesicles also contain serine proteases, which are thought to enter the target cell [cytosol](#) through the perforin channels. One of the proteases, called *granzyme B*, cleaves, and thereby activates, one or more members of the [caspase family](#) of proteases that mediate apoptosis. These caspases then activate other caspases, producing a proteolytic cascade that helps kill the cell (discussed in Chapter 17) ([Figure 24-46A](#)). Mice in which the perforin [gene](#) is inactivated cannot generate microbe-specific cytotoxic T cells and show increased susceptibility to certain viral and intracellular bacterial infections.

In the second killing strategy, the [cytotoxic T cell](#) also activates a death-inducing [caspase](#) cascade in the target cell but does it less directly. A homotrimeric [protein](#) on the cytotoxic T cell surface called **Fas ligand** binds to transmembrane [receptor](#) proteins on the target cell called **Fas**. The binding alters the Fas proteins so that their clustered cytosolic tails recruit procaspase-8 into the [complex](#) via an [adaptor protein](#). The recruited procaspase-8 molecules cross-cleave and activate each other to begin the caspase cascade that leads to [apoptosis](#). Cytotoxic T cells apparently use this killing strategy to help contain an [immune response](#) once it is well underway, by killing excessive effector lymphocytes, especially effector T cells: if the [gene](#) encoding either Fas or Fas ligand is inactivated by [mutation](#), effector

lymphocytes accumulate in vast numbers in the spleen and [lymph](#) nodes, which become enormously enlarged.

Effector Helper T Cells Help Activate Macrophages, B Cells, and Cytotoxic T Cells

In contrast to cytotoxic T cells, [helper T cells](#) are crucial for defense against both extracellular and intracellular pathogens. They help stimulate B cells to make antibodies that help inactivate or eliminate extracellular pathogens and their toxic products. They activate macrophages to destroy any intracellular [pathogen](#) multiplying within the [macrophage's](#) phagosomes, and they help activate cytotoxic T cells to kill infected target cells.

Once a [helper T cell](#) has been activated by an [antigen-presenting cell](#) to become an [effector cell](#), it can then help activate other cells. It does this both by secreting a variety of cytokines and by displaying costimulatory proteins on its surface. When activated by an antigen-presenting cell, a naïve helper T cell can differentiate into either of two distinct types of effector helper cell, called T_H1 and T_H2 . T_H1 cells mainly help activate macrophages and cytotoxic T cells, whereas T_H2 cells mainly help activate B cells. As we discuss later, the nature of the invading [pathogen](#) and the types of innate **immune** responses it elicits largely determine which type of helper T cell develops. This, in turn, determines the nature of the adaptive **immune** responses mobilized to fight the invaders.

T Cells Recognize Foreign Peptides Bound to MHC Proteins

As discussed earlier, both cytotoxic T cells and helper T cells are initially activated in peripheral lymphoid organs by recognizing foreign [antigen](#) on the surface of an [antigen-presenting cell](#), usually a [dendritic cell](#). The antigen is in the form of peptide fragments that are generated by the degradation of foreign [protein](#) antigens inside the antigen-presenting cell. The recognition process depends on the presence in the antigen-presenting cell of **MHC proteins**, which bind these fragments, carry them to the cell surface, and present them there, along with a co-stimulatory signal, to the T cells. Once activated, effector T cells then recognize the same peptide-MHC [complex](#) on the surface of the target cell they influence, which may be a B cell, a [cytotoxic T cell](#), or an infected [macrophage](#) in the case of a [helper T cell](#), or a [virus](#)-infected cell in the case of a cytotoxic T cell.

MHC proteins are encoded by a large [complex](#) of genes called the **major histocompatibility complex (MHC)**. There are two main structurally and functionally distinct classes of MHC proteins: *class I MHC proteins*, which present foreign peptides to cytotoxic T cells, and *class II MHC proteins*, which present foreign peptides to helper cells.

Before examining the different mechanisms by which [protein](#) antigens are processed for display to the two main classes of T cells, we must look more closely at the **MHC** proteins themselves, which have such an important role in T cell function.

MHC Proteins Were Identified in Transplantation Reactions Before Their Functions Were Known

MHC proteins were initially identified as the main antigens recognized in **transplantation reactions**. When organ grafts are exchanged between adult individuals, either of the same species (*allografts*) or of different species (*xenografts*), they are usually rejected. In the 1950s, skin grafting experiments between different strains of mice demonstrated that *graft rejection* is an [adaptive immune response](#) to the foreign antigens on the surface of the grafted cells. Rejection is mediated mainly by T cells, which react against genetically “foreign” versions of cell-surface proteins called *histocompatibility molecules* (from the Greek word *histos*, meaning “tissue”). The MHC proteins encoded by the clustered genes of the major histocompatibility [complex](#) (MHC) are by far the most important of these. MHC proteins are expressed on the cells of all higher vertebrates. They were first demonstrated in mice, where they are called *H-2 antigens* (histocompatibility-2 antigens). In humans they are called *HLA antigens* (human-leucocyte-associated antigens) because they were first demonstrated on leucocytes (white blood cells).

Three remarkable properties of **MHC** proteins baffled immunologists for a long time. First, MHC proteins are overwhelmingly the preferred antigens recognized in T-cell-mediated transplantation reactions. Second, an unusually large fraction of T cells are able to recognize foreign MHC proteins: whereas fewer than 0.001% of an individual's T cells respond to a typical viral [antigen](#), more than 0.1% of them respond to a single foreign MHC antigen. Third, some of the genes that code for MHC proteins are the most [polymorphic](#) known in higher vertebrates. That is, within a species, there is an extraordinarily large number of *alleles* (alternative forms of the same [gene](#)) present (in some cases more than 200), without any one [allele](#) predominating. As each individual has at least 12 genes encoding MHC proteins (see later), it is very rare for two unrelated individuals to have an identical set of MHC proteins. This makes it very difficult to match donor and recipient for organ transplantation unless they are closely related.

Of course, a vertebrate does not need to protect itself against invasion by foreign vertebrate cells. So the apparent obsession of its T cells with foreign **MHC** proteins and the extreme polymorphism of these molecules were a great puzzle. The puzzle was solved only after it was discovered that (1) MHC proteins bind fragments of foreign proteins and display them on the surface of host cells for T cells to recognize, and (2) T cells respond to foreign MHC proteins in the same way they respond to self MHC proteins that have foreign [antigen](#) bound to them.

Class I and Class II MHC Proteins Are Structurally Similar Heterodimers

Class I and class II **MHC** proteins have very similar overall structures. They are both transmembrane heterodimers with extracellular N-terminal domains that bind [antigen](#) for presentation to T cells.

Class I MHC proteins consist of a transmembrane α chain, which is encoded by a class I MHC [gene](#), and a small extracellular [protein](#) called β_2 -microglobulin. The β_2 -microglobulin does not span the [membrane](#) and is encoded by a gene that does not lie in the MHC gene cluster. The α chain is folded into three extracellular globular domains (α_1 , α_2 , α_3), and the α_3 [domain](#) and the β_2 -microglobulin, which are closest to the membrane, are both similar to an [Ig](#) domain. The two N-terminal domains of the α chain, which are farthest from the

membrane, contain the [polymorphic](#) (variable) amino acids that are recognized by T cells in transplantation reactions. These domains bind a peptide and present it to cytotoxic T cells.

Like class I [MHC](#) proteins, **class II MHC proteins** are heterodimers with two conserved [Ig](#)-like domains close to the [membrane](#) and two [polymorphic](#) (variable) N-terminal domains farthest from the membrane. In these proteins, however, both chains (α and β) are encoded by genes within the MHC, and both span the membrane. The two polymorphic domains bind a peptide and present it to helper T cells.

The presence of [Ig](#)-like domains in class I and class II proteins suggests that [MHC](#) proteins and antibodies have a common evolutionary history. The locations of the genes that encode class I and class II MHC proteins in humans are shown in, where we illustrate how an individual can make six types of class I MHC proteins and more than six types of class II proteins.

In addition to the classic class I [MHC](#) proteins, there are many *nonclassical class I MHC proteins*, which form dimers with β 2-microglobulin. These proteins are not [polymorphic](#), but some of them present specific microbial antigens, including some lipids and glycolipids, to T cells. The functions of most of them, however, are unknown.

An MHC Protein Binds a Peptide and Interacts with a T Cell Receptor

Any individual can make only a small number of different [MHC](#) proteins, which together must be able to present peptide fragments from almost any foreign [protein](#) to T cells. Thus, unlike an antibody [molecule](#), each MHC protein has to be able to bind a very large number of different peptides. The structural basis for this versatility has emerged from x-ray crystallographic analyses of MHC proteins.

a class I [MHC protein](#) has a single peptide-[binding site](#) located at one end of the [molecule](#), facing away from the [plasma membrane](#). This site consists of a deep groove between two long α helices; the groove narrows at both ends so that it is only large enough to accommodate an extended peptide about 8–10 amino acids long. In fact, when a class I MHC protein was first analyzed by x-ray crystallography in 1987, this groove contained bound peptides that had co-crystallized with the MHC protein, suggesting that once a peptide binds to this site it does not normally dissociate.

A typical peptide binds in the groove of a class I [MHC protein](#) in an extended [conformation](#), with its terminal [amino group](#) bound to an invariant pocket at one end of the groove and its terminal [carboxyl group](#) bound to an invariant pocket at the other end of the groove. Other amino acids (called “anchor amino acids”) in the peptide bind to “specificity pockets” in the groove formed by [polymorphic](#) portions of the MHC protein. The side chains of other amino acids of the peptide point outward, in a position to be recognized by receptors on cytotoxic T cells. Because the conserved pockets at the ends of the binding groove recognize features of the peptide backbone that are common to all peptides, each allelic form of a class I MHC protein can bind a large variety of peptides of diverse sequence. At the same time, the differing specificity pockets along the groove, which bind particular [amino acid](#) side chains of the peptide, ensure that each allelic form binds and presents a distinct characteristic set of peptides. Thus, the six types of class I MHC proteins in an individual can present a broad

range of foreign peptides to the cytotoxic T cells, but in each individual they do so in slightly different ways.

Class II [MHC](#) proteins have a three-dimensional structure that is very similar to that of class I proteins, but their [antigen](#)-binding groove does not narrow at the ends, so it can accommodate longer peptides, which are usually 13–17 amino acids long. Moreover, the peptide is not bound at its ends. It is held in the groove by parts of its peptide backbone that bind to invariant pockets formed by conserved amino acids that line all class II MHC peptide-binding grooves, as well as by the side chains of anchor amino acids that bind to variable specificity pockets in the groove. A class II MHC binding groove can accommodate a more heterogeneous set of peptides than can a class I MHC groove. Thus, although an individual makes only a small number of types of class II proteins, each with its own unique peptide-binding groove, together these proteins can bind and present an enormous variety of foreign peptides to helper T cells, which have a crucial role in almost all adaptive **immune** responses.

The way in which the T cell [receptor](#) recognizes a peptide fragment bound to an [MHC protein](#) is revealed by x-ray crystallographic analyses of complexes formed between a soluble receptor and a soluble MHC protein with peptide in its binding groove. (The soluble proteins for these experiments are produced by [recombinant DNA](#) technology.) In each case studied, the T cell receptor fits diagonally across the peptide-binding groove and binds through its V_α and V_β hypervariable loops to both the walls of the groove and the peptide. Soluble MHC-peptide complexes are now widely used to detect T cells with a particular specificity; they are usually cross-linked into tetramers to increase their [avidity](#) for T cell receptors.

MHC Proteins Help Direct T Cells to Their Appropriate Targets

Class I [MHC](#) proteins are expressed on virtually all nucleated cells. This is presumably because effector cytotoxic T cells must be able to focus on and kill any cell in the body that happens to become infected with an intracellular microbe such as a [virus](#). Class II proteins, by contrast, are normally confined largely to cells that take up foreign antigens from the extracellular fluid and interact with helper T cells. These include dendritic cells, which initially activate helper T cells, as well as the targets of effector helper T cells, such as macrophages and B cells. Because dendritic cells express both class I and class II MHC proteins, they can activate both cytotoxic and helper T cells.

It is important that effector cytotoxic T cells focus their attack on cells that *make* the foreign antigens (such as viral proteins), while helper T cells focus their help mainly on cells that have taken up foreign antigens from the extracellular fluid. Since the former type of target cell is always a menace, while the latter type is essential for the body's **immune** defenses, it is vitally important that T cells never confuse the two target cells and misdirect their cytotoxic and helper functions. Therefore, in addition to the [antigen receptor](#) that recognizes a peptide-[MHC complex](#), each of the two major classes of T cells also expresses a *co-receptor* that recognizes a separate, invariant part of the appropriate class of MHC [protein](#). These two co-receptors, called [CD4](#) and [CD8](#), help direct helper T cells and cytotoxic T cells, respectively, to their appropriate targets, as we now discuss. The properties of class I and class II MHC proteins are compared in.

CD4 and CD8 Co-receptors Bind to Nonvariable Parts of MHC Proteins

The affinity of T cell receptors for peptide-MHC complexes on an [antigen-presenting cell](#) or target cell is usually too low to mediate a functional interaction between the two cells by itself. T cells normally require *accessory receptors* to help stabilize the interaction by increasing the overall strength of the cell-cell adhesion. Unlike T cell receptors or MHC proteins, the accessory receptors do not bind foreign antigens and are invariant.

When accessory receptors also have a direct role in activating the T cell by generating their own intracellular signals, they are called **co-receptors**. The most important and best understood of the co-receptors on T cells are the [CD4](#) and [CD8](#) proteins, both of which are single-pass transmembrane proteins with extracellular [Ig](#)-like domains. Like T cell receptors, they recognize [MHC](#) proteins, but, unlike T cell receptors, they bind to nonvariable parts of the [protein](#), far away from the peptide-binding groove. [CD4](#) is expressed on helper T cells and binds to class II MHC proteins, whereas [CD8](#) is expressed on cytotoxic T cells and binds to class I MHC proteins. Thus, CD4 and CD8 contribute to T cell recognition by helping to focus the cell on particular MHC proteins, and thus on particular types of cells—helper T cells on dendritic cells, macrophages, and B cells, and cytotoxic cells on any nucleated host cell displaying a foreign peptide on a class I MHC protein. The cytoplasmic tail of these transmembrane proteins is associated with a member of the [Src family](#) of cytoplasmic tyrosine protein kinases called *Lck*, which phosphorylates various intracellular proteins on tyrosines and thereby participates in the activation of the T cell. Antibodies to CD4 and CD8 are widely used as tools to distinguish between the two main classes of T cells, in both humans and experimental animals.

Ironically, the AIDS [virus](#) ([HIV](#)) makes use of [CD4](#) molecules (as well as [chemokine](#) receptors) to enter helper T cells. It is the eventual depletion of helper T cells that renders AIDS patients susceptible to infection by microbes that are not normally dangerous. As a result, most AIDS patients die of infection within several years of the onset of symptoms, unless they are treated with a combination of powerful anti-HIV drugs. HIV also uses CD4 and chemokine receptors to enter macrophages, which also have both of these receptors on their surface.

Before a cytotoxic or [helper T cell](#) can recognize a foreign [protein](#), the protein has to be processed inside an [antigen-presenting cell](#) or target cell so that it can be displayed as peptide-MHC complexes on the cell surface. We first consider how a [virus](#)-infected antigen-presenting cell or target cell processes viral proteins for presentation to a [cytotoxic T cell](#). We then discuss how ingested foreign proteins are processed for presentation to a helper T cell.

Cytotoxic T Cells Recognize Fragments of Foreign Cytosolic Proteins in Association with Class I MHC Proteins

One of the first, and most dramatic, demonstrations that [MHC](#) proteins present foreign antigens to T cells came from an experiment performed in the 1970s. It was found that effector cytotoxic T cells from a [virus](#)-infected mouse could kill cultured cells infected with

the same virus only if these target cells expressed some of the same class I MHC proteins as the infected mouse. This experiment demonstrated that the T cells of any individual that recognize a specific [antigen](#) do so only when that antigen is associated with the allelic forms of MHC proteins expressed by that individual, a phenomenon known as *MHC restriction*.

The chemical nature of the viral antigens recognized by cytotoxic T cells was not discovered for another 10 years. In experiments on cells infected with influenza [virus](#), it was unexpectedly found that some of the effector cytotoxic T cells activated by the virus specifically recognize internal proteins of the virus that would not be accessible in the intact virus particle. Subsequent evidence indicated that the T cells were recognizing degraded fragments of the internal viral proteins that were bound to class I [MHC](#) proteins on the infected cell surface. Because a T cell can recognize tiny amounts of [antigen](#) (as few as one hundred peptide-MHC complexes), only a small fraction of the fragments generated from viral proteins have to bind to class I MHC proteins and get to the cell surface to attract an attack by an effector [cytotoxic T cell](#).

The viral proteins are synthesized in the [cytosol](#) of the infected cell. As discussed in Chapter 3, proteolytic degradation in the cytosol is mainly mediated by an ATP- and [ubiquitin](#)-dependent mechanism that operates in *proteasomes*—large [proteolytic enzyme](#) complexes constructed from many different [protein](#) subunits. Although all proteasomes are probably able to generate peptide fragments that can bind to class I [MHC](#) proteins, some proteasomes are thought to be specialized for this purpose, as they contain two subunits that are encoded by genes located within the MHC chromosomal region. Even bacterial proteasomes cut proteins into peptides of about the length that fits into the groove of a class I MHC protein, suggesting that the MHC groove evolved to fit this length of peptide.

How do peptides generated in the [cytosol](#) make contact with the peptide-binding groove of class I [MHC](#) proteins in the [lumen](#) of the [endoplasmic reticulum](#)? The answer was discovered through observations on [mutant](#) cells in which class I MHC proteins are not expressed at the cell surface but are instead degraded within the cell. The mutant genes in these cells proved to encode subunits of a [protein](#) belonging to the family of *ABC transporters*, which we discuss in Chapter 11. This transporter protein is located in the [ER membrane](#) and uses the energy of ATP hydrolysis to [pump](#) peptides from the cytosol into the [ER lumen](#). The genes encoding its two subunits are in the MHC chromosomal region, and, if either [gene](#) is inactivated by [mutation](#), cells are unable to supply peptides to class I MHC proteins. The class I MHC proteins in such mutant cells are degraded in the cell because peptide binding is normally required for the proper folding of these proteins. Until it binds a peptide, a class I MHC protein remains in the ER, tethered to an ABC transporter by a chaperone protein.

In cells that are not infected, peptide fragments come from the cells' own cytosolic and nuclear proteins that are degraded in the processes of normal [protein](#) turnover and quality control mechanisms. (Surprisingly, more than 30% of the proteins made by mammalian cells are apparently faulty and are degraded in proteasomes soon after they are synthesized.) These peptides are pumped into the [ER](#) and are carried to the cell surface by class I [MHC](#) proteins. They are not antigenic because the cytotoxic T cells that could recognize them have been eliminated or inactivated during T cell [development](#), as we discuss later.

When cytotoxic T cells and some helper T cells are activated by [antigen](#) to become effector cells, they secrete the [cytokine](#) **interferon- γ (IFN- γ)**, which greatly enhances anti-viral responses. The IFN- γ acts on infected cells in two ways. It blocks viral replication, and it

increases the [expression](#) of many genes within the [MHC](#) chromosomal region. These genes include those that encode class I (and class II) MHC proteins, the two specialized [proteasome](#) subunits, and the two subunits of the peptide transporter located in the [ER](#). Thus, all of the machinery required for presenting viral antigens to cytotoxic T cells is coordinately called into action by IFN- γ , creating a positive feedback that amplifies the [immune response](#) and culminates in the death of the infected cells.

Helper T Cells Recognize Fragments of Endocytosed Foreign Protein Associated with Class II MHC Proteins

Unlike cytotoxic T cells, helper T cells do not act directly to kill infected cells so as to eliminate microbes. Instead, they stimulate macrophages to be more effective in destroying intracellular microorganisms, and they help B cells and cytotoxic T cells to respond to microbial antigens.

Like the viral proteins presented to cytotoxic T cells, the proteins presented to helper T cells on [antigen](#)-presenting cells or target cells are degraded fragments of foreign proteins. The fragments are bound to class II [MHC](#) proteins in much the same way that [virus](#)-derived peptides are bound to class I MHC proteins. But both the source of the peptide fragments presented and the route they take to find the MHC proteins are different from those of peptide fragments presented by class I MHC proteins to cytotoxic T cells.

Rather than being derived from foreign [protein](#) synthesized in the [cytosol](#) of a cell, the foreign peptides presented to helper T cells are derived from endosomes. Some come from extracellular microbes or their products that the [antigen-presenting cell](#) has endocytosed and degraded in the acidic environment of its endosomes. Others come from microbes growing within the endocytic [compartment](#) of the antigen-presenting cell. These peptides do not have to be pumped across a [membrane](#) because they do not originate in the cytosol; they are generated in a compartment that is topologically equivalent to the extracellular space. They never enter the [lumen](#) of the [ER](#), where the class II [MHC](#) proteins are synthesized and assembled, but instead bind to preassembled class II heterodimers in a special endosomal compartment. Once the peptide has bound, the class II MHC protein alters its [conformation](#), trapping the peptide in the binding groove for presentation at the cell surface to helper T cells.

A newly synthesized class II [MHC protein](#) must avoid clogging its binding groove prematurely in the [ER lumen](#) with peptides derived from endogenously synthesized proteins. A special [polypeptide](#), called the **invariant chain**, ensures this by associating with newly synthesized class II MHC heterodimers in the ER. Part of its polypeptide chain lies within the peptide-binding groove of the MHC protein, thereby blocking the groove from binding other peptides in the lumen of the ER. The invariant chain also directs class II MHC proteins from the *trans* Golgi network to a late endosomal [compartment](#). Here, the invariant chain is cleaved by proteases, leaving only a short fragment bound in the peptide-binding groove of the MHC protein. This fragment is then released (catalyzed by a class II-MHC-like protein called HLA-DM), freeing the MHC protein to bind peptides derived from endocytosed proteins ([Figure 24-60](#)). In this way, the functional differences between class I and class II MHC proteins are ensured—the former presenting molecules that come from the [cytosol](#), the latter presenting molecules that come from the endocytic compartment.

Most of the class I and class II [MHC](#) proteins on the surface of a target cell have peptides derived from self proteins in their binding groove. For class I proteins, the fragments derive from degraded cytosolic and nuclear proteins. For class II proteins, they mainly derive from degraded proteins that originate in the [plasma membrane](#) or extracellular fluid and are endocytosed. Only a small fraction of the 10^5 or so class II MHC proteins on the surface of an [antigen-presenting cell](#) have foreign peptides bound to them. This is sufficient, however, because only a hundred or so of such molecules are required to stimulate a [helper T cell](#), just as in the case of peptide-class-I-MHC complexes stimulating a [cytotoxic T cell](#).

Potentially Useful T Cells Are Positively Selected in the Thymus

We have seen that T cells recognize [antigen](#) in association with self [MHC](#) proteins but not in association with foreign MHC proteins: that is, T cells show *MHC restriction*. This restriction results from a process of **positive selection** during T cell [development](#) in the thymus. In this process, those immature T cells that will be capable of recognizing foreign peptides presented by self MHC proteins are selected to survive, while the remainder, which would be of no use to the animal, undergo [apoptosis](#). Thus, MHC restriction is an acquired property of the [immune system](#) that emerges as T cells develop in the thymus.

The most direct way to study the selection process is to follow the fate of a set of developing T cells of known specificity. This can be done by using transgenic mice that express a specific pair of rearranged α and β T cell [receptor](#) genes derived from a T cell [clone](#) of known [antigen](#) and [MHC](#) specificity. Such experiments show that the transgenic T cells mature in the thymus and populate the peripheral lymphoid organs only if the transgenic mouse also expresses the same allelic form of MHC [protein](#) as is recognized by the transgenic T cell receptor. If the mouse does not express the appropriate MHC protein, the transgenic T cells die in the thymus. Thus, the survival and maturation of a T cell depend on a match between its receptor and the MHC proteins expressed in the thymus. Similar experiments using transgenic mice in which MHC [expression](#) is confined to specific cell types in the thymus indicate that it is MHC proteins on epithelial cells in the cortex of the thymus that are responsible for this positive selection process. After positively selected T cells leave the thymus, their continued survival depends on their continual stimulation by self-peptide-MHC complexes; this stimulation is enough to promote cell survival but not enough to activate the T cells to become effector cells.

As part of the positive selection process in the thymus, developing T cells that express receptors recognizing class I [MHC](#) proteins are selected to become cytotoxic cells, while T cells that express receptors recognizing class II MHC proteins are selected to become helper cells. Thus, genetically engineered mice that lack cell-surface class I MHC proteins specifically lack cytotoxic T cells, whereas mice that lack class II MHC proteins specifically lack helper T cells. The cells that are undergoing positive selection initially express both [CD4](#) and [CD8](#) co-receptors, and these are required for the selection process: without CD4, helper T cells fail to develop, and without CD8, cytotoxic T cells fail to develop.

Positive selection still leaves a large problem to be solved. If developing T cells with receptors that recognize self peptides associated with self [MHC](#) proteins were to mature in the thymus and migrate to peripheral lymphoid tissues, they might wreak havoc. A second, *negative selection* process in the thymus is required to help avoid this potential disaster.

Many Developing T Cells That Could Be Activated by Self Peptides Are Eliminated in the Thymus

As discussed previously, a fundamental feature of the adaptive [immune system](#) is that it can distinguish self from nonself and normally does not react against self molecules. An important mechanism in achieving this state of *immunological self tolerance* is the [deletion](#) in the thymus of developing self-reactive T cells—that is, T cells whose receptors bind strongly enough to the [complex](#) of a self peptide and a self [MHC protein](#) to become activated. Because, as we discuss later, most B cells require helper T cells to respond to [antigen](#), the elimination of self-reactive helper T cells also helps ensure that self-reactive B cells that escape B cell tolerance [induction](#) are harmless.

It is not enough, therefore, for the thymus to select *for* T cells that recognize self [MHC](#) proteins; it must also select *against* T cells that could be activated by self MHC proteins complexed with self peptides. In other words, it must pick out for survival just those T cells that will be capable of responding to self MHC proteins complexed with foreign peptides, even though these peptides are not present in the developing thymus. It is thought that these T cells bind weakly in the thymus to self MHC proteins that are carrying self peptides mismatched to the T cell receptors. Thus, the required goal can be achieved by (1) ensuring the death of T cells that bind *strongly* to the self-peptide-MHC complexes in the thymus while (2) promoting the survival of those that bind weakly and (3) permitting the death of those that do not bind at all. Process 2 is the positive selection we have just discussed. Process 1 is called **negative selection**. In both death processes, the cells that die undergo [apoptosis](#).

The most convincing evidence for negative selection derives once again from experiments with transgenic mice. After the introduction of T cell [receptor](#) transgenes encoding a receptor that recognizes a male-specific peptide [antigen](#), for example, large numbers of mature T cells expressing the transgenic receptor are found in the thymus and peripheral lymphoid organs of female mice. Very few, however, are found in male mice, where the cells die in the thymus before they have a chance to mature. Like positive selection, negative selection requires the interaction of a T cell receptor and a [CD4](#) or [CD8](#) co-receptor with an appropriate [MHC protein](#). Unlike positive selection, however, which occurs mainly on the surface of thymus epithelial cells, negative selection occurs on the surface of thymus dendritic cells and macrophages, which, as we have seen, function as antigen-presenting cells in peripheral lymphoid organs.

The [deletion](#) of self-reactive T cells in the thymus cannot eliminate all potentially self-reactive T cells, as some self molecules are not present in the thymus. Thus, some potentially self-reactive T cells are deleted or functionally inactivated after they leave the thymus, presumably because they recognize self peptides bound to [MHC](#) proteins on the surface of dendritic cells that have not been activated by microbes and therefore do not provide a costimulatory signal. As we discuss later, [antigen](#) recognition without costimulatory signals can delete or inactivate a T or B cell.

Some potentially self-reactive T cells, however, are not deleted or inactivated. Instead, special *regulatory* (or *suppressor*) *T cells* are thought to keep them from responding to their self antigens by secreting inhibitory cytokines such as TGF- β (discussed in Chapter 15).

These self-reactive T cells may sometimes escape from this suppression and cause autoimmune diseases.

The Function of MHC Proteins Explains Their Polymorphism

The role of [MHC](#) proteins in binding foreign peptides and presenting them to T cells provides an explanation for the extensive polymorphism of these proteins. In the evolutionary war between pathogenic microbes and the adaptive [immune system](#), microbes tend to change their antigens to avoid associating with MHC proteins. When a microbe succeeds, it is able to sweep through a population as an epidemic. In such circumstances, the few individuals that produce a new MHC [protein](#) that can associate with an [antigen](#) of the altered microbe have a large selective advantage. In addition, individuals with two different alleles at any given MHC [locus](#) (heterozygotes) have a better chance of resisting infection than those with identical alleles at the locus, as they have a greater capacity to present peptides from a wide range of microbes and parasites. Thus, selection will tend to promote and maintain a large diversity of MHC proteins in the population. Strong support for this hypothesis, that infectious diseases have provided the driving force for MHC polymorphism, has come from studies in West Africa. Here, it is found that individuals with a specific MHC [allele](#) have a reduced susceptibility to a severe form of [malaria](#). Although the allele is rare elsewhere, it is found in 25% of the West African population where this form of malaria is common.

If greater [MHC](#) diversity means greater resistance to infection, why do we each have so few MHC genes encoding these molecules? Why have we not evolved strategies for increasing the diversity of MHC proteins—by [alternative RNA splicing](#), for example, or by the [genetic recombination](#) mechanisms used to diversify antibodies and T cell receptors? Presumably, the limits exist because each time a new MHC [protein](#) is added to the repertoire, the T cells that recognize self peptides in association with the new MHC protein must be eliminated to maintain self tolerance. The elimination of these T cells would counteract the advantage of adding the new MHC protein. Thus, the number of MHC proteins we express may represent a balance between the advantages of presenting a wide diversity of foreign peptides to T cells against the disadvantages of severely restricting the T cell repertoire during negative selection in the thymus. This explanation is supported by computer modeling studies.

Summary

There are two main functionally distinct classes of T cells: cytotoxic T cells kill infected cells directly by inducing them to undergo [apoptosis](#), while helper T cells help activate B cells to make antibody responses and macrophages to destroy microorganisms that either invaded the [macrophage](#) or were ingested by it. Helper T cells also help activate cytotoxic T cells. Both classes of T cells express cell-surface, antibodylike receptors, which are encoded by genes that are assembled from multiple [gene](#) segments during T cell [development](#) in the thymus. These receptors recognize fragments of foreign proteins that are displayed on the surface of host cells in association with [MHC](#) proteins. Both cytotoxic and helper T cells are activated in peripheral lymphoid organs by [antigen](#)-presenting cells, which express peptide-MHC complexes, costimulatory proteins, and various cell-cell adhesion molecules on their cell surface.

Class I and class II [MHC](#) proteins have crucial roles in presenting foreign [protein](#) antigens to cytotoxic and helper T cells, respectively. Whereas class I proteins are expressed on almost all vertebrate cells, class II proteins are normally restricted to those cell types that interact with helper T cells, such as dendritic cells, macrophages, and B lymphocytes. Both classes of MHC proteins have a single peptide-binding groove, which binds small peptide fragments derived from proteins. Each MHC protein can bind a large and characteristic set of peptides, which are produced intracellularly by protein degradation: class I MHC proteins generally bind fragments produced in the [cytosol](#), while class II MHC proteins bind fragments produced in the endocytic [compartment](#). After they have formed inside the target cell, the peptide-MHC complexes are transported to the cell surface. Complexes that contain a peptide derived from a foreign protein are recognized by T cell receptors, 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 nonpolymorphic regions of MHC proteins on the target cell: helper cells express CD4, which recognizes class II MHC proteins, while cytotoxic T cells express CD8, which recognizes class I MHC proteins.

The T cell [receptor](#) repertoire is shaped mainly by a combination of positive and negative selection processes that operate during T cell [development](#) in the thymus. These processes help to ensure that only T cells with potentially useful receptors survive and mature, while the others die by [apoptosis](#). T cells that will be able to respond to foreign peptides complexed with self [MHC](#) proteins are positively selected, while many T cells that could react strongly with self peptides complexed with self MHC proteins are eliminated. T cells with receptors that could react strongly with self antigens not present in the thymus are eliminated, functionally inactivated, or actively kept suppressed after they leave the thymus.

Helper T Cells and Lymphocyte Activation

Helper T cells are arguably the most important cells in adaptive immunity, as they are required for almost all adaptive **immune** responses. They not only help activate B cells to secrete antibodies and macrophages to destroy ingested microbes, but they also help activate cytotoxic T cells to kill infected target cells. As dramatically demonstrated in AIDS patients, without helper T cells we cannot defend ourselves even against many microbes that are normally harmless.

Helper T cells themselves, however, can only function when activated to become effector cells. They are activated on the surface of [antigen](#)-presenting cells, which mature during the innate **immune** responses triggered by an infection. The innate responses also dictate what kind of [effector cell](#) a [helper T cell](#) will develop into and thereby determine the nature of the [adaptive immune response](#) elicited.

In this final [section](#), we discuss the multiple signals that help activate a T cell and how a [helper T cell](#), once activated to become an [effector cell](#), helps activate other cells. We also consider how innate **immune** responses determine the nature of adaptive responses by stimulating helper T cells to differentiate into either T_H1 or T_H2 effector cells.

Costimulatory Proteins on Antigen-Presenting Cells Help Activate T Cells

To activate a cytotoxic or [helper T cell](#) to proliferate and differentiate into an [effector cell](#), an [antigen-presenting cell](#) provides two kinds of signals. *Signal 1* is provided by a foreign peptide bound to an [MHC protein](#) on the surface of the presenting cell. This peptide-MHC [complex](#) signals through the T cell [receptor](#) and its associated proteins. *Signal 2* is provided by costimulatory proteins, especially the **B7 proteins** (CD80 and CD86), which are recognized by the co-receptor protein [CD28](#) on the surface of the T cell. The [expression](#) of B7 proteins on an antigen-presenting cell is induced by pathogens during the innate response to an infection. Effector T cells act back to promote the expression of B7 proteins on antigen-presenting cells, creating a positive feedback loop that amplifies the T cell response.

Signal 2 is thought to amplify the intracellular signaling process triggered by signal 1. If a T cell receives signal 1 without signal 2, it may undergo [apoptosis](#) or become altered so that it can no longer be activated, even if it later receives both signals. This is one mechanism by which a T cell can become *tolerant* to self antigens.

The T cell [receptor](#) does not act on its own to transmit signal 1 into the cell. It is associated with a [complex](#) of invariant transmembrane proteins called **CD3**, which transduces the binding of the peptide-MHC complex into intracellular signals. In addition, the [CD4](#) and [CD8](#) co-receptors play important parts in the signaling process, as illustrated in.

The combined actions of signal 1 and signal 2 stimulate the T cell to proliferate and begin to differentiate into an [effector cell](#) by a curiously indirect mechanism. In culture, they cause the T cells to stimulate their own proliferation and [differentiation](#) by inducing the cells to secrete a [cytokine](#) called **interleukin-2 (IL-2)** and simultaneously to synthesize high affinity cell-surface receptors that bind it. The binding of IL-2 to the IL-2 receptors activates intracellular signaling pathways that turn on genes that help the T cells to proliferate and differentiate into effector cells ([Figure 24-65](#)). As discussed in Chapter 15, there are advantages to such an autocrine mechanism. It helps ensure that T cells differentiate into effector cells only when substantial numbers of them respond to [antigen](#) simultaneously in the same location, such as in a [lymph](#) node during an infection. Only then do IL-2 levels rise high enough to be effective.

Once bound to the surface of an [antigen-presenting cell](#), a T cell increases the strength of the binding by activating an [integrin](#) adhesion [protein](#) called *lymphocyte-function-associated protein 1 (LFA-1)*. Activated LFA-1 now binds more strongly to its [Ig-like ligand](#), *intracellular adhesion molecule 1 (ICAM-1)*, on the surface of the presenting cell. This increased adhesion enables the T cell to remain bound to the antigen-presenting cell long enough for the T cell to become activated.

The activation of a T cell is controlled by negative feedback. During the activation process, the cell starts to express another cell-surface [protein](#) called *CTLA-4*, which acts to inhibit intracellular signaling. It resembles [CD28](#), but it binds to B7 proteins on the surface of the [antigen-presenting cell](#) with much higher affinity than does CD28, and, when it does, it holds the activation process in check. Mice with a disrupted *CTLA-4* [gene](#) die from a massive accumulation of activated T cells.

Most of the T (and B) effector cells produced during an [immune response](#) must be eliminated after they have done their job. As [antigen](#) levels fall and the response subsides, effector cells are deprived of the antigen and [cytokine](#) stimulation that they need to survive, and the majority die by [apoptosis](#). Only memory cells and some long-lived effector cells survive.

Before considering how effector helper T cells help activate macrophages and B cells, we need to discuss the two functionally distinct subclasses of effector helper T cells, T_H1 and T_H2 cells, and how they are generated.

The Subclass of Effector Helper T Cell Determines the Nature of the Adaptive Immune Response

When a [antigen-presenting cell](#) activates a naïve [helper T cell](#) in a peripheral lymphoid tissue, the T cell can differentiate into either a T_H1 or T_H2 effector helper cell. These two types of functionally distinct subclasses of effector helper T cells can be distinguished by the cytokines they secrete. If the cell differentiates into a **T_H1 cell**, it will secrete *interferon- γ* (*IFN- γ*) and *tumor necrosis factor- α* (*TNF- α*) and will activate macrophages to kill microbes located within the macrophages' phagosomes. It will also activate cytotoxic T cells to kill infected cells. Although, in these ways, T_H1 cells mainly defend an animal against intracellular pathogens, they may also stimulate B cells to secrete specific subclasses of IgG antibodies that can coat extracellular microbes and activate complement.

If the naïve T helper cell differentiates into a T_H2 cell, by contrast, it will secrete *interleukins* 4, 5, 10, and 13 (*IL-4*, *IL-5*, *IL-10*, and *IL-13*) and will mainly defend the animal against extracellular pathogens. A T_H2 cell can stimulate B cells to make most classes of antibodies, including IgE and some subclasses of IgG antibodies that bind to mast cells, basophils, and eosinophils. These cells release local mediators that cause sneezing, coughing, or diarrhea and help expel extracellular microbes and larger parasites from epithelial surfaces of the body.

Thus, the decision of naïve helper T cells to differentiate into T_H1 or T_H2 effector cells influences the type of [adaptive immune response](#) that will be mounted against the [pathogen](#)—whether it will be dominated by [macrophage](#) activation or by antibody production. The specific cytokines present during the process of [helper T cell](#) activation influence the type of [effector cell](#) produced. Microbes at a site of infection not only stimulate dendritic cells to make cell-surface B7 costimulatory proteins; they also stimulate them to produce cytokines. The dendritic cells then migrate to a peripheral [lymphoid organ](#) and activate naïve helper T cells to differentiate into either T_H1 or T_H2 effector cells, depending on the cytokines the dendritic cells produce. Some intracellular bacteria, for example, stimulate dendritic cells to produce *IL-12*, which encourages T_H1 [development](#), and thereby macrophage activation. As expected, mice that are deficient in either *IL-12* or its [receptor](#) are much more susceptible to these bacterial infections than are normal mice. Many parasitic [protozoa](#) and worms, by contrast, stimulate the production of cytokines that encourage T_H2 development, and thereby antibody production and eosinophil activation, leading to parasite expulsion.

Once a T_H1 or T_H2 [effector cell](#) develops, it inhibits the [differentiation](#) of the other type of [helper T cell](#). *IFN- γ* produced by T_H1 cells inhibits the [development](#) of T_H2 cells, while

IL-4 and IL-10 produced by T_H2 cells inhibit the development of T_H1 cells. Thus, the initial choice of response is reinforced as the response proceeds.

The importance of the T_H1/T_H2 decision is illustrated by individuals infected with *Mycobacterium leprae*, the bacterium that causes leprosy. The bacterium replicates mainly within macrophages and causes either of two forms of disease, depending mainly on the genetic make-up of the infected individual. In some patients, the *tuberculoid* form of the disease occurs. T_H1 cells develop and stimulate the infected macrophages to kill the bacteria. This produces a local [inflammatory response](#), which damages skin and nerves. The result is a chronic disease that progresses slowly but does not kill the host. In other patients, by contrast, the *lepromatous* form of the disease occurs. T_H2 cells develop and stimulate the production of antibodies. As the antibodies cannot get through the [plasma membrane](#) to attack the intracellular bacteria, the bacteria proliferate unchecked and eventually kill the host.

T_H1 Cells Help Activate Macrophages at Sites of Infection

T_H1 cells are preferentially induced by [antigen](#)-presenting cells that harbor microbes in intracellular vesicles. The bacteria that cause tuberculosis for example, replicate mainly in phagosomes inside macrophages, where they are protected from antibodies. They are also not readily attacked by cytotoxic T cells, which mainly recognize foreign antigens that are produced in the [cytosol](#). The bacteria can survive in phagosomes because they inhibit both the fusion of the phagosomes with lysosomes and the acidification of the phagosomes that is necessary to activate lysosomal hydrolases. Infected dendritic cells recruit helper T cells to assist in the killing of such microbes. The dendritic cells migrate to peripheral lymphoid organs, where they stimulate the production of T_H1 cells, which then migrate to sites of infection to help activate infected macrophages to kill the microbes harboring in their phagosomes.

T_H1 effector cells use two signals to activate a [macrophage](#). They secrete IFN- γ , which binds to IFN- γ receptors on the macrophage surface, and they display the costimulatory [protein CD40 ligand](#), which binds to CD40 on the macrophage. (We see later that CD40 ligand is also used by helper T cells to activate B cells.) Once activated, the macrophage can kill the microbes it contains: lysosomes can now fuse more readily with the phagosomes, unleashing a hydrolytic attack, and the activated macrophage makes oxygen radicals and [nitric oxide](#), both of which are highly toxic to the microbes. Because dendritic cells also express CD40, the T_H1 cells at sites of infection can also help activate them. As a result, the dendritic cells increase their production of class II [MHC](#) proteins, B7 costimulatory proteins, and various cytokines, especially IL-12. This makes them more effective at stimulating helper T cells to differentiate into T_H1 effector cells in peripheral lymphoid organs, providing a positive feedback loop that increases the production of T_H1 cells and, thereby, the activation of macrophages.

T_H1 effector cells stimulate an [inflammatory response](#) by recruiting more phagocytic cells into the infected site. They do so in three ways:

1.

They secrete cytokines that act on the bone marrow to increase the production of monocytes ([macrophage](#) precursors that circulate in the blood) and neutrophils.

2.

They secrete other cytokines that activate endothelial cells lining local blood vessels to express cell adhesion molecules that cause monocytes and neutrophils in the blood to adhere there.

3.

They secrete chemokines that direct the migration of the adherent monocytes and neutrophils out of the bloodstream into the site of infection.

T_H1 cells can also help activate cytotoxic T cells in peripheral lymphoid organs by stimulating dendritic cells to produce more costimulatory proteins. In addition, they can help effector cytotoxic T cells kill [virus](#)-infected target cells, by secreting IFN- γ , which increases the efficiency with which target cells process viral antigens for presentation to cytotoxic T cells (see [Figure 24-59](#)). An effector T_H1 cell can also directly kill some cells itself, including effector lymphocytes: by expressing *Fas* [ligand](#) on its surface, it can induce effector T or B cells that express cell-surface *Fas* to undergo [apoptosis](#).

Both T_H1 and T_H2 cells can help stimulate B cells to proliferate and differentiate into either antibody-secreting effector cells or memory cells. They can also stimulate B cells to switch the class of antibody they make, from IgM (and IgD) to one of the secondary classes of antibody. Before considering how helper T cells do this, we need to discuss the role of the B cell [antigen receptor](#) in the activation of B cells.

Antigen Binding Provides Signal 1 to B Cells

Like T cells, B cells require two types of extracellular signals to become activated. Signal 1 is provided by [antigen](#) binding to the antigen [receptor](#), which is a [membrane](#)-bound antibody [molecule](#). Signal 2 is usually provided by a [helper T cell](#). Like a T cell, if a B cell receives the first signal only, it is usually eliminated or functionally inactivated, which is one way in which B cells become tolerant to self antigens.

Signaling through the B cell [antigen receptor](#) works in much the same way as signaling through the T cell receptor. The receptor is associated with two invariant [protein](#) chains, $Ig\alpha$ and $Ig\beta$, which help convert antigen binding to the receptor into intracellular signals. When antigen cross-links its receptors on the surface of a B cell, it causes the receptors and its associated invariant chains to cluster into small aggregates. This aggregation leads to the assembly of an intracellular signaling [complex](#) at the site of the clustered receptors and to the initiation of a [phosphorylation](#) cascade.

Just as the [CD4](#) and [CD8](#) co-receptors on T cells enhance the efficiency of signaling through the T cell [receptor](#), so a co-receptor [complex](#) that binds complement proteins greatly enhances the efficiency of signaling through the B cell [antigen](#) receptor and its associated invariant chains. If a microbe activates the [complement system](#) (discussed in Chapter 25), complement proteins are often deposited on the microbe surface, greatly increasing the B cell response to the microbe. Now, when the microbe clusters antigen receptors on a B cell, the *complement-binding co-receptor complexes* are brought into the cluster, increasing the

strength of signaling. As expected, antibody responses are greatly reduced in mice lacking either one of the required complement components or complement receptors on B cells.

Later in the [immune response](#), by contrast, when IgG antibodies decorate the surface of the microbe, a different co-[receptor](#) comes into play to dampen down the B cell response. These are *Fc receptors*, which bind the tails of the IgG antibodies. They recruit [phosphatase](#) enzymes into the signaling [complex](#) that decrease the strength of signaling. In this way the Fc receptors on B cells act as inhibitory co-receptors, just as the CTLA-4 proteins do on T cells. Thus, the co-receptors on a T cell or B cell allow the cell to gain additional information about the [antigen](#) bound to its receptors and thereby make a more informed decision as to how to respond.

Unlike T cell receptors, the [antigen](#) receptors on B cells do more than just bind antigen and transmit signal 1. They deliver the antigen to an endosomal [compartment](#) where the antigen is degraded to peptides, which are returned to the B cell surface bound to class II [MHC](#) proteins. The peptide-class-II-MHC complexes are then recognized by effector helper T cells, which can now deliver signal 2. Signal 1 prepares the B cell for its interaction with a [helper T cell](#) by increasing the [expression](#) of both class II MHC proteins and receptors for signal 2.

Helper T Cells Provide Signal 2 to B Cells

Whereas [antigen](#)-presenting cells such as dendritic cells and macrophages are omnivorous and ingest and present antigens nonspecifically, a B cell generally presents only an antigen that it specifically recognizes. In a primary antibody response, naïve helper T cells are activated in a peripheral [lymphoid organ](#) by binding to a foreign peptide bound to a class II [MHC protein](#) on the surface of a [dendritic cell](#). Once activated, the effector [helper T cell](#) can then activate a B cell that specifically displays the same [complex](#) of foreign peptide and class II MHC protein on its surface.

The display of [antigen](#) on the B cell surface reflects the selectivity with which it takes up foreign proteins from the extracellular fluid. These foreign proteins are selected by the antigen receptors on the surface of the B cell and are ingested by [receptor-mediated endocytosis](#). They are then degraded and recycled to the cell surface in the form of peptides bound to class II [MHC](#) proteins. Thus, the [helper T cell](#) activates those B cells with receptors that specifically recognize the antigen that initially activated the T cell, although the T and B cells usually recognize distinct antigenic determinants on the antigen. In secondary antibody responses, memory B cells themselves can act as antigen-presenting cells and activate helper T cells, as well as being the subsequent targets of the effector helper T cells. The mutually reinforcing actions of helper T cells and B cells lead to an [immune response](#) that is both intense and highly specific.

Once a [helper T cell](#) has been activated to become an [effector cell](#) and contacts a B cell, the contact initiates an internal rearrangement of the helper cell [cytoplasm](#). The T cell orients its [centrosome](#) and Golgi apparatus toward the B cell, as described previously for an effector [cytotoxic T cell](#) contacting its target cell. In this case, however, the orientation is thought to enable the effector helper T cell to provide signal 2 by directing both [membrane-bound](#) and secreted signal molecules onto the B cell surface. The membrane-bound [signal molecule](#) is the [transmembrane protein](#) CD40 [ligand](#), which we encountered earlier and is expressed on the surface of effector helper T cell, but not on nonactivated naïve or memory

helper T cells. It is recognized by the CD40 protein on the B cell surface. The interaction between CD40 ligand and CD40 is required for helper T cells to activate B cells to proliferate and differentiate into memory or antibody-secreting effector cells. Individuals that lack CD40 ligand are severely immunodeficient. They are susceptible to the same infections that affect AIDS patients, whose helper T cells have been destroyed.

Secreted signals from helper T cells also help B cells to proliferate and differentiate and, in some cases, to switch the class of antibody they produce. *Interleukin-4 (IL-4)* is one such signal. Produced by T_H2 cells, it collaborates with CD40 [ligand](#) in stimulating B cell proliferation and [differentiation](#), and it promotes switching to IgE antibody production. Mice deficient in IL-4 production are severely impaired in their ability to make IgE.

Some antigens can stimulate B cells to proliferate and differentiate into antibody-secreting effector cells without help from T cells. Most of these *T-cell-independent antigens* are microbial polysaccharides that do not activate helper T cells. Some activate B cells directly by providing both signal 1 and signal 2. Others are large polymers with repeating, identical antigenic determinants; their multipoint binding to B cell [antigen](#) receptors can generate a strong enough signal 1 to activate the B cell directly, without signal 2. Because T-cell-independent antigens do not activate helper T cells, they fail to induce B cell memory, [affinity maturation](#), or [class switching](#), all of which require help from T cells. They therefore mainly stimulate the production of low-affinity (but high-[avidity](#)) IgM antibodies. Most B cells that make antibodies without T cell help belong to a distinct B cell lineage. They are called *B1 cells* to distinguish them from *B2 cells*, which require T cell help. B1 cells seem to be especially important in defense against intestinal pathogens.

Immune Recognition Molecules Belong to an Ancient Superfamily

Most of the proteins that mediate cell-cell recognition or [antigen](#) recognition in the [immune system](#) contain [Ig](#) or Ig-like domains, suggesting that they have a common evolutionary history. Included in this [Ig superfamily](#) are antibodies, T cell receptors, [MHC](#) proteins, the [CD4](#), [CD8](#), and [CD28](#) co-receptors, and most of the invariant [polypeptide](#) chains associated with B and T cell receptors, as well as the various Fc receptors on lymphocytes and other white blood cells. All of these proteins contain one or more Ig or Ig-like domains. In fact, about 40% of the 150 or so polypeptides that have been characterized on the surface of white blood cells belong to this superfamily. Many of these molecules are dimers or higher oligomers in which Ig or Ig-like domains of one chain interact with those in another.

The amino acids in each [Ig-like domain](#) are usually encoded by a separate [exon](#). It seems likely that the entire [gene](#) superfamily evolved from a gene coding for a single Ig-like domain—similar to that encoding β_2 -microglobulin (see [Figure 24-50A](#)) or the Thy-1 [protein](#) that may have mediated cell-cell interactions. There is evidence that such a primordial gene arose before vertebrates diverged from their invertebrate ancestors about 400 million years ago. New family members presumably arose by exon and gene duplications.

The multiple [gene](#) segments that encode antibodies and T cell receptors may have arisen when a [transposable element](#), or transposon (discussed in Chapter 5), inserted into an [exon](#) of a gene encoding an [Ig](#) family member in an ancestral [lymphocyte](#)-like cell. The transposon may have contained the ancestors of the *rag* genes, which, as discussed earlier, encode the

proteins that initiate *V(D)J* joining; the finding that the RAG proteins can act as transposons in a test tube strongly supports this view. Once the transposon had inserted into the exon, the gene could be expressed only if the transposon was excised by the RAG proteins and the two ends of the exon were rejoined, much as occurs when the *V* and *J* gene segments of an Ig [light chain](#) gene are assembled. A second insertion of the transposon into the same exon may then have divided the gene into three segments, equivalent to the present-day *V*, *D*, and *J* gene segments. Subsequent duplication of either the individual gene segments or the entire split gene may have generated the arrangements of gene segments that characterize the adaptive **immune** systems of present-day vertebrates.

Adaptive **immune** systems evolved to defend vertebrates against infection by pathogens. Pathogens, however, evolve more quickly, and they have acquired remarkably sophisticated strategies to counter these defenses.

Summary

Naïve T cells require at least two signals for activation. Both are provided by an [antigen-presenting cell](#), which is usually a [dendritic cell](#): signal 1 is provided by [MHC](#)-peptide complexes binding to T cell receptors, while signal 2 is mainly provided by B7 costimulatory proteins binding to [CD28](#) on the T cell surface. If the T cell receives only signal 1, it is usually deleted or inactivated. When helper T cells are initially activated on a dendritic cell, they can differentiate into either T_H1 or T_H2 effector cells, depending on the cytokines in their environment: T_H1 cells activate macrophages, cytotoxic T cells, and B cells, while T_H2 cells mainly activate B cells. In both cases, the effector helper T cells recognize the same [complex](#) of foreign peptide and class II [MHC protein](#) on the target cell surface as they initially recognized on the dendritic cell that activated them. They activate their target cells by a combination of [membrane](#)-bound and secreted signal proteins. The membrane-bound signal is [CD40 ligand](#). Like T cells, B cells require two simultaneous signals for activation. Antigen binding to the B cell antigen receptors provides signal 1, while effector helper T cells provide signal 2 in the form of [CD40 ligand](#) and various cytokines.

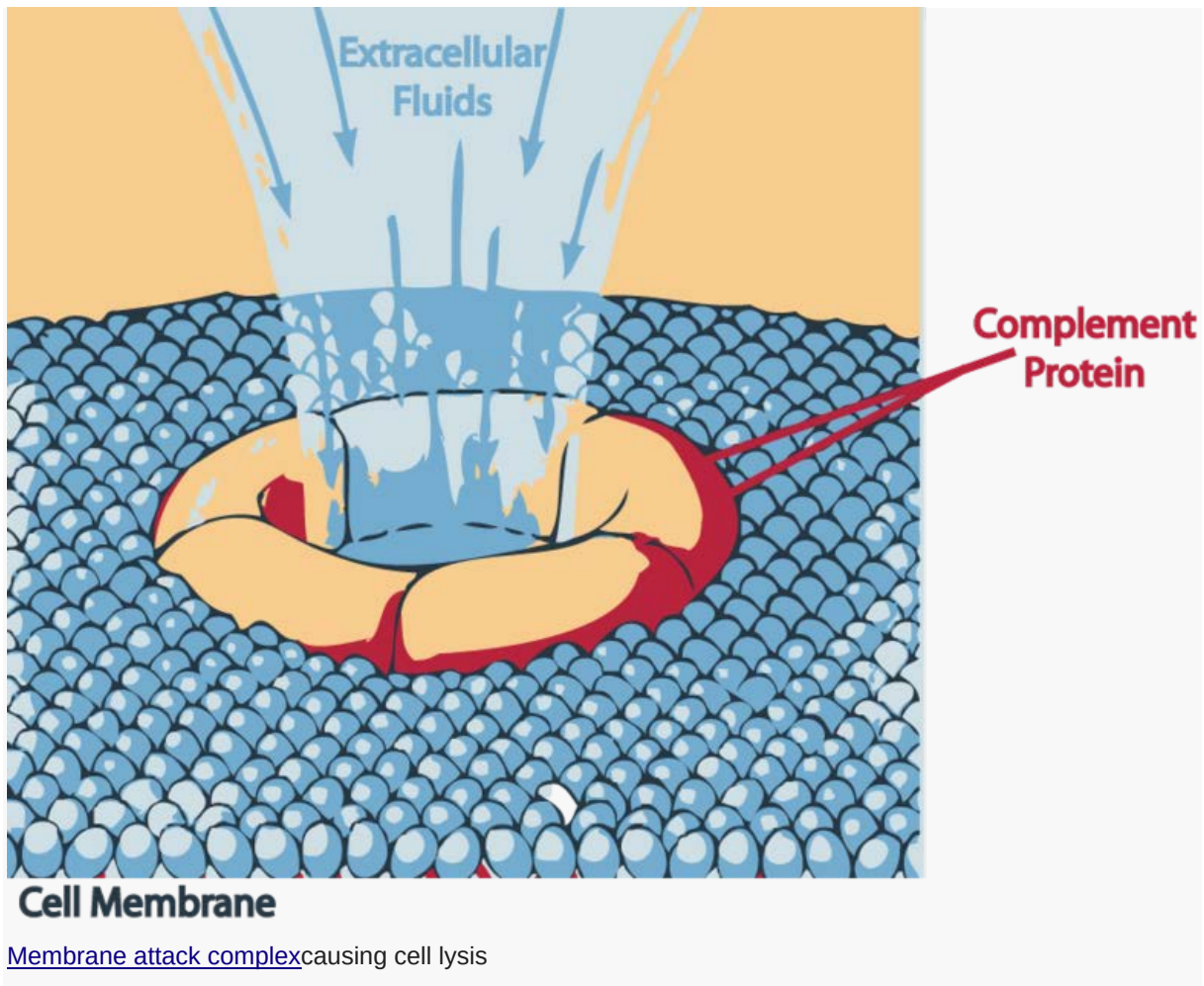
Most of the proteins involved in cell-cell recognition and [antigen](#) recognition in the [immune system](#), including antibodies, T cell receptors, and [MHC](#) proteins, as well as the various co-receptors discussed in this chapter, belong to the ancient [Ig superfamily](#). This superfamily is thought to have evolved from a primordial [gene](#) encoding a single Ig-like [domain](#).

The **complement system**

The **complement system** is a part of the [immune system](#) that helps or [complements](#) the ability of [antibodies](#) and [phagocytic](#) cells to clear [pathogens](#) from an organism. It is part of the [innate](#)

Ehrlich believed that each antigen-specific amboceptor has its own specific complement, whereas Bordet believed that there is only one type of complement. In the early 20th century, this controversy was resolved when it became understood that complement can act in combination with specific antibodies, or on its own in a non-specific way.

Functions



The following are the basic functions of complement:

1. **Opsonization** – enhancing phagocytosis of antigens
2. **Chemotaxis** – attracting macrophages and neutrophils
3. **Cell Lysis** – rupturing membranes of foreign cells
4. **Agglutination** – clustering and binding of pathogens together (sticking)

Overview

The proteins and glycoproteins that constitute the complement system are synthesized by hepatocytes. But significant amounts are also produced by tissue macrophages, blood monocytes, and epithelial cells of the genitourinary tract and gastrointestinal tract. The three pathways of activation all generate homologous variants of the [protease C3-convertase](#). The

classical complement pathway typically requires antigen:antibody complexes ([immune complexes](#)) for activation (specific immune response), whereas the alternative and mannose-binding lectin pathways can be activated by C3 hydrolysis or antigens without the presence of antibodies (non-specific immune response). In all three pathways, C3-convertase cleaves and activates component [C3](#), creating C3a and C3b, and causing a cascade of further cleavage and activation events. C3b binds to the surface of pathogens, leading to greater internalization by [phagocytic cells](#) by [opsonization](#). [C5a](#) is an important [chemotactic protein](#), helping recruit inflammatory cells. C3a is the precursor of an important [cytokine](#) (adipokine) named [ASP](#) (although this is not universally accepted ^[3]) and is usually rapidly cleaved by [carboxypeptidase B](#). Both C3a and C5a have [anaphylatoxin](#) activity, directly triggering [degranulation](#) of [mast cells](#) as well as increasing vascular permeability and [smooth muscle](#) contraction.^[3] C5b initiates the [membrane attack pathway](#), which results in the [membrane attack complex](#)(MAC), consisting of C5b, [C6](#), [C7](#), [C8](#), and polymeric [C9](#).^[4] MAC is the cytolytic endproduct of the complement cascade; it forms a transmembrane channel, which causes [osmotic](#) lysis of the target cell. [Kupffer cells](#) and other macrophage cell types help clear complement-coated pathogens. As part of the innate immune system, elements of the complement cascade can be found in species earlier than vertebrates; most recently in the protostome [horseshoe crab](#) species, putting the origins of the system back further than was previously thought.

Classical pathway

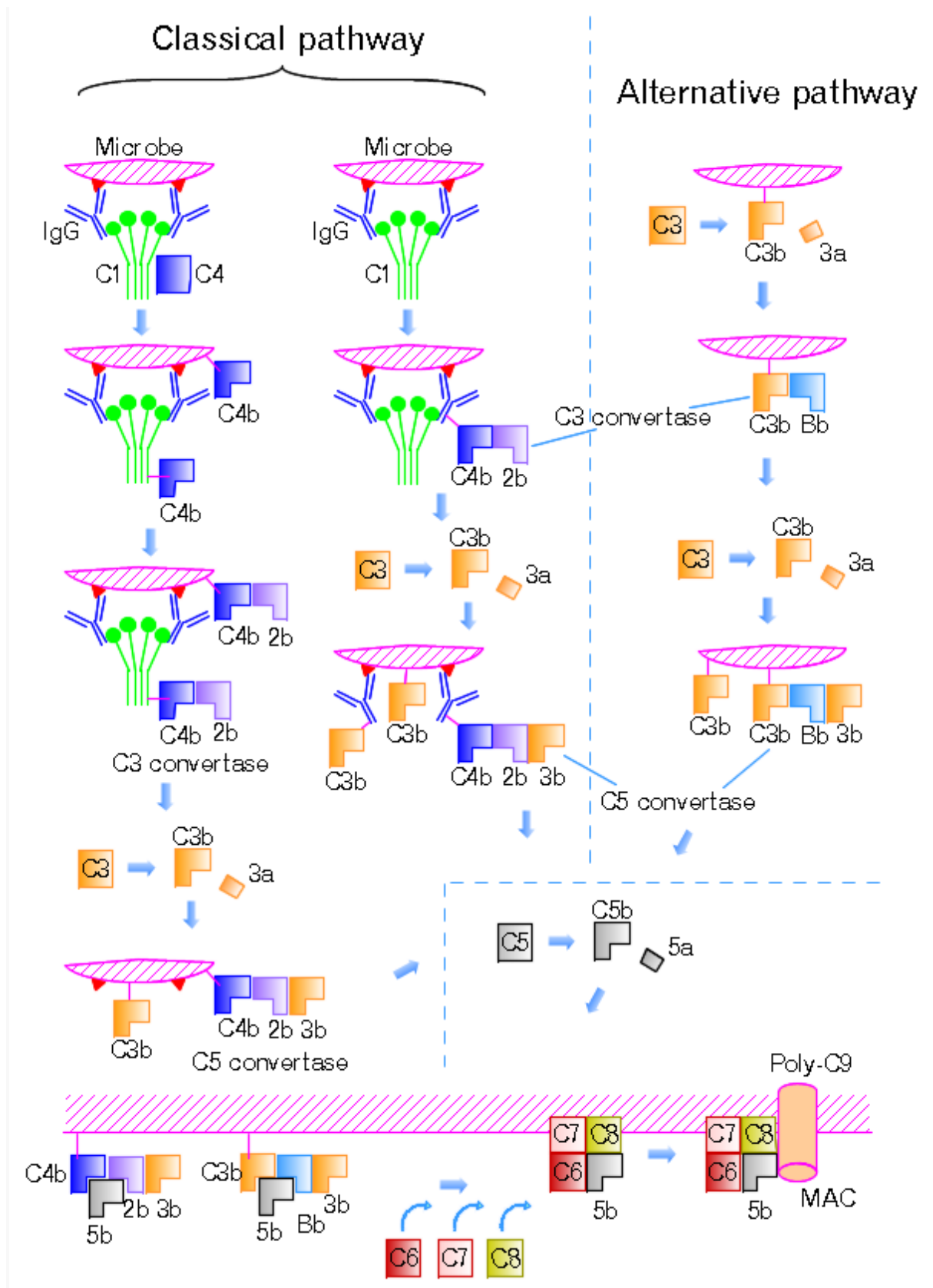


Figure 2. The classical and alternative complement pathways

C2 fragment nomenclature

Different assignment for the fragments C2a and C2b, as to which is larger or smaller, is found below in several current text books in immunology; however, we might safely make assignment that the former is smaller. In a literature below, in the publishing year of as early as 1994,^[5] they commented that the larger fragment of C2 should be designated C2b. In the 4th edition of their book, they say that:^[6]

"It is also useful to be aware that the larger active fragment of C2 was originally designated C2a, and is still called that in some texts and research papers. Here, for consistency, we shall call all large fragments of complement **b**, so the larger fragment of C2 will be designated C2b. In the classical and lectin pathways the C3 convertase enzyme is formed from membrane-bound C4b with C2b" (p. 341).

This nomenclature is used in another literature:^[7]

"(Note that, in older texts, the smaller fragment is often called C2b, and the larger one is called C2a for historical reason.)" (p. 332).

The assignment is mixed in the latter literature, though.

Literature^{[8][9][10][11][12][13][14][15][16]} can be found where the larger and smaller fragments are assigned to be C2a and C2b, respectively, and literature^{[5][6][17][18][19]} can be found where the opposite assignment is made. However, due to the widely established convention, C2b here is the larger fragment, which, in the classical pathway, forms C4b2b (classically C4b2a). It may be noteworthy that, in a series of editions of Janeway's book, 1st to 7th, in the latest edition^[15] they withdraw the stance to indicate the larger fragment of C2 as C2b.

The [classical pathway](#) is triggered by activation of the C1-complex. The C1-complex is composed of 1 molecule of C1q, 2 molecules of C1r and 2 molecules of C1s, or $C1qr^2s^2$. This occurs when C1q binds to [IgM](#) or [IgG](#) complexed with [antigens](#). A single pentameric IgM can initiate the pathway, while several, ideally six, IgGs are needed. This also occurs

when [C1q](#) binds directly to the surface of the pathogen. Such binding leads to conformational changes in the C1q molecule, which leads to the activation of two C1r molecules. C1r is a serine protease. They then cleave [C1s](#) (another serine protease). The C1r²s² component now splits [C4](#) and then [C2](#), producing C4a, C4b, C2a, and C2b. C4b and C2a bind to form the classical pathway C3-convertase (C4b2a complex), which promotes cleavage of C3 into C3a and C3b; C3b later joins with C4b2a (the C3 convertase) to make C5 convertase (C4b2a3b complex). The inhibition of C1r and C1s is controlled by [C1-inhibitor](#).

C3-convertase can be inhibited by [Decay accelerating factor](#) (DAF), which is bound to erythrocyte plasma membranes via a [GPI](#) anchor.

[Paroxysmal nocturnal hemoglobinuria](#) is caused by complement breakdown of RBCs due to an inability to make GPI. Thus the RBCs are not protected by GPI anchored proteins such as DAF.

Alternative pathway

The [alternative pathway](#) is continuously activated at a low level, analogous to a car engine at idle, as a result of spontaneous [C3](#) hydrolysis due to the breakdown of the internal thioester bond (C3 is mildly unstable in aqueous environment). The alternative pathway does not rely on pathogen-binding antibodies like the other pathways.^[2] C3b that is generated from C3 by a C3 convertase enzyme complex in the fluid phase is rapidly inactivated by [factor H](#) and [factor I](#), as is the C3b-like C3 that is the product of spontaneous cleavage of the internal thioester. In contrast, when the internal thioester of C3 reacts with a hydroxyl or amino group of a molecule on the surface of a cell or pathogen, the C3b that is now covalently bound to the surface is protected from factor H-mediated inactivation. The surface-bound C3b may now bind [factor B](#) to form C3bB. This complex in the presence of [factor D](#) will be cleaved into Ba and Bb. Bb will remain associated with C3b to form C3bBb, which is the alternative pathway C3 convertase. The C3bBb complex is stabilized by binding oligomers of [factor P](#). The stabilized C3 convertase, C3bBbP, then acts enzymatically to cleave much more C3, some of which becomes covalently attached to the same surface as C3b. This newly bound C3b recruits more B, D and P activity and greatly amplifies the complement activation. When complement is activated on a cell surface, the activation is limited by endogenous complement regulatory proteins, which include [CD35](#), [CD46](#), [CD55](#) and [CD59](#), depending on the cell. Pathogens, in general, don't have complement regulatory proteins (there are many exceptions, which reflect adaptation of microbial pathogens to vertebrate immune defenses). Thus, the alternative complement pathway is able to distinguish self from non-self on the basis of the surface expression of complement regulatory proteins. Host cells don't accumulate cell surface C3b (and the proteolytic fragment of C3b called iC3b) because this is prevented by the complement regulatory proteins, while foreign cells, pathogens and abnormal surfaces may be heavily decorated with C3b and iC3b. Accordingly, the alternative complement pathway is one element of [innate immunity](#).

Once the alternative C3 convertase enzyme is formed on a pathogen or cell surface, it may bind covalently another C3b, to form C3bBbC3bP, the C5 convertase. This enzyme then cleaves C5

to C5a, a potent [anaphylatoxin](#), and C5b. The C5b then recruits and assembles C6, C7, C8 and multiple C9 molecules to assemble the [membrane attack complex](#). This creates a hole or pore in the membrane that can kill or damage the pathogen or cell.

Lectin pathway

The [lectin](#) pathway is homologous to the classical pathway, but with the opsonin, [mannose-binding lectin](#) (MBL), and [ficolins](#), instead of C1q. This pathway is activated by binding of MBL to mannose residues on the pathogen surface, which activates the MBL-associated serine proteases, MASP-1, and MASP-2 (very similar to C1r and C1s, respectively), which can then split C4 into C4a and C4b and C2 into C2a and C2b. C4b and C2a then bind together to form the classical C3-convertase, as in the classical pathway. Ficolins are homologous to MBL and function via MASP in a similar way. Several single-nucleotide polymorphisms have been described in M-ficolin in humans, with effect on ligand-binding ability and serum levels. Historically, the larger fragment of C2 was named C2a, but it is now referred as C2b.^[20] In invertebrates without an adaptive immune system, ficolins are expanded and their binding specificities diversified to compensate for the lack of pathogen-specific recognition molecules.

Activation of complements by antigen-associated antibody

In the classical pathway, C1 binds with its C1q subunits to Fc fragments (made of CH₂ region) of IgG or IgM, which has formed a complex with antigens. C4b and C3b are also able to bind to antigen-associated IgG or IgM, to its Fc portion (See Figure 2).

Such immunoglobulin-mediated binding of the complement may be interpreted as that the complement uses the ability of the immunoglobulin to detect and bind to non-self antigens as its guiding stick. The complement itself is able to bind non-self pathogens after detecting their [pathogen-associated molecular patterns](#) (PAMPs), however, utilizing specificity of antibody, complements are able to detect non-self enemies much more specifically. There must be mechanisms that complements bind to Ig but would not focus its function to Ig but to the antigen.

Figure 2 shows the classical and the alternative pathways with the late steps of complement activation schematically. Some components have a variety of binding sites. In the classical pathway C4 binds to Ig-associated C1q and C1r²s² enzyme cleave C4 to C4b and 4a. C4b binds to C1q, antigen-associated Ig (specifically to its Fc portion), and even to the microbe surface. C3b binds to antigen-associated Ig and to the microbe surface. Ability of C3b to bind to antigen-associated Ig would work effectively against antigen-antibody immune complexes to make them soluble. In the figure, C2b refers to the larger of the C2 fragments.

Regulation

The complement system has the potential to be extremely damaging to host tissues, meaning its activation must be tightly regulated. The complement system is regulated by [complement control proteins](#), which are present at a higher concentration in the blood plasma than the complement proteins themselves. Some complement control proteins are present on the membranes of self-cells preventing them from being targeted by complement. One example is [CD59](#), also known

as [protectin](#), which inhibits C9 polymerisation during the formation of the [membrane attack complex](#). The classical pathway is inhibited by [C1-inhibitor](#), which binds to C1 to prevent its activation.

Neoplastic Transformations

Proto-oncogene

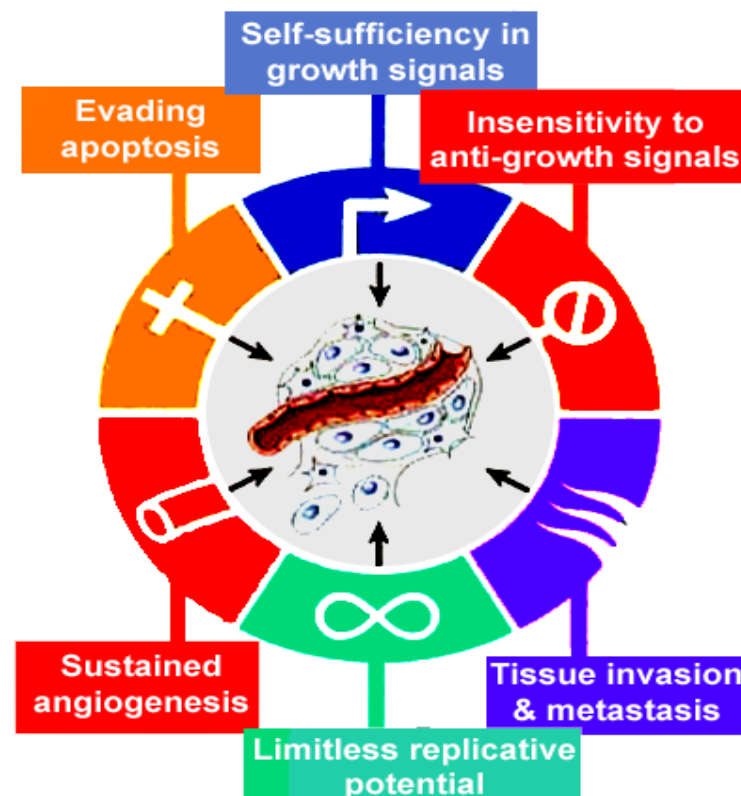
A proto-oncogene is a normal gene that can become an oncogene, either after mutation or increased expression. Proto-oncogenes code for proteins that help to regulate cell growth and differentiation. Proto-oncogenes are often involved in signal transduction and execution of mitogenic signals, usually through their protein products. Upon activation, a proto-oncogene (or its product) becomes a tumor inducing agent, an oncogene.

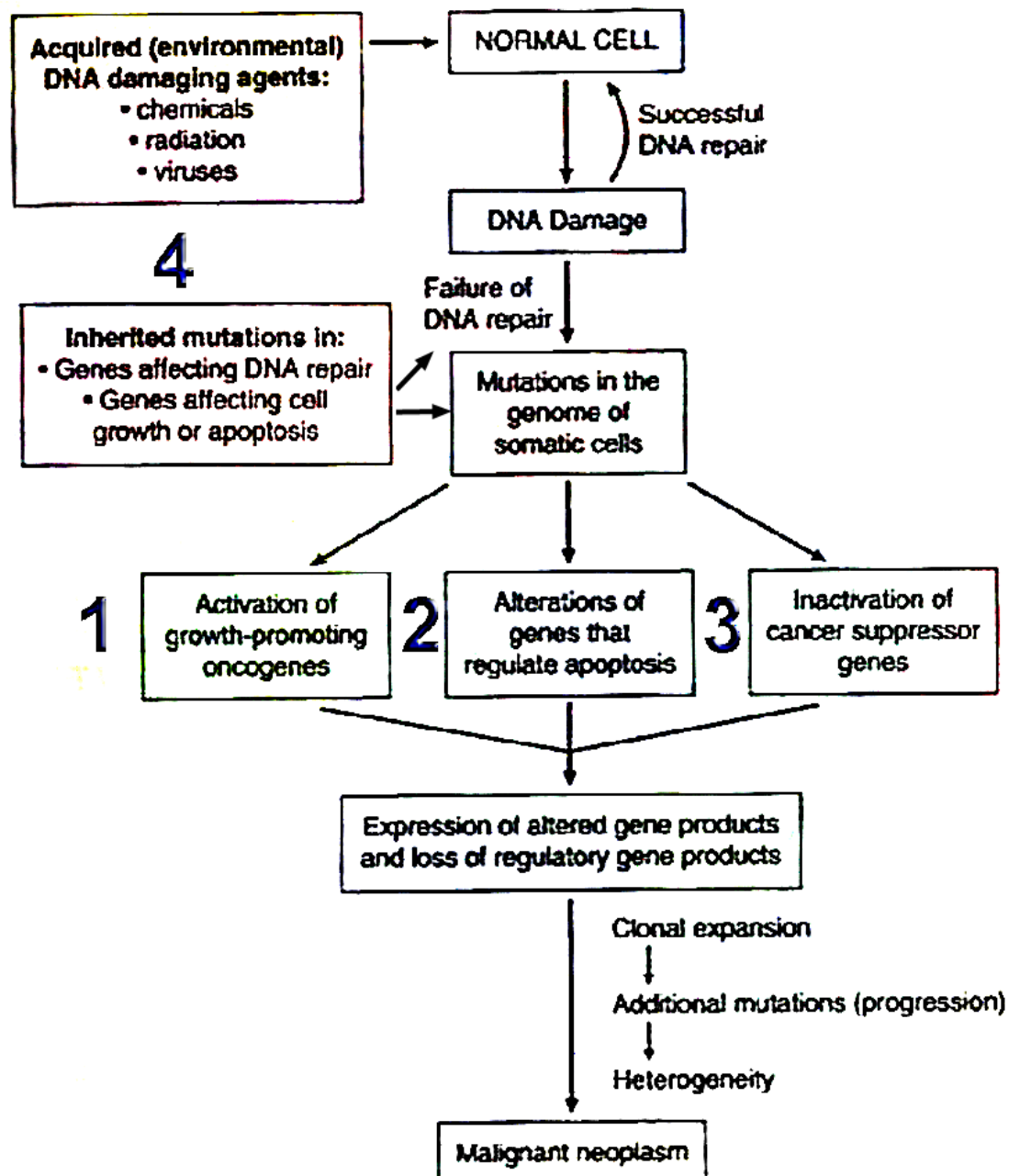
Some oncogenes, usually involved in early stages of cancer development, increase the chance that a normal cell develops into a tumor cell, possibly resulting in cancer

Classes of Oncogenes

- i. Growth Promoting Oncogenes
- ii. Growth-Inhibiting Cancer Suppressor Genes
- iii. DNA Repairing Genes
- iv. Apoptosis Genes

Neoplastic Transformations





i. Growth Promoting Oncogenes

Autonomous Cell Growth in Cancer Cells Is Promoted By Following Elements

- Growth Factors
- Growth Factor Receptors
- Signal-Transducing Proteins (secondary messengers)
- Nuclear Transcription Factors
- Cyclins and Cyclin Dependent Kinases (CDKs)

a) Growth Factors

Some cancer cells acquire the ability through mutations to produce growth factors to which they are responsive (autocrine signaling).

Many glioblastomas secrete PDGF and sarcomas make TGF- α . Similar loops are fairly common many types of cancers.

b) Growth Factor Receptors

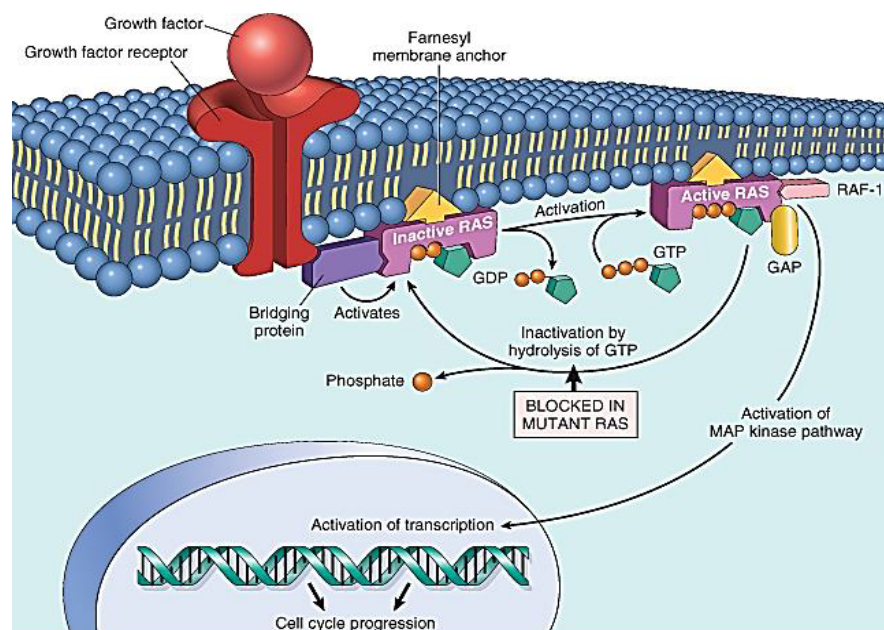
Mutated Growth Factor Receptors continue to deliver continuous signals in the absence of the ligand.

EGF receptor ERBB1 is over expressed in 80% of SCC of lung.

ERBB2/HER2 is amplified in 25%-30% of breast cancers and adenocarcinomas of lung, ovary and salivary glands.

These tumours are sensitive to small amounts of growth factors.

c) Signal-Transducing Proteins

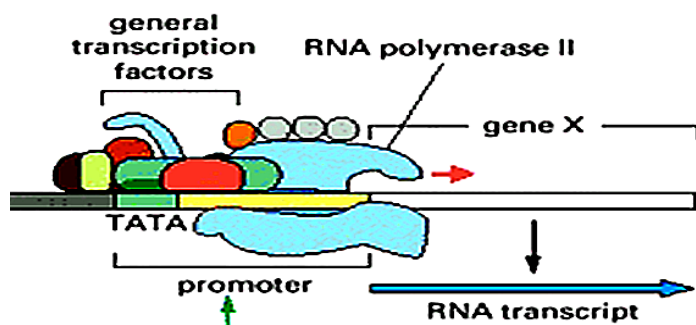


Mutation of RAS is the most common oncogene abnormality, approximately 30% of all human tumors contain mutated version of RAS gene.

GAP/neurofibramin 1 (a RAS deactivating protein) is mutated in neurofibramatosis type 1.

BCR-ABL hybrid gene activates RAS pathways (like p53 Normal ABL gene also promotes apoptosis via NLS domain).

d) Nuclear Transcription Factors

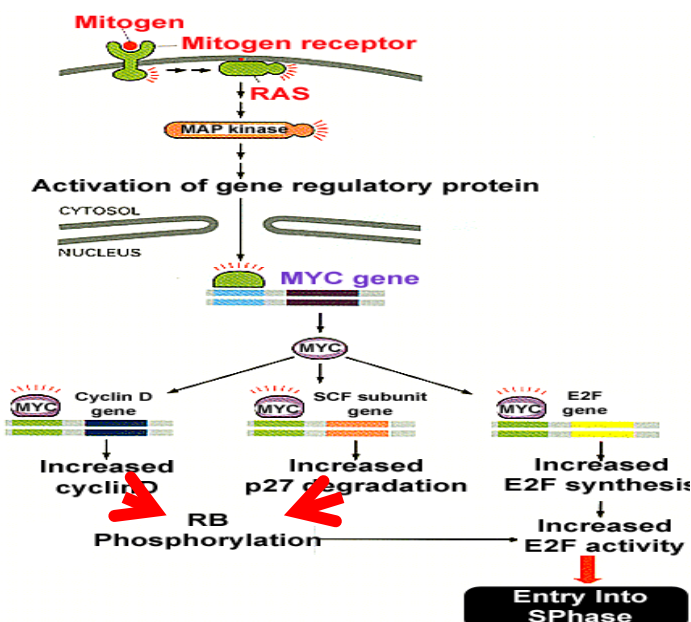


MYC and the polymerase assemble on a specific DNA sequence and drive transcription of genes.

MYC protooncogene is expressed in virtually all the cells, and the MYC protein is rapidly induced when cells receive a signal to divide.

MYC binds DNA and causes transcriptional activation of many growth related proteins.

Dysregulation of of MYC gene occurs in Burkitt Lymphoma (B-cell tumor); MYC is also amplified in many other cancers.



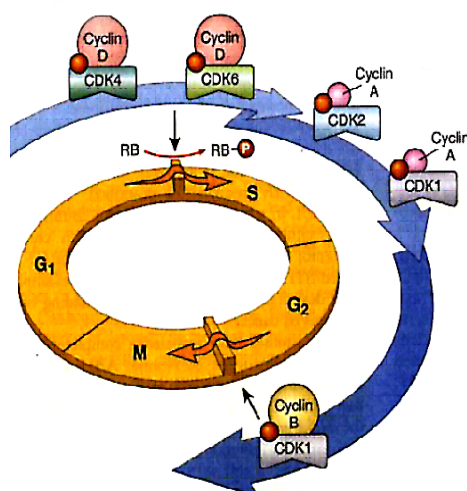
The binding of mitogens to cell-surface receptors leads to the activation of Ras and a MAP kinase cascade.

One effect of this pathway is the increased production of the gene regulatory protein Myc.

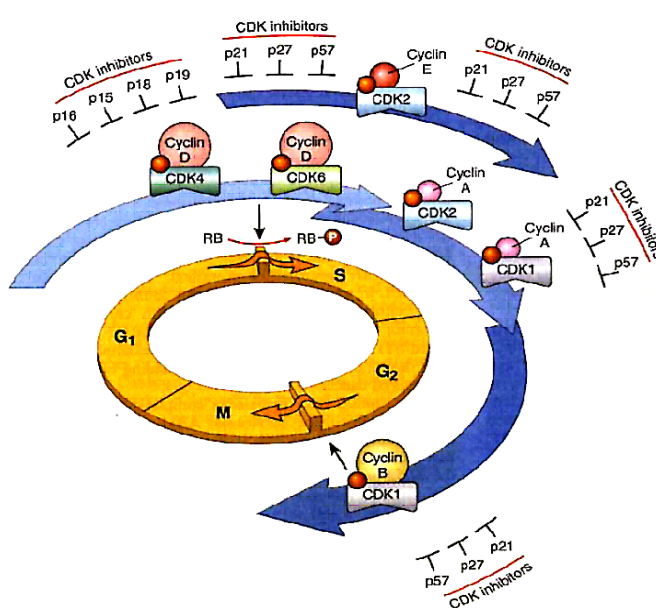
Myc increases the transcription of several genes, including the gene encoding cyclin D.

e) Growth Promoting Oncogenes Cyclins & CDKs

CDKs Phosphorylate crucial target proteins & are expressed constitutively during cell cycle in inactive form however cyclins are synthesized during specific phases of cell cycle.



Activated CDK4 and CDK6 phosphorylate RB which overcome the hurdle of G₁ to S hurdle.

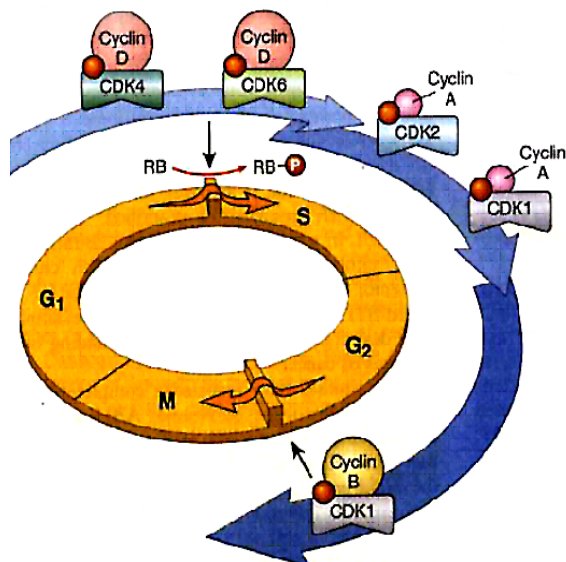


The activity of CDKs is regulated by CDK inhibitors (CKIs).

Cyclin D genes are over expressed in many cancers, e.g. affecting the breast, esophagus, liver & subset of lymphomas.

Mutations affecting other cyclins and CDKs also occur in some neoplasms.

TGF- β stimulates production of p15 and inhibits the transcription of CDK2, CDK4 and cyclin A.



Activated CDK4 and CDK6 phosphorylate RB which overcome the hurdle of G1 to S hurdle

ii. Growth-Inhibiting Cancer Suppressor Genes

The signals and the signal-transducing pathways for growth inhibition are less well understood than those for growth promotion.

The antigrowth signals may cause dividing cells to go into G0 (dormancy) where they remain signalled not to divide or they may cause cells to enter postmitotic differentiation.

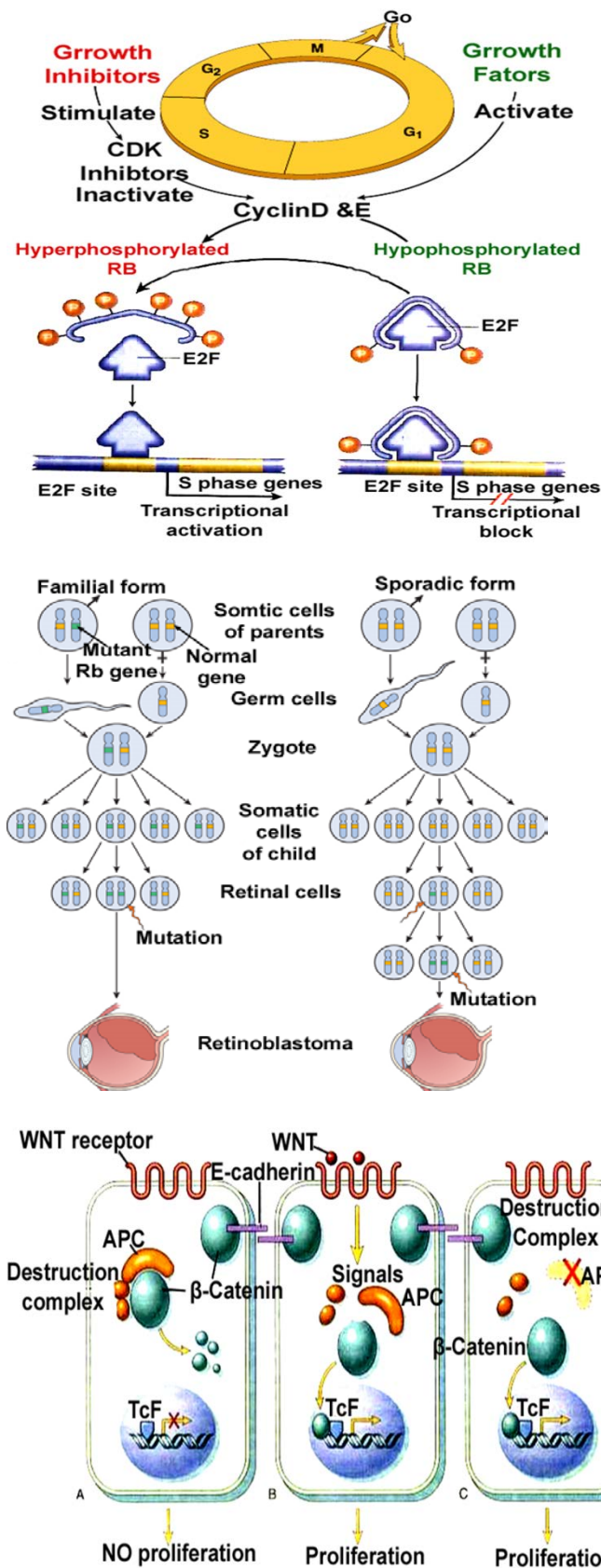
Activated retinoblastoma (RB) protein prevents cells from advancing from G1 to S phase.

RB is the first Tumor Suppressor Gene. RB gene product is a DNA binding protein and is ubiquitously expressed.

Hypophosphorylated (active) RB serves as a break for G1 to S phase transition, Hyperphosphorylated (inactive) RB allows cells to divide.

Hypophosphorylated RB binds transcriptional factor E2F when RB is phosphorylated by CDK4, CDK6 and CDK2 it releases E2F.

RB target of some oncogenic viruses neutralizing.



iii. DNA Repairing Genes

p53 has many anti-cancer mechanisms:

It can activate DNA repair proteins when DNA has sustained damage.

p53 can also arrest the cell cycle at the G1/S regulation point on DNA damage recognition (if it holds the cell here for long enough, the DNA repair proteins will have time to fix the damage & the cell will be allowed to continue the cell cycle.)

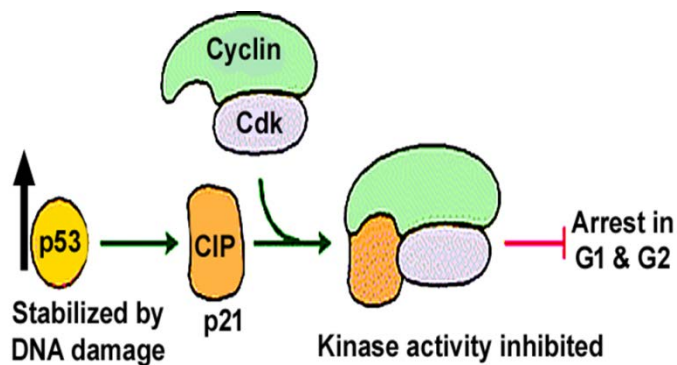
It can initiate apoptosis, the programmed cell death, if the DNA damage proves to be irreparable.

p53 is one of the most commonly mutated genes in human cancers (it can sense DNA damage via other proteins e.g. ATM).

p53 has an anti-proliferative effects and also regulates apoptosis.

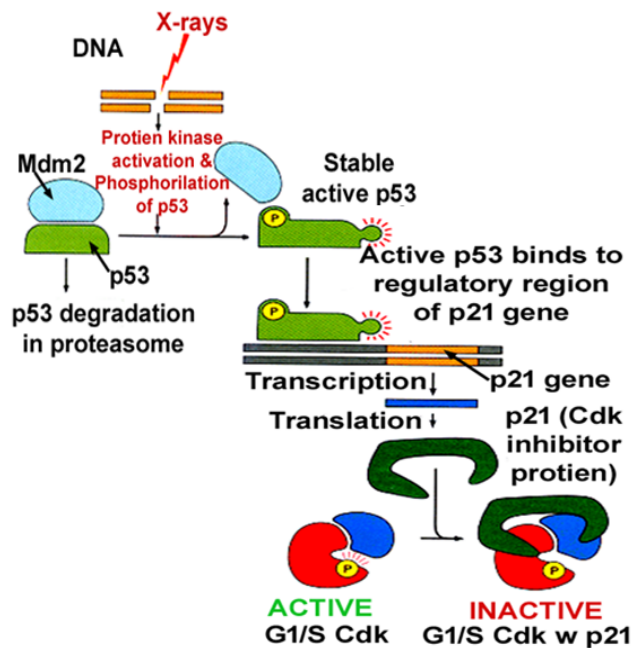
The normally unstable p53 protein is stabilized by damaged DNA, so its concentration increases.

Acting as a transcription factor, p53 induces expression of p21, a cyclin-kinase inhibitor that inhibits all Cdk (Cdk1-, Cdk2-, Cdk4-, and Cdk6-cyclin complexes).



Binding of p21 to these Cdk-cyclin complexes leads to cell cycle arrest in G1 and G2.

TP53/ P53

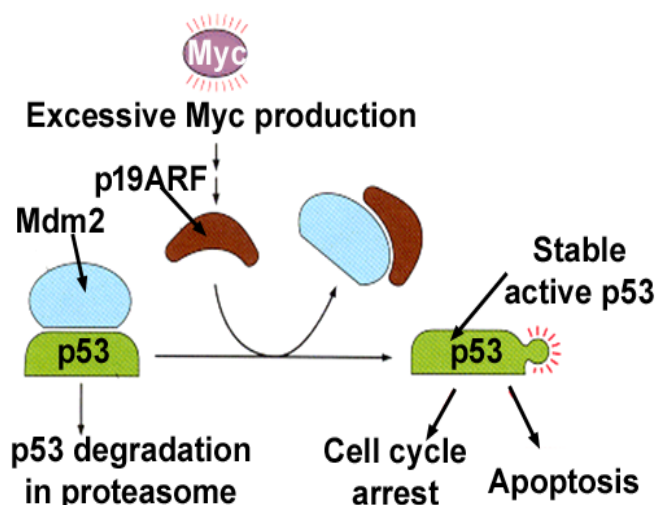


When DNA is damaged, protein kinases that phosphorylate p53 are activated.

Mdm2 normally binds to p53 and promotes its ubiquitylation and destruction in proteasomes.

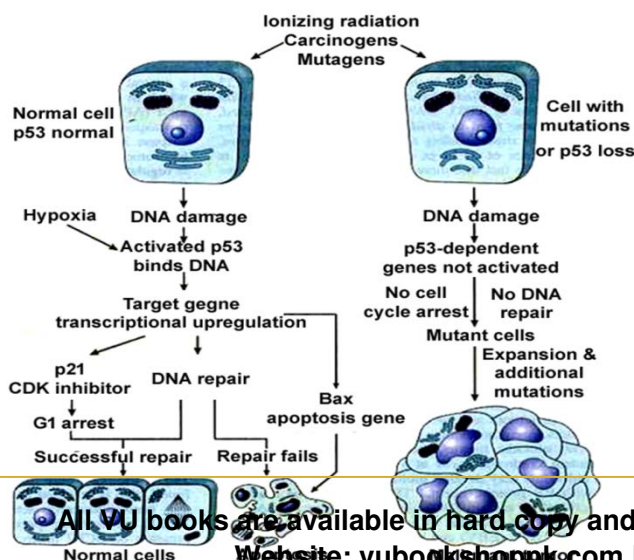
Phosphorylation of p53 blocks its binding to Mdm2; as a result, p53 accumulates to high levels and stimulates transcription of the gene that encodes the CKI protein p21.

The p21 binds and inactivates G1/S-Cdk and S-Cdk complexes, arresting the cell in G1. In some cases, DNA damage also induces either the phosphorylation of Mdm2 or a decrease in Mdm2 production, which causes an increase in p53.



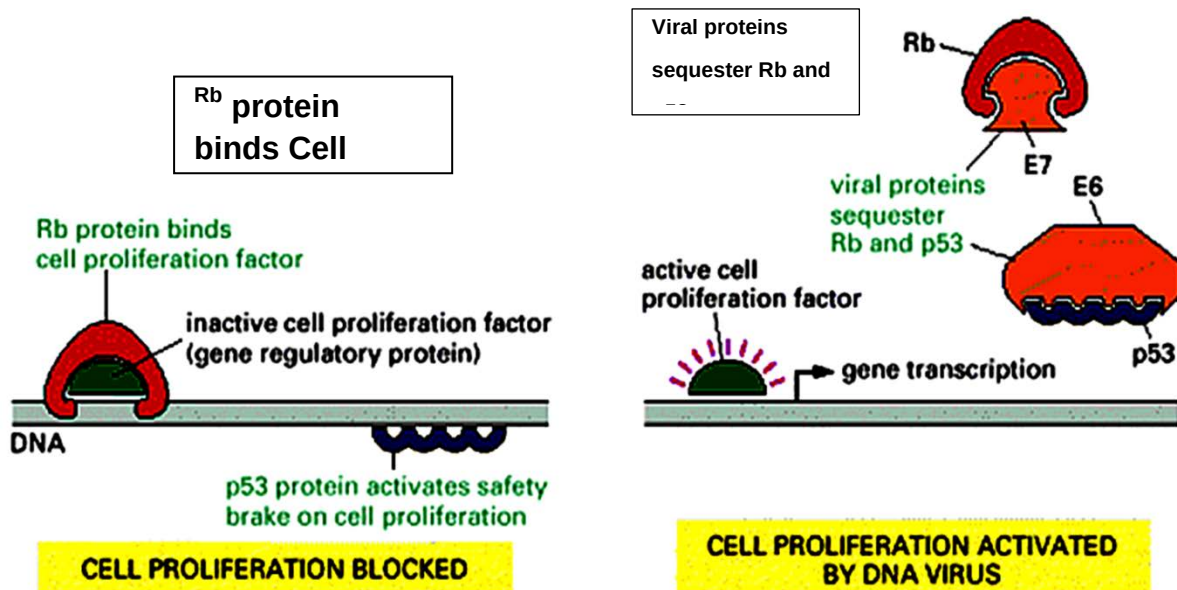
Abnormally high levels of Myc cause the activation of p19ARF, which binds and inhibits Mdm2 and thereby causes increased p53 levels.

Depending on the cell type p53 then causes either cell-cycle arrest or apoptosis.



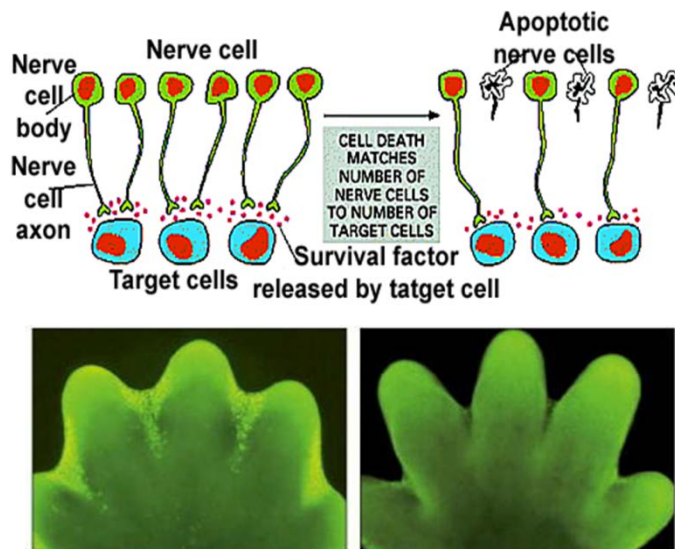
Li-Fraumeni syndrome: One of the P53 alleles is mutated which causes 25 fold increased chance of developing a malignant tumor by age 50

The Viral Protein Targets



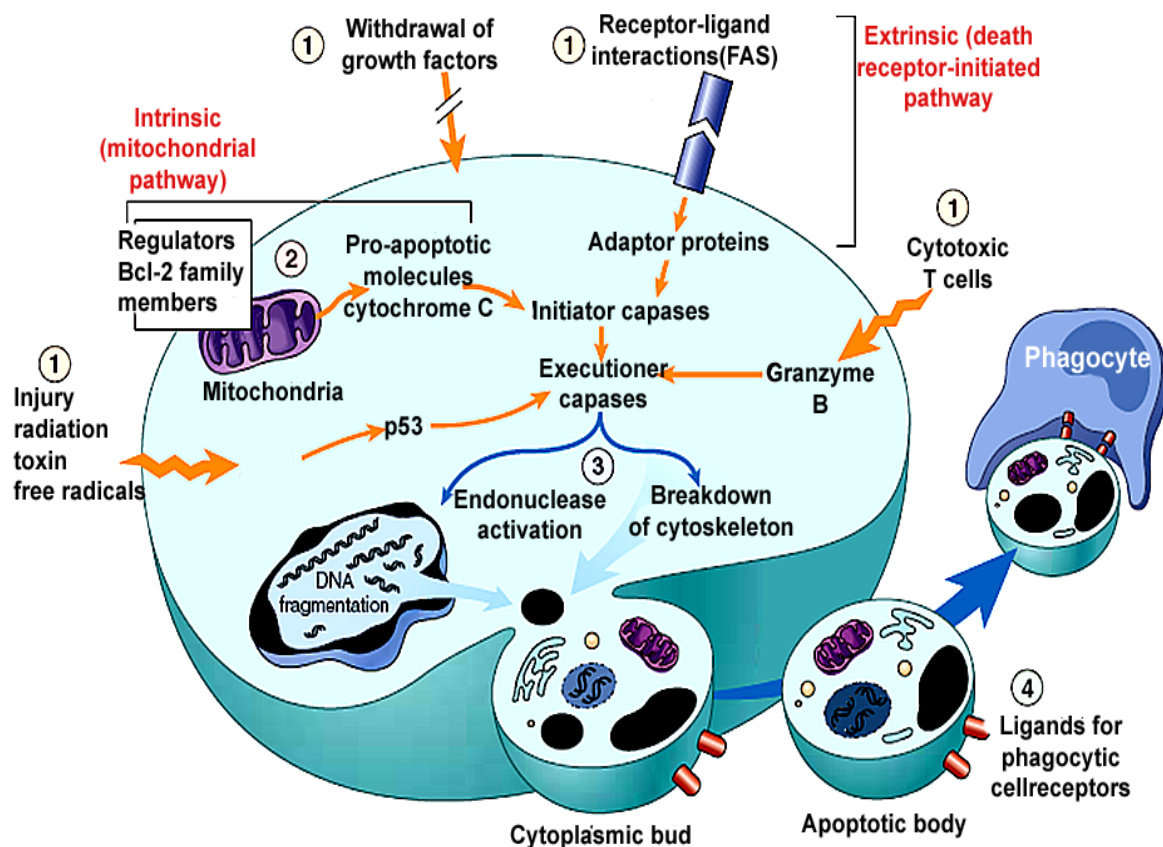
Papillomavirus uses two viral proteins, E6 and E7, to sequester the host cell's p53 and Rb. The SV40 virus uses a single dual-purpose protein called large T antigen, for the same purpose.

iv. Apoptosis Genes



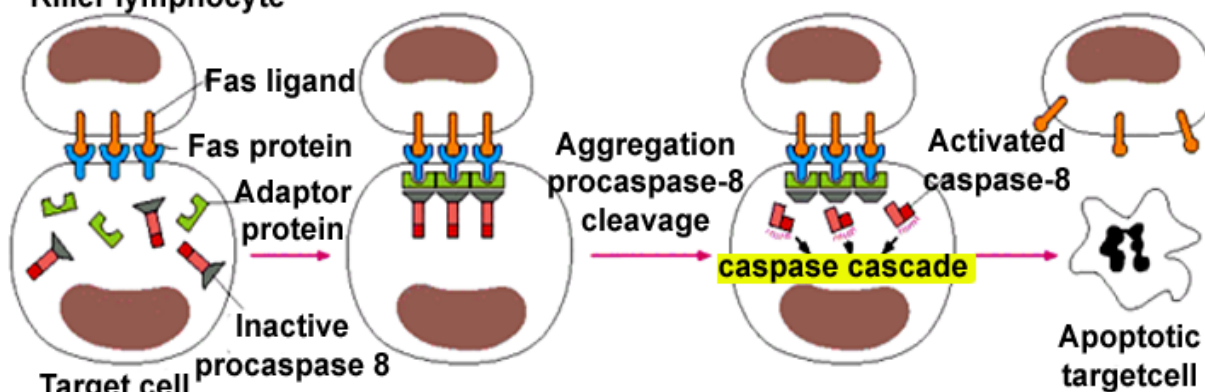
Billions of bone cells die in a healthy adult human

In adults, cell death exactly balances cell division



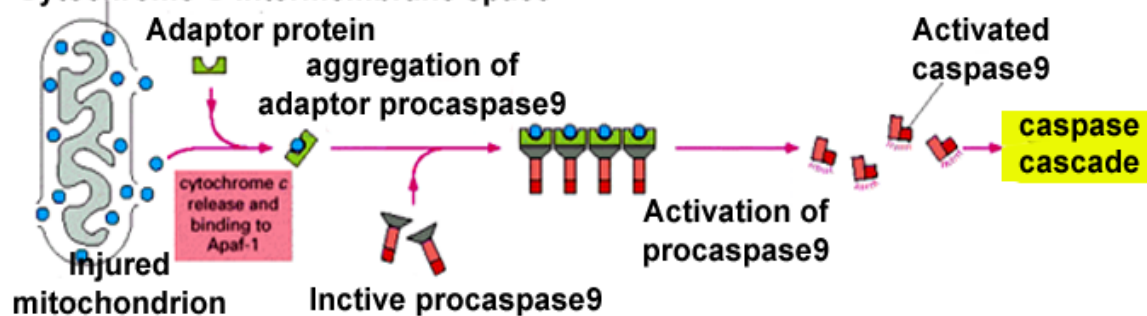
(A) Activation of apoptosis from outside the cell

Killer lymphocyte

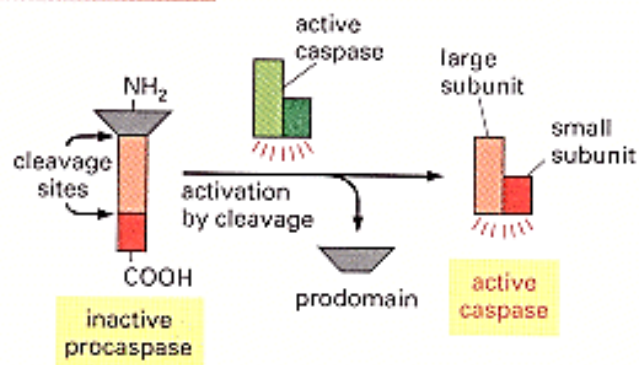


(B) Activation of apoptosis from inside the cell

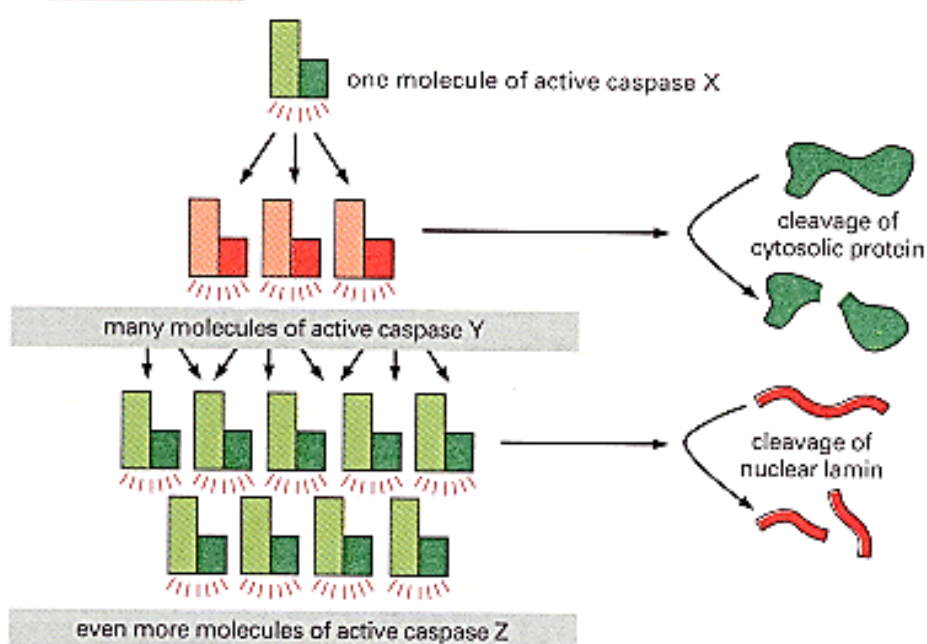
Cytochrome C intermembrane space

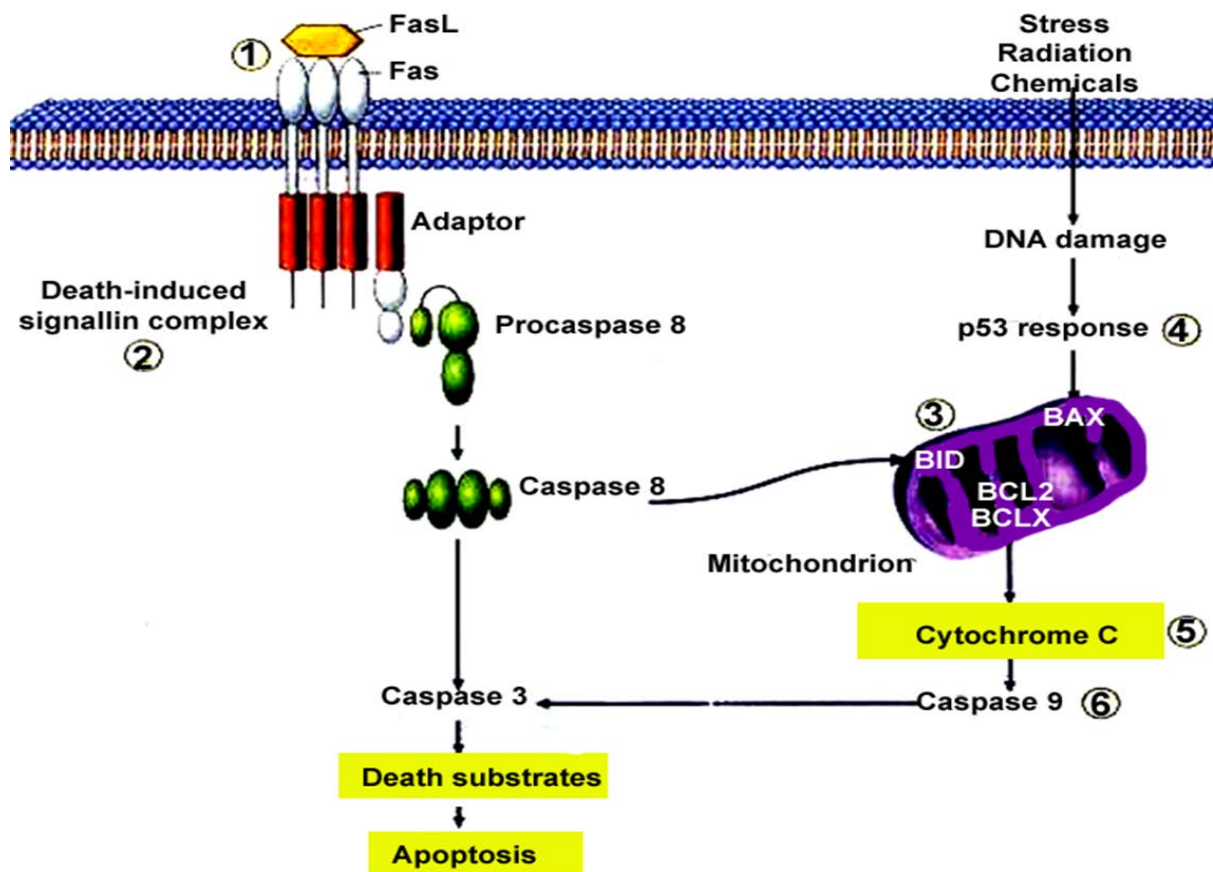


(A) procaspase activation



(B) caspase cascade





Protein Sorting

The Compartmentalization of Cells

In this introductory [section](#) we present a brief overview of the compartments of the cell and the relationships between them. In doing so, we organize the organelles conceptually into a small number of discrete families and discuss how proteins are directed to specific organelles and how they cross [organelle](#) membranes.

the ER and dispatches them to a variety of destinations, usually covalently modifying them *en route*.

Mitochondria and (in plants) *chloroplasts* generate most of the ATP used by cells to drive reactions that require an input of free energy; chloroplasts are a specialized version of *plastids*, which can also have other functions in plant cells, such as the storage of food or pigment molecules. *Lysosomes* contain digestive enzymes that degrade defunct intracellular organelles, as well as macromolecules and particles taken in from outside the cell by [endocytosis](#). On their way to lysosomes, endocytosed material must first pass through a series of organelles called *endosomes*. *Peroxisomes* are small vesicular compartments that contain enzymes utilized in a variety of oxidative reactions.

In general, each [membrane](#)-enclosed [organelle](#) performs the same set of [basic](#) functions in all cell types. But to serve the specialized functions of cells, these organelles will vary in abundance and can have additional properties that differ from cell type to cell type.

On average, the [membrane](#)-enclosed compartments together occupy nearly half the volume of a cell, and a large amount of intracellular membrane is required to make them all. In liver and pancreatic cells, for example, the [endoplasmic reticulum](#) has a total membrane surface area that is, respectively, 25 times and 12 times that of the [plasma membrane](#). In terms of its area and mass, the plasma membrane is only a minor membrane in most eucaryotic.

Membrane-enclosed organelles often have characteristic positions in the [cytosol](#). In most cells, for example, the Golgi apparatus is located close to the [nucleus](#), whereas the network of [ER](#) tubules extends from the nucleus throughout the entire cytosol. These characteristic distributions depend on interactions of the organelles with the [cytoskeleton](#). The localization of both the ER and the Golgi apparatus, for instance, depends on an intact [microtubule](#) array; if the microtubules are experimentally depolymerized with a drug, the Golgi apparatus fragments and disperses throughout the cell, and the ER network collapses toward the cell center (discussed in Chapter 16).

The Topological Relationships of Membrane-enclosed Organelles Can Be Interpreted in Terms of Their Evolutionary Origins

To understand the relationships between the compartments of the cell, it is helpful to consider how they might have evolved. The precursors of the first eucaryotic cells are thought to have been simple organisms that resembled bacteria, which generally have a [plasma membrane](#) but no internal membranes. The plasma membrane in such cells therefore provides all membrane-dependent functions, including the pumping of ions, ATP synthesis, [protein](#) secretion, and [lipid](#) synthesis. Typical present-day eucaryotic cells are 10–30 times larger in linear dimension and 1000–10,000 times greater in volume than a typical bacterium such as *E. coli*. The profusion of internal membranes can be seen in part as an [adaptation](#) to this increase in size: the eucaryotic cell has a much smaller ratio of surface area to volume, and its area of plasma membrane is presumably too small to sustain the many vital functions for which

membranes are required. The extensive [internal membrane](#) systems of a eucaryotic cell alleviate this imbalance.

The evolution of internal membranes evidently accompanied the specialization of [membrane](#) function. Consider, for example, the generation of [thylakoid vesicles](#) in chloroplasts. These vesicles form during the [development](#) of chloroplasts from *proplastids* in the green leaves of plants. Proplastids are small precursor organelles that are present in all immature plant cells. They are surrounded by a double membrane and develop according to the needs of the differentiated cells: they develop into chloroplasts in leaf cells, for example, and into organelles that store [starch](#), [fat](#), or pigments in other cell types ([Figure 12-3A](#)). In the process of differentiating into chloroplasts, specialized membrane patches form and pinch off from the [inner membrane](#) of the proplastid. The vesicles that pinch off form a new specialized [compartment](#), the *thylakoid*, that harbors all of the [chloroplast](#)'s photosynthetic machinery.

Other compartments in eucaryotic cells may have originated in a conceptually similar way. Pinching off of specialized intracellular [membrane](#) structures from the [plasma membrane](#), for example, would create organelles with an interior that is topologically equivalent to the exterior of the cell. We shall see that this topological relationship holds for all of the organelles involved in the secretory and endocytic pathways, including the [ER](#), Golgi apparatus, endosomes, and lysosomes. We can therefore think of all of these organelles as members of the same family. As we discuss in detail in the next chapter, their interiors communicate extensively with one another and with the outside of the cell via **transport vesicles** that bud off from one [organelle](#) and fuse with another.

As described in Chapter 14, mitochondria and plastids differ from the other [membrane](#)-enclosed organelles in containing their own genomes. The nature of these genomes, and the close resemblance of the proteins in these organelles to those in some present-day bacteria, strongly suggest that mitochondria and plastids evolved from bacteria that were engulfed by other cells with which they initially lived in [symbiosis](#) (discussed in Chapters 1 and 14). According to the hypothetical scheme shown in [Figure 12-4B](#), the [inner membrane](#) of mitochondria and plastids corresponds to the original [plasma membrane](#) of the bacterium, while the [lumen](#) of these organelles evolved from the bacterial [cytosol](#). As might be expected from such an endocytic origin, these two organelles are surrounded by a double membrane, and they remain isolated from the extensive vesicular traffic that connects the interiors of most of the other membrane-enclosed organelles to each other and to the outside of the cell.

The evolutionary scheme described above groups the intracellular compartments in eucaryotic cells into four distinct families: (1) the [nucleus](#) and the [cytosol](#), which communicate through *nuclear pore complexes* and are thus topologically continuous (although functionally distinct); (2) all organelles that function in the secretory and endocytic pathways—including the [ER](#), Golgi apparatus, endosomes, lysosomes, the numerous classes of **transport intermediates** such as **transport vesicles**, and possibly peroxisomes; (3) the mitochondria; and (4) the plastids (in plants only).

Proteins Can Move Between Compartments in Different Ways

All proteins begin being synthesized on ribosomes in the [cytosol](#), except for the few that are synthesized on the ribosomes of mitochondria and plastids. Their subsequent fate depends on their [amino acid](#) sequence, which can contain [sorting signals](#) that direct their delivery to locations outside the cytosol. Most proteins do not have a [sorting signal](#) and consequently remain in the cytosol as permanent residents. Many others, however, have specific **sorting** signals that direct their **transport** from the cytosol into the [nucleus](#), the [ER](#), mitochondria, plastids, or peroxisomes; **sorting** signals can also direct the **transport** of proteins from the ER to other destinations in the cell.

To understand the general principles by which **sorting** signals operate, it is important to distinguish three fundamentally different ways by which proteins move from one [compartment](#) to another. These three mechanisms are described below, and their sites of action in the cell are outlined.

1.

In **gated transport**, the **protein** traffic between the [cytosol](#) and [nucleus](#) occurs between topologically equivalent spaces, which are in continuity through the nuclear pore complexes. The nuclear pore complexes function as selective gates that actively **transport** specific macromolecules and macromolecular assemblies, although they also allow free [diffusion](#) of smaller molecules.

2.

In **transmembrane transport**, [membrane](#)-bound **protein** *translocators* directly **transport** specific proteins across a membrane from the [cytosol](#) into a space that is topologically distinct. The transported **protein molecule** usually must unfold to snake through the translocator. The initial **transport** of selected proteins from the cytosol into the [ER lumen](#) or from the cytosol into mitochondria, for example, occurs in this way.

3.

In **vesicular transport**, [membrane](#)-enclosed **transport** intermediates—which may be small, spherical **transport** vesicles or larger, irregularly shaped [organelle](#) fragments—ferry proteins from one [compartment](#) to another. The **transport** vesicles and fragments become loaded with a cargo of molecules derived from the [lumen](#) of one compartment as they pinch off from its membrane; they discharge their cargo into a second compartment by fusing with that compartment ([Figure 12-7](#)). The transfer of soluble proteins from the [ER](#) to the Golgi apparatus, for example, occurs in this way. Because the transported proteins do not cross a membrane, **vesicular transport** can move proteins only between compartments that are topologically equivalent.

Each of the three modes of **protein** transfer is usually guided by **sorting** signals in the transported **protein** that are recognized by [complementary receptor](#) proteins. If a large **protein** is to be imported into the [nucleus](#), for example, it must possess

a **sorting signal** that is recognized by receptor proteins that guide it through the **nuclear pore complex**. If a **protein** is to be transferred directly across a **membrane**, it must possess a **sorting** signal that is recognized by the translocator in the membrane to be crossed. Likewise, if a **protein** is to be loaded into a certain type of **vesicle** or retained in certain organelles, its **sorting** signal must be recognized by a complementary receptor in the appropriate membrane.

Signal Sequences and Signal Patches Direct Proteins to the Correct Cellular Address

There are at least two types of **sorting** signals in proteins. One type resides in a continuous stretch of **amino acid** sequence, typically 15–60 residues long. Some of these **signal sequences** are removed from the finished **protein** by specialized **signal peptidases** once the **sorting** process has been completed. The other type consists of a specific three-dimensional arrangement of atoms on the **protein's** surface that forms when the **protein** folds up. The amino acid residues that comprise this **signal patch** can be distant from one another in the linear amino acid sequence, and they generally persist in the finished **protein** (Figure 12-8). Signal sequences are used to direct proteins from the **cytosol** into the **ER**, mitochondria, chloroplasts, and peroxisomes, and they are also used to **transport** proteins from the **nucleus** to the cytosol and from the Golgi apparatus to the ER. The **sorting** signals that direct proteins into the nucleus from the cytosol can be either short signal sequences or longer sequences that are likely to fold into signal patches. Signal patches also direct newly synthesized degradative enzymes into lysosomes.

Each **signal sequence** specifies a particular destination in the cell. Proteins destined for initial transfer to the **ER** usually have a signal sequence at their **N terminus**, which characteristically includes a sequence composed of about 5–10 hydrophobic amino acids. Many of these proteins will in turn pass from the ER to the Golgi apparatus, but those with a specific sequence of four amino acids at their **C terminus** are recognized as ER residents and are returned to the ER. Proteins destined for mitochondria have signal sequences of yet another type, in which positively charged amino acids alternate with hydrophobic ones. Finally, many proteins destined for peroxisomes have a signal peptide of three characteristic amino acids at their C terminus.

Some specific signal sequences are presented in. The importance of each of these signal sequences for **protein** targeting has been shown by experiments in which the peptide is transferred from one **protein** to another by genetic engineering techniques. Placing the N-terminal **ER signal sequence** at the beginning of a cytosolic **protein**, for example, redirects the **protein** to the ER. Signal sequences are therefore both necessary and sufficient for **protein** targeting. Even though their **amino acid** sequences can vary greatly, the signal sequences of all proteins having the same destination are functionally interchangeable, and physical properties, such as hydrophobicity, often seem to be more important in the signal-recognition process than the exact amino acid sequence.

Signal patches are far more difficult to analyze than signal sequences, so less is known about their structure. Because they often result from a [complex](#) three-dimensional **protein**-folding pattern, they cannot be easily transferred experimentally from one **protein** to another.

Both types of **sorting** signals are recognized by [complementary sorting](#) receptors that guide proteins to their appropriate destination, where the receptors unload their cargo. The receptors function catalytically: after completing one round of targeting, they return to their point of origin to be reused. Most **sorting** receptors recognize classes of proteins rather than just an individual **protein** species. They therefore can be viewed as public transportation systems dedicated to delivering groups of components to their correct location in the cell.

Most Membrane-enclosed Organelles Cannot Be Constructed From Scratch: They Require Information in the Organelle Itself

When a cell reproduces by division, it has to duplicate its [membrane](#)-enclosed organelles. In general, cells do this by enlarging the existing organelles by incorporating new molecules into them; the enlarged organelles then divide and are distributed to the two daughter cells. Thus, each daughter cell inherits from its mother a complete set of specialized cell membranes. This inheritance is essential because a cell could not make such membranes from scratch. If the [ER](#) were completely removed from a cell, for example, how could the cell reconstruct it? As we shall discuss later, the membrane proteins that define the ER and perform many of its functions are themselves products of the ER. A new ER could not be made without an existing ER or, at the very least, a membrane that specifically contains the **protein** translocators required to import selected proteins into the ER from the [cytosol](#) (including the ER-specific translocators themselves). The same is true for mitochondria, plastids, and peroxisomes.

Thus, it seems that the information required to construct a [membrane](#)-enclosed [organelle](#) does not reside exclusively in the [DNA](#) that specifies the organelle's proteins. *Epigenetic* information in the form of at least one distinct **protein** that preexists in the organelle membrane is also required, and this information is passed from parent cell to progeny cell in the form of the organelle itself. Presumably, such information is essential for the propagation of the cell's compartmental organization, just as the information in DNA is essential for the propagation of the cell's [nucleotide](#) and [amino acid](#) sequences.

As we discuss in more detail in Chapter 13, however, the [ER](#) sheds a constant stream of [membrane](#) vesicles that incorporate only specific proteins and therefore have a different composition from the ER itself. Similarly, the [plasma membrane](#) constantly produces specialized endocytic vesicles. Thus, some membrane-enclosed compartments can form from other organelles and do not have to be inherited at [cell division](#).

Summary

Eucaryotic cells contain intracellular membranes that enclose nearly half the cell's total volume in separate intracellular compartments called organelles. The main types

of [membrane](#)-enclosed organelles present in all eucaryotic cells are the [endoplasmic reticulum](#), Golgi apparatus, [nucleus](#), mitochondria, lysosomes, endosomes, and peroxisomes; plant cells also contain plastids, such as chloroplasts. Each [organelle](#) contains a distinct set of proteins that mediate its unique functions.

Each newly synthesized [organelle protein](#) must find its way from a [ribosome](#) in the [cytosol](#), where it is made, to the organelle where it functions. It does so by following a specific pathway, guided by signals in its [amino acid](#) sequence that function as signal sequences or signal patches. Signal sequences and patches are recognized by [complementary sorting](#) receptors that deliver the **protein** to the appropriate target organelle. Proteins that function in the cytosol do not contain **sorting** signals and therefore remain there after they are synthesized.

During [cell division](#), organelles such as the [ER](#) and mitochondria are distributed intact to each daughter cell. These organelles contain information that is required for their construction so that they cannot be made from scratch.

The Transport of Molecules between the Nucleus and the Cytosol

The [nuclear envelope](#) encloses the [DNA](#) and defines the *nuclear compartment*. This envelope consists of two concentric membranes that are penetrated by nuclear pore complexes. Although the inner and outer nuclear membranes are continuous, they maintain distinct [protein](#) compositions. The [inner nuclear membrane](#) contains specific proteins that act as binding sites for [chromatin](#) and for the **protein** meshwork of the [nuclear lamina](#) that provides structural support for this [membrane](#). The [inner membrane](#) is surrounded by the [outer nuclear membrane](#), which is continuous with the membrane of the [ER](#). Like the membrane of the ER that will be described later in this chapter, the [outer nuclear membrane](#) is studded with ribosomes engaged in **protein** synthesis. The proteins made on these ribosomes are transported into the space between the inner and outer nuclear membranes (the *perinuclear space*), which is continuous with the [ER lumen](#).

Bidirectional traffic occurs continuously between the [cytosol](#) and the [nucleus](#). The many proteins that function in the nucleus—including histones, [DNA](#) and [RNA](#) polymerases, [gene](#) regulatory proteins, and RNA-processing proteins—are selectively imported into the nuclear [compartment](#) from the cytosol, where they are made. At the same time, tRNAs and mRNAs are synthesized in the nuclear compartment and then exported to the cytosol. Like the import process, the export process is selective; mRNAs, for example, are exported only after they have been properly modified by RNA-processing reactions in the nucleus. In some cases the **transport** process is [complex](#): ribosomal proteins, for instance, are made in the cytosol, imported into the nucleus—where they assemble with newly made ribosomal RNA into particles—and are then exported again to the cytosol as part of a ribosomal [subunit](#). Each of these steps requires selective **transport** across the [nuclear envelope](#).

Nuclear Pore Complexes Perforate the Nuclear Envelope

The [nuclear envelope](#) of all eucaryotes is perforated by large, elaborate structures known as **nuclear pore complexes**. In animal cells, each [complex](#) has an estimated molecular mass of about 125 million and is thought to be composed of more than 50 different proteins, called [nucleoporins](#), that are arranged with a striking octagonal symmetry.

In general, the more active the [nucleus](#) is in transcription, the greater the number of pore complexes its envelope contains. The [nuclear envelope](#) of a typical mammalian cell contains 3000–4000 pore complexes. If the cell is synthesizing [DNA](#), it needs to import about 10^6 [histone](#) molecules from the [cytosol](#) every 3 minutes to package the newly made DNA into [chromatin](#), which means that, on average, each pore [complex](#) needs to **transport** about 100 histone molecules per minute. If the cell is growing rapidly, each complex also needs to **transport** about 6 newly assembled large and small ribosomal subunits per minute from the nucleus, where they are produced, to the cytosol, where they are used. And that is only a very small part of the total traffic that passes through the pore complexes.

Each pore [complex](#) contains one or more open [aqueous](#) channels through which small water-soluble molecules can passively diffuse. The effective size of these channels has been [determined](#) by injecting labeled water-soluble molecules of different sizes into the [cytosol](#) and then measuring their rate of [diffusion](#) into the [nucleus](#). Small molecules (5000 daltons or less) diffuse in so fast that the [nuclear envelope](#) can be considered to be freely permeable to them. A [protein](#) of 17,000 daltons takes 2 minutes to equilibrate between the cytosol and the nucleus, whereas proteins larger than 60,000 daltons are hardly able to enter the nucleus at all. A quantitative analysis of such data suggests that the [nuclear pore complex](#) contains a pathway for free diffusion equivalent to a water-filled cylindrical channel about 9 [nm](#) in diameter and 15 nm long; such a channel would occupy only a small fraction of the total volume of the pore complex.

Because many cell proteins are too large to pass by [diffusion](#) through the nuclear pore complexes, the [nuclear envelope](#) enables the nuclear [compartment](#) and the [cytosol](#) to maintain different complements of proteins. Mature cytosolic ribosomes, for example, are about 30 [nm](#) in diameter and thus cannot diffuse through the 9 nm channels; their exclusion from the [nucleus](#) ensures that [protein](#) synthesis is confined to the cytosol. But how does the nucleus export newly made ribosomal subunits or import large molecules, such as [DNA](#) and [RNA](#) polymerases, which have [subunit](#) molecular weights of 100,000–200,000 daltons? As we discuss next, these and many other **protein** and RNA molecules bind to specific [receptor](#) proteins that ferry them actively through nuclear pore complexes.

Nuclear Localization Signals Direct Nuclear Proteins to the Nucleus

When proteins are experimentally extracted from the [nucleus](#) and reintroduced into the [cytosol](#) (e.g., through experimentally induced perforations in the [plasma membrane](#)), even the very large ones reaccumulate efficiently in the nucleus. The selectivity of this nuclear import process resides in **nuclear localization signals (NLSs)**, which are present only in nuclear proteins. The signals have been precisely defined in numerous nuclear proteins by using [recombinant DNA](#) technology. As mentioned earlier, they can be either signal sequences or signal patches. In many nuclear proteins they consist of one or two short sequences that are rich in the positively charged amino acids lysine and, the precise sequence varying for different nuclear proteins. Other nuclear proteins contain different signals, some of which are not yet characterized.

The signals characterized this far can be located almost anywhere in the [amino acid](#) sequence and are thought to form loops or patches on the [protein](#) surface. Many function even when linked as short peptides to lysine side chains on the surface of a cytosolic **protein**, suggesting that the precise location of the signal within the amino acid sequence of a nuclear **protein** is not important.

The **transport** of nuclear proteins through nuclear pore complexes can be directly visualized by coating gold particles with a [nuclear localization signal](#), injecting the particles into the [cytosol](#), and then following their fate by [electron](#) microscopy. Studies with various sizes of gold beads indicate that the opening can dilate up to about 26nm in diameter during the **transport** process. A structure in the center of the [nuclear pore complex](#) seems to function like a close-fitting diaphragm that opens just the right amount to let **transport** substrates pass. The molecular basis of the gating mechanism remains a mystery.

The mechanism of macromolecular **transport** across nuclear pore complexes is fundamentally different from the **transport** mechanisms involved in [protein](#) transfer across the membranes of other organelles, because it occurs through a large [aqueous](#) pore rather than through a **protein** transporter spanning one or more [lipid](#) bilayers. For this reason, nuclear proteins can be transported through a pore [complex](#) while they are in a fully folded [conformation](#). Likewise, a newly formed ribosomal [subunit](#) is transported out of the [nucleus](#) as an assembled particle. By contrast, proteins have to be extensively unfolded during their **transport** into most other organelles, as we discuss later. In the [electron microscope](#), however, very large particles traversing the pore seem to become constricted as they squeeze through the [nuclear pore complex](#), indicating that at least some of them must undergo restructuring during **transport**.

Nuclear Import Receptors Bind Nuclear Localization Signals and Nucleoporins

To initiate nuclear import, most nuclear localization signals must be recognized by **nuclear import receptors**, which are encoded by a family of related genes. Each family member encodes a [receptor protein](#) that is specialized for the **transport** of a group of nuclear proteins sharing structurally similar nuclear localization signals.

The import receptors are soluble cytosolic proteins that bind both to the [nuclear localization signal](#) on the [protein](#) to be transported and to nucleoporins, some of which form the tentaclelike fibrils that extend into the [cytosol](#) from the rim of the nuclear pore complexes. The fibrils and many other nucleoporins contain a large number of short amino-[acid](#) repeats that contain phenylalanine and glycine and are therefore called *FG-repeats* (named after the one-letter code for amino acids, discussed in Chapter 5). FG-repeats serve as binding sites for the import receptors. They are thought to line the path through the nuclear pore complexes taken by the import receptors and their bound cargo proteins. These **protein** complexes move along the path by repeatedly binding, dissociating, and then re-binding to adjacent repeat sequences. Once in the [nucleus](#), the import receptors dissociate from their cargo and are returned to the cytosol.

Nuclear import receptors do not always bind to nuclear proteins directly. Additional adaptor proteins are sometimes used that bridge between the import receptors and the nuclear localization signals on the proteins to be transported. Surprisingly, the adaptor proteins are structurally related to nuclear import receptors, suggesting a common evolutionary origin. The combined use of import receptors and adaptors allows a cell to recognize the broad repertoire of nuclear localization signals that are displayed on nuclear proteins.

Nuclear Export Works Like Nuclear Import, But in Reverse

The nuclear export of large molecules, such as new ribosomal subunits and [RNA](#) molecules, also occurs through nuclear pore complexes and depends on a selective **transport** system. The **transport** system relies on [nuclear export signals](#) on the macromolecules to be exported, as well as on [complementary nuclear export receptors](#). These receptors bind both the export signal and nucleoporins to guide their cargo through the pore [complex](#) to the [cytosol](#).

Nuclear export receptors are structurally related to nuclear import receptors, and they are encoded by the same [gene](#) family of [nuclear transport receptors](#), or *karyopherins*. In [yeast](#), there are 14 genes encoding members of this family; in animal cells the number is significantly larger. From their [amino acid](#) sequence alone, it is often not possible to distinguish whether a particular family member works as a nuclear import or nuclear export [receptor](#). It comes as no surprise, therefore, that the import and export **transport** systems work in similar ways but in opposite directions: the import receptors bind their cargo molecules in the [cytosol](#), release them in the [nucleus](#), and are then exported to the cytosol for reuse, while the export receptors function in reverse.

If gold spheres similar to those used in the experiments shown in are coated with small [RNA](#) molecules ([tRNA](#) or ribosomal 5S RNA) and injected into the [nucleus](#) of a cultured cell, they are rapidly transported through the nuclear pore complexes into the [cytosol](#). Using two sizes of gold particles, one coated with RNA and injected into the nucleus and the other coated with nuclear localization signals and injected into the cytosol, it can be shown that a single pore [complex](#) conducts traffic in both directions. How a pore

complex coordinates the bidirectional flow of macromolecules to avoid congestion and head-on collisions is not known.

The Ran GTPase Drives Directional Transport Through Nuclear Pore Complexes

The import of nuclear proteins through the pore [complex](#) concentrates specific proteins in the [nucleus](#), thereby increasing order in the cell, which must consume energy (discussed in Chapter 2). The energy is thought to be provided by the hydrolysis of GTP by the monomeric [GTPase Ran](#). [Ran](#) is found in both the [cytosol](#) and the nucleus, and it is required for both the nuclear import and export systems.

Like other GTPases, [Ran](#) is a molecular switch that can exist in two conformational states, depending on whether GDP or GTP is bound (discussed in Chapter 3). Conversion between the two states is triggered by two Ran-specific regulatory proteins: a cytosolic [GTPase-activating protein](#) (*GAP*) that triggers GTP hydrolysis and thus converts Ran-GTP to Ran-GDP, and a nuclear *guanine exchange factor* (*GEF*) that promotes the exchange of GDP for GTP and thus converts Ran-GDP to Ran-GTP. Because *Ran-GAP* is located in the [cytosol](#) and *Ran-GEF* is located in the [nucleus](#), the cytosol primarily contains Ran-GDP, and the nucleus primarily contains Ran-GTP.

This gradient of the two conformational forms of [Ran](#) drives [nuclear transport](#) in the appropriate direction. Docking of nuclear import receptors to FG-repeats on the cytosolic side of the [nuclear pore complex](#), for example, occurs only when these receptors are loaded with an appropriate cargo. The import receptors with their bound cargo then move along tracks lined by FG-repeat sequences until they reach the nuclear side of the pore complex, where Ran-GTP binding causes the import receptors to release their cargo. By favoring cargo-dependent loading of import receptors onto the FG-repeat track in the [cytosol](#) and Ran-GTP-dependent cargo release in the [nucleus](#), the nuclear localization of Ran-GTP imposes directionality.

Having discharged its cargo in the [nucleus](#), the empty import [receptor](#) with [Ran](#)-GTP bound is transported back through the pore [complex](#) to the [cytosol](#). There, two cytosolic proteins, *Ran Binding Protein* and [Ran-GAP](#) collaborate to convert Ran-GTP to Ran-GDP. The *Ran Binding Protein* first displaces Ran-GTP from the import receptor, which allows [Ran-GAP](#) to trigger [Ran](#) to hydrolyze its bound GTP. The Ran-GDP then dissociates from the *Ran Binding Protein* and is reimported into the nucleus, thereby completing the cycle.

Nuclear export occurs by a similar mechanism, except that [Ran](#)-GTP in the [nucleus](#) promotes cargo binding to the export [receptor](#) and the binding of the loaded receptor to the nuclear side of the pore [complex](#). Once in the [cytosol](#), [Ran](#) encounters [Ran-GAP](#) and *Ran Binding Protein* and hydrolyses its bound GTP. The export receptor then releases both its cargo and Ran-GDP in the cytosol and dissociates from the pore complex, and free export receptors are returned to the nucleus to complete the cycle.

Transport Between the Nucleus and Cytosol Can Be Regulated by Controlling Access to the Transport Machinery

Some proteins, such as those that bind newly made mRNAs in the [nucleus](#), contain both nuclear localization and nuclear export signals. These proteins continually shuttle between the nucleus and the [cytosol](#). The steady-state localization of such *shuttling proteins* is [determined](#) by the relative rates of their import and export. If the rate of import exceeds the rate of export, a [protein](#) will be located primarily in the nucleus. Conversely, if the rate of export exceeds the rate of import, a **protein** will be located primarily in the cytosol. Thus, changing the rate of import, export, or both, can change the location of a **protein**.

Some shuttling proteins move continuously in and out of the [nucleus](#). In other cases, however, the **transport** is stringently controlled. As discussed in Chapter 7, the activity of some [gene](#) regulatory proteins is controlled by keeping them out of the nuclear [compartment](#) until they are needed there. In many cases, this control depends on the regulation of nuclear localization and export signals; these can be turned on or off, often by [phosphorylation](#) of adjacent amino acids.

Other [gene](#) regulatory proteins are bound to inhibitory cytosolic proteins that either anchor them in the [cytosol](#) (through interactions with the [cytoskeleton](#) or with specific organelles), or mask their nuclear localization signals so that they are unable to interact with nuclear import receptors. When the cell receives an appropriate stimulus, the [gene regulatory protein](#) is released from its cytosolic anchor or mask and is transported into the [nucleus](#). One important example is the latent gene regulatory **protein** that controls the [expression](#) of proteins involved in [cholesterol metabolism](#). The **protein** is made and stored in an inactive form as a [transmembrane protein](#) in the [ER](#). When deprived of cholesterol, the cell activates specific proteases that cleave the **protein**, releasing its cytosolic [domain](#). This domain is then imported into the nucleus, where it activates the transcription of genes required for cholesterol import and synthesis.

Cells control the export of [RNA](#) from the [nucleus](#) in a similar way. Messenger RNAs become bound to proteins that are loaded onto the RNA as transcription and splicing proceed. These proteins contain nuclear export signals that are recognized by export receptors that guide the RNA out of the nucleus through nuclear pore complexes. Upon entry into the [cytosol](#), the proteins coating the RNA are stripped off and rapidly returned to the nucleus. Other RNAs, such as snRNAs and tRNAs, are exported by different sets of nuclear export receptors.

Incompletely processed pre-mRNAs are actively retained in the [nucleus](#), anchored to the nuclear transcription and splicing machinery, which releases an [RNA molecule](#) only after its processing is completed. Genetic studies in [yeast](#) show that a [mutant](#) pre-mRNA that cannot properly engage with the splicing machinery is improperly exported as an unspliced molecule.

The Nuclear Envelope Is Disassembled During Mitosis

The [nuclear lamina](#) is a meshwork of interconnected [protein](#) subunits called [nuclear lamins](#). The lamins are a special class of [intermediate filament](#) proteins (discussed in Chapter 16) that polymerize into a two-dimensional lattice ([Figure 12-20](#)). The [nuclear lamina](#) gives shape and stability to the [nuclear envelope](#), to which it is anchored by attachment to both the nuclear pore complexes and integral [membrane](#) proteins of the [inner nuclear membrane](#). The lamina also interacts directly with [chromatin](#), which itself interacts with the integral membrane proteins of the inner nuclear membrane. Together with the lamina, these membrane proteins provide structural links between the [DNA](#) and the nuclear envelope.

When a [nucleus](#) disassembles during [mitosis](#), the [nuclear lamina](#) depolymerizes. The disassembly is at least partly a consequence of direct [phosphorylation](#) of the nuclear lamins by the [cyclin](#)-dependent kinase activated at the onset of mitosis (discussed in Chapter 17). At the same time, proteins of the [inner nuclear membrane](#) are phosphorylated, and the nuclear pore complexes disassemble and disperse in the [cytosol](#). Nuclear envelope membrane proteins—no longer tethered to the pore complexes, lamina, or [chromatin](#)—diffuse throughout the [ER](#) membrane. Together, these events break down the barriers that normally separate the nucleus and cytosol, and these nuclear proteins that are not bound to membranes or chromosomes intermix completely with the cytosol of the dividing cell.

Later in [mitosis](#) (in late [anaphase](#)), the [nuclear envelope](#) reassembles on the surface of the chromosomes, as [inner nuclear membrane](#) proteins and dephosphorylated lamins rebind to [chromatin](#). [ER](#) membranes wrap around groups of chromosomes and continue fusing until a sealed nuclear envelope is reformed. During this process, the nuclear pore complexes also reassemble and start actively reimporting proteins that contain nuclear localization signals. Because the nuclear envelope is initially closely applied to the surface of the chromosomes, the newly formed [nucleus](#) excludes all proteins except those initially bound to the mitotic chromosomes and those that are selectively imported through nuclear pore complexes. In this way, all other large proteins are kept out of the newly assembled nucleus.

Nuclear localization signals are not cleaved off after **transport** into the [nucleus](#). This is presumably because nuclear proteins need to be imported repeatedly, once after every [cell division](#). In contrast, once a [protein molecule](#) has been imported into any of the other [membrane](#)-enclosed organelles, it is passed on from generation to generation within that [compartment](#) and need never be translocated again; the [signal sequence](#) on these molecules is often removed after **protein translocation**.

Summary

The [nuclear envelope](#) consists of an inner and an [outer nuclear membrane](#). The [outer membrane](#) is continuous with the [ER](#) membrane, and the space between it and the [inner membrane](#) is continuous with the [ER lumen](#). [RNA](#) molecules, which are made in the [nucleus](#), and ribosomal subunits, which are assembled there, are exported to the [cytosol](#), while all the proteins that function in the nucleus are synthesized in the cytosol and are then imported. The

extensive traffic of materials between the nucleus and cytosol occurs through nuclear pore complexes, which provide a direct passageway across the nuclear envelope.

Proteins containing nuclear localization signals are actively transported inward through the nuclear pore complexes, while [RNA](#) molecules and newly made ribosomal subunits contain nuclear export signals that direct their [active transport](#) outward through the pore complexes. Some proteins, including nuclear import and export receptors, continually shuttle between the [cytosol](#) and [nucleus](#). The [GTPase Ran](#), provides directionality for [nuclear transport](#). The **transport** of nuclear proteins and RNA molecules through the pore complexes can be regulated by denying these molecules access to the **transport** machinery. Because nuclear localization signals are not removed, nuclear proteins can be imported repeatedly, as is required each time that the nucleus reassembles after [mitosis](#).

The Transport of Proteins into Mitochondria and Chloroplasts

As discussed in Chapter 14, mitochondria and chloroplasts are double-[membrane](#)-enclosed organelles. They specialize in the synthesis of ATP, using energy derived from [electron transport](#) and [oxidative phosphorylation](#) in mitochondria and from [photosynthesis](#) in chloroplasts. Although both organelles contain their own [DNA](#), ribosomes, and other components required for [protein](#) synthesis, most of their proteins are encoded in the cell [nucleus](#) and imported from the [cytosol](#). Moreover, each imported **protein** must reach the particular [organelle](#) subcompartment in which it functions.

There are two subcompartments in mitochondria: the internal [matrix space](#) and the [intermembrane space](#). These compartments are formed by the two concentric mitochondrial membranes: the [inner membrane](#), which forms extensive invaginations, the *cristae*, and encloses the [matrix space](#), and the [outer membrane](#), which is in contact with the [cytosol](#) (Figure 12-22A). Chloroplasts have the same two subcompartments plus an additional subcompartment, the [thylakoid space](#), which is surrounded by the [thylakoid membrane](#). Each of the subcompartments in mitochondria and chloroplasts contains a distinct set of proteins.

New mitochondria and chloroplasts are produced by the growth of preexisting organelles followed by fission (discussed in Chapter 14). Their growth depends mainly on the import of proteins from the [cytosol](#). This requires that proteins be translocated across a number of membranes in succession and end up in the appropriate place. How this occurs is the subject of this [section](#).

Translocation into the Mitochondrial Matrix Depends on a Signal Sequence and Protein Translocators

Proteins imported into the matrix of **mitochondria** are usually taken up from the cytosol within seconds or minutes of their release from ribosomes. Thus, in contrast to the protein translocation into the ER described later, mitochondrial proteins are first fully synthesized as precursor proteins in the cytosol and then translocated into mitochondria by aposttranslational mechanism. Most of the mitochondrial precursor proteins have a signal sequence at their N terminus that is rapidly removed after import by a protease (the signal peptidase) in the mitochondrial matrix. The signal sequences are both necessary and sufficient for import of the proteins that contain them: through the use of genetic engineering techniques, these signals can be linked to any cytosolic **protein** to direct the **protein** into the mitochondrial matrix. Sequence comparisons and physical studies of different matrix signal sequences suggest that their common feature is the propensity to fold into an amphipathic α helix, in which positively charged residues are clustered on one side of the helix, while uncharged hydrophobic residues are clustered on the opposite side. This configuration—rather than a precise amino acid sequence—is recognized by specific receptor proteins that initiate **protein** translocation.

Protein translocation across mitochondrial membranes is mediated by multi-subunit **protein** complexes that function as protein translocators: the TOM complex functions across the outer membrane, and two TIM complexes, the TIM23 and TIM22 complexes, function across the inner membrane (Figure 12-24). TOM and TIM stand for translocase of the outer and inner mitochondrial membranes, respectively. These complexes contain some components that act as receptors for mitochondrial precursor proteins and other components that form the translocation channel. The TOM complex is required for the import of all nucleus-encoded mitochondrial proteins. It initially transports their signal sequences into the intermembrane space and helps to insert transmembrane proteins into the outer membrane. The TIM23 complex then transports some of these proteins into the matrix space, while helping to insert transmembrane proteins into the inner membrane. The TIM22 complex mediates the insertion of a subclass of inner membrane proteins, including the carrier protein that transports ADP, ATP, and phosphate. A third protein translocator in the inner mitochondrial membrane, the **OXA complex**, mediates the insertion of inner membrane proteins that are synthesized within the mitochondria. It also helps to insert some proteins that are initially transported into the matrix by the TOM and TIM complexes.

Mitochondrial Precursor Proteins Are Imported as Unfolded Polypeptide Chains

Almost everything we know about the molecular mechanism of protein import into mitochondria has been learned from analyses of cell-free, reconstituted **transport** systems. Mitochondria are first purified by differential centrifugation of homogenized cells and are then incubated with radiolabeled mitochondrial precursor proteins, which are generally taken up rapidly and efficiently. By changing the incubation conditions, it is possible to establish the biochemical requirements for the **transport**.

Mitochondrial precursor proteins do not fold into their native structures after they are synthesized; instead, they remain unfolded through interactions with other proteins in the [cytosol](#). Some of these interacting proteins are general *chaperone proteins* belonging to the *hsp70 family* (discussed in Chapter 6), whereas others are dedicated to mitochondrial precursor proteins and bind directly to their signal sequences. All these interacting proteins help to prevent the precursor proteins from aggregating or folding up spontaneously before they engage with the [TOM complex](#) in the outer mitochondrial [membrane](#). As a first step in the import process, the mitochondrial precursor proteins bind to import [receptor](#) proteins of the TOM complex, which recognize the mitochondrial signal sequences. The interacting proteins are then stripped off, and the unfolded [polypeptide](#) chain is fed—[signal sequence](#) first—into the [translocation](#) channel.

Mitochondrial Precursor Proteins Are Imported into the Matrix at Contact Sites That Join the Inner and Outer Membranes

In principle, a [protein](#) could reach the mitochondrial matrix by crossing the two membranes one at a time, or it could pass through both at once. To distinguish between these possibilities, a cell-free mitochondrial import system was cooled to a low temperature, arresting the proteins at an intermediate step in the [translocation](#) process. The proteins that accumulated at this step already had their N-terminal [signal sequence](#) removed by the matrix [signal peptidase](#), indicating that their [N terminus](#) must be in the [matrix space](#). Yet, the bulk of the **protein** could still be attacked from outside the mitochondria by externally added proteolytic enzymes ([Figure 12-25](#)). This result demonstrates that the precursor proteins can pass through both mitochondrial membranes at once to enter the matrix. It is thought that the [TOM complex](#) first transports the mitochondrial targeting signal across the [outer membrane](#). Once it reaches in the [intermembrane space](#), the targeting signal binds to a TIM complex, opening the channel in the complex through which the [polypeptide](#) chain either enters the matrix or inserts into the [inner membrane](#). Electron microscopists have noted numerous *contact sites* at which the inner and outer mitochondrial membranes are closely apposed, and it seems likely that translocation occurs at or near these sites.

Although the functions of the TOM and [TIM complexes](#) are usually coupled to allow **protein transport** across both membranes at the same time, both **protein** types of translocator can work independently. The [TOM complex](#) in isolated outer membranes, for example, can translocate the [signal sequence](#) of precursor proteins across the [membrane](#). Similarly, mitochondria with experimentally disrupted outer membranes, and therefore with the TIM23 complex exposed on their surface, efficiently import precursor proteins into the [matrix space](#). Despite the independent functional roles of the TOM and TIM translocators, the two mitochondrial membranes at contact sites may be permanently held together by the TIM23 complex, which spans both membranes.

ATP Hydrolysis and a H⁺ Gradient are Used to Drive Protein Import into Mitochondria

Directional **transport** requires energy. In most biological systems, energy is supplied by ATP hydrolysis. Mitochondrial **protein** import is fueled by ATP hydrolysis at two discrete sites, one outside the mitochondria and one in the matrix ([Figure 12-27](#)). In addition, another energy source is required: an electrochemical H^+ gradient across the inner mitochondrial [membrane](#).

The first requirement for energy occurs at the initial stage of the [translocation](#) process, when the unfolded precursor **protein**, associated with chaperone proteins, interacts with the mitochondrial import receptors. As discussed in Chapter 6, the release of newly synthesized polypeptides from the hsp70 family of chaperone proteins requires ATP hydrolysis. Experimentally, the requirement for hsp70 and ATP in the [cytosol](#) can be bypassed if the precursor **protein** is artificially unfolded prior to adding it to purified mitochondria.

Once the [signal sequence](#) has passed through the [TOM complex](#) and has become bound to either TIM complex, further [translocation](#) through the TIM requires an electrochemical H^+ gradient across the [inner membrane](#). The [electrochemical gradient](#) is maintained by the pumping of H^+ from the matrix to the [intermembrane space](#), driven by [electron transport](#) processes in the inner membrane. By contrast, the outer mitochondrial membrane, like that of Gram-negative bacteria, contains a pore-forming **protein** called porin and is thus freely permeable to inorganic ions and metabolites (but not to most proteins), so that [ion](#) gradients cannot be maintained across it. The energy in the electrochemical H^+ gradient across the inner membrane is not only used to help drive most of the cell's ATP synthesis; it is also used to drive the translocation of the targeting signals through the [TIM complexes](#). The precise mechanism by which this occurs is not known, but it is possible that the electrical components of the gradient helps to drive the positively charged signal sequence into the matrix by electrophoresis.

Hsp70 chaperone proteins in the [matrix space](#) also have a role in the [translocation](#) process, and they are the third point in the import process at which ATP is consumed, as we discuss next.

Repeated Cycles of ATP Hydrolysis by Mitochondrial Hsp70 Complete the Import Process

We know that **mitochondrial hsp70** is crucial to the import process, because mitochondria containing [mutant](#) forms of the **protein** fail to import precursor proteins. Like its cytosolic cousin, mitochondrial hsp70 has a high affinity for unfolded [polypeptide](#) chains and it binds tightly to an imported **protein** as soon as it emerges from the translocator in the matrix. The hsp70 then releases the **protein** in an ATP-dependent step. This energy-driven cycle of binding and subsequent release is thought to provide the final driving force needed to complete **protein** import after a **protein** has initially inserted into the TIM23 [complex](#).

Two models have been proposed to explain how ATP hydrolysis by mitochondrial hsp70 drives **protein** import. In both models, hsp70 proteins are closely associated with the TIM23 [complex](#), which deposits them onto the translocating [polypeptide](#) chain as it emerges

into the matrix. In the *thermal ratchet model*, the emerging chain slides back and forth in the TIM23 [translocation](#) channel by thermal motion. Each time a sufficiently long portion of the chain is exposed in the matrix, an hsp70 [molecule](#) binds to it, preventing further backsliding and thereby making the movement directional. Thus, a hand-over-hand binding of multiple hsp70 proteins translocates the polypeptide chain into the matrix. In the *cross-bridge ratchet model* ([Figure 12-28B](#)), the hsp70 proteins that bind to the emerging polypeptide chain undergo a conformational change, driven by ATP hydrolysis, that actively pulls a segment of the polypeptide chain into the matrix. A new hsp70 molecule can then bind to the segment just pulled in and repeat the cycle. In both models, therefore, hsp70 functions as a ratchet that prevents backsliding of the emerging polypeptide chain.

After the initial interaction with mitochondrial hsp70, many imported proteins are passed on to another chaperone [protein](#), *mitochondrial hsp60*. As discussed in Chapter 6, hsp60 provides a chamber for the unfolded [polypeptide](#) chain that facilitates its folding by binding and releasing it through cycles of ATP hydrolysis.

Protein Transport into the Inner Mitochondrial Membrane and the Intermembrane Space Requires Two Signal Sequences

Proteins that are integrated into the inner mitochondrial [membrane](#) or that operate in the [intermembrane space](#) are initially transported from the [cytosol](#) by the same mechanism that transports proteins into the matrix. In some cases they are first transferred into the matrix (see [Figure 12-26](#)). A hydrophobic [amino acid](#) sequence, however, is strategically placed after the N-terminal [signal sequence](#) that guides import into the matrix. Once the N-terminal signal sequence has been removed by the matrix [signal peptidase](#), the hydrophobic sequence functions as a new N-terminal signal sequence to translocate the [protein](#) from the matrix into or across the [inner membrane](#), using the OXA [complex](#) as the translocator. The OXA complex is also used to insert proteins encoded in the mitochondrion into the inner membrane. Closely related translocators are found in the plasma membranes of bacteria and in the thylakoids of chloroplasts, where they are thought to help to insert membrane proteins by a similar mechanism.

An alternative route to the [inner membrane](#) avoids excursion into the matrix altogether. In this case, the TIM23 translocator in the inner membrane binds to the hydrophobic sequence that follows the N-terminal [signal sequence](#) and initiates import, causing it to act as a *stop-transfer sequence* that prevents further [translocation](#) across the inner membrane. After the N-terminal signal sequence has been cleaved off, the remainder of the [protein](#) is pulled through the [TOM complex](#) in the [outer membrane](#) into the [intermembrane space](#). Different proteins use one or the other of these two pathways to the inner membrane or intermembrane space.

Proteins destined for the [intermembrane space](#) are first inserted via their hydrophobic [signal sequence](#) into the [inner membrane](#), and then cleaved by a [signal peptidase](#) in the intermembrane space to release the mature [polypeptide](#) chain as a soluble [protein](#). Many of these proteins attach as peripheral membrane proteins to the outer surface of the inner

membrane, where they form subunits of **protein** complexes that also contain transmembrane proteins.

Mitochondria are the principal site of ATP synthesis in the cell, but they also contain many metabolic enzymes, such as those of the citric [acid](#) cycle. Thus, in addition to proteins, mitochondria must also **transport** small metabolites across their membranes. While the [outer membrane](#) contains porins that make the membrane freely permeable to small molecules, the [inner membrane](#) does not. Instead, the **transport** of a vast number of small molecules across the inner membrane is mediated by a family of metabolite-specific carrier proteins. In [yeast](#) cells, these proteins comprise a family of 35 different proteins, of which the most abundant are those that **transport** ADP and ATP, or phosphate. These carrier proteins in the inner membrane are multipass transmembrane proteins, which do not have cleavable signal sequences at their N termini but, instead, contain internal signal sequences. These proteins cross the [TOM complex](#) in the outer membrane and are inserted into the inner membrane by the TIM22 complex. Their integration into the inner membrane requires the electrochemical H^+ gradient, but not mitochondrial hsp70 or ATP. The energetically favorable partitioning of the hydrophobic transmembrane regions into the inner membrane is likely to help drive integration.

Two Signal Sequences Are Required to Direct Proteins to the Thylakoid Membrane in Chloroplasts

Protein transport into [chloroplasts](#) resembles **transport** into mitochondria in many respects. Both processes occur posttranslationally, use separate [translocation](#) complexes in each [membrane](#), occur at contact sites, require energy, and use [amphipathic](#) N-terminal signal sequences that are removed after use. With the exception of some of the chaperone molecules, however, the [protein](#) components that form the translocation complexes are different. Moreover, whereas mitochondria harness the electrochemical H^+ gradient across their [inner membrane](#) to drive **transport**, chloroplasts, which have an electrochemical H^+ gradient across their [thylakoid](#) membrane but not their inner membrane, use the hydrolysis of GTP and ATP to power import across their double membrane. The functional similarities may thus result from convergent evolution, reflecting the common requirements for translocation across a double-membrane system.

Although the signal sequences for import into chloroplasts superficially resemble those for import into mitochondria, mitochondria and chloroplasts are both present in the same plant cells, so proteins must choose appropriately between them. In plants, for example, a bacterial [enzyme](#) can be directed specifically to mitochondria if it is experimentally joined to an N-terminal [signal sequence](#) of a mitochondrial [protein](#); the same enzyme joined to an N-terminal signal sequence of a [chloroplast protein](#) ends up in chloroplasts. The different signal sequences can therefore be distinguished by the import receptors on each [organelle](#).

Chloroplasts have an extra [membrane](#)-enclosed [compartment](#), the [thylakoid](#). Many [chloroplast](#) proteins, including the [protein](#) subunits of the photosynthetic system and of

the [ATP synthase](#) (discussed in Chapter 14), are embedded in the [thylakoid](#) membrane. Like the precursors of some mitochondrial proteins, these proteins are transported from the [cytosol](#) to their final destination in two steps. First, they pass across the double membrane at contact sites into the [matrix space](#) of the chloroplast, called the [stroma](#), and then they are translocated into the thylakoid membrane (or across this membrane into the thylakoid space) ([Figure 12-30A](#)). The precursors of these proteins have a hydrophobic thylakoid [signal sequence](#) following the N-terminal chloroplast signal sequence. After the N-terminal signal sequence has been used to import the **protein** into the [stroma](#), it is removed by a stromal [signal peptidase](#) (analogous to the matrix signal peptidase in mitochondria). This [cleavage](#) unmasks the thylakoid signal sequence, which then initiates **transport** across the thylakoid membrane. There are at least four routes for proteins to cross or become integrated into the thylakoid membrane, distinguished by their need for different stromal chaperones and energy sources.

Summary

Although mitochondria and chloroplasts have their own genetic systems, they produce only a small proportion of their own proteins. Instead, the two organelles import most of their proteins from the [cytosol](#), using similar mechanisms. In both cases, proteins are imported in an unfolded state. Proteins are translocated into the mitochondrial [matrix space](#) by passing through the TOM and [TIM complexes](#) at sites of adhesion between the outer and inner membranes known as contact sites. Translocation into mitochondria is driven by both ATP hydrolysis and an electrochemical H^+ gradient across the [inner membrane](#), whereas [translocation](#) into chloroplasts is driven solely by the hydrolysis of GTP and ATP.

Chaperone proteins of the cytosolic hsp70 family maintain the precursor proteins in an unfolded, [translocation](#)-competent state. A second set of hsp70 proteins in the matrix or [stroma](#) bind to the incoming [polypeptide](#) chain to pull it into the [organelle](#). Only proteins that contain a specific [signal sequence](#) are translocated into mitochondria or chloroplasts. The signal sequence is usually located at the [N terminus](#) and is cleaved off after import. Some imported proteins also contain an internal signal sequence that guides their further **transport**. **Transport** across or into the [inner membrane](#) can occur as a second step if a hydrophobic signal sequence is unmasked when the first signal sequence is removed. In chloroplasts, import from the stroma into the [thylakoid](#) likewise requires a second signal sequence and can occur by one of several routes.

The Endoplasmic Reticulum

All eucaryotic cells have an [endoplasmic reticulum \(ER\)](#). Its [membrane](#) typically constitutes more than half of the total membrane of an average animal cell. The ER is organized into a netlike labyrinth of branching tubules and flattened sacs extending throughout the [cytosol](#) ([Figure 12-35](#)). The tubules and sacs are all thought to interconnect, so that the ER membrane forms a continuous sheet enclosing a single internal space. This highly convoluted space is called the [ER lumen](#) or the *ER cisternal space*, and it often occupies

more than 10% of the total cell volume (see [Table 12-1](#)). The ER membrane separates the [ER lumen](#) from the cytosol, and it mediates the selective transfer of molecules between these two compartments.

The [ER](#) has a central role in [lipid](#) and [protein](#) biosynthesis. Its [membrane](#) is the site of production of all the transmembrane proteins and lipids for most of the cell's organelles, including the ER itself, the Golgi apparatus, lysosomes, endosomes, secretory vesicles, and the [plasma membrane](#). The ER membrane makes a major contribution to mitochondrial and peroxisomal membranes by producing most of their lipids. In addition, almost all of the proteins that will be secreted to the cell exterior—plus those destined for the [lumen](#) of the ER, Golgi apparatus, or lysosomes—are initially delivered to the [ER lumen](#).

Membrane-bound Ribosomes Define the Rough ER

The [ER](#) captures selected proteins from the [cytosol](#) as they are being synthesized. These proteins are of two types: *transmembrane proteins*, which are only partly translocated across the ER [membrane](#) and become embedded in it, and *water-soluble proteins*, which are fully translocated across the ER membrane and are released into the [ER lumen](#). Some of the transmembrane proteins function in the ER, but many are destined to reside in the [plasma membrane](#) or the membrane of another [organelle](#). The water-soluble proteins are destined either for the lumen of an organelle or for secretion. All of these proteins, regardless of their subsequent fate, are directed to the ER membrane by the same kind of [signal sequence](#) and are translocated across it by similar mechanisms.

In mammalian cells, the import of proteins into the [ER](#) begins before the [polypeptide](#) chain is completely synthesized—that is, import is a [co-translational](#) process. This distinguishes the process from the import of proteins into mitochondria, chloroplasts, nuclei, and peroxisomes, which are [posttranslational](#) processes. Since one end of the [protein](#) is usually translocated into the ER as the rest of the polypeptide chain is being made, the protein is never released into the [cytosol](#) and therefore is never in danger of folding up before reaching the translocator in the ER [membrane](#). Thus, in contrast to the [posttranslational](#) import of proteins into mitochondria and chloroplasts, chaperone proteins are not required to keep the protein unfolded. The [ribosome](#) that is synthesizing the protein is directly attached to the ER membrane. These membrane-bound ribosomes coat the surface of the ER, creating regions termed [rough endoplasmic reticulum](#), or **rough ER**.

There are therefore two spatially separate populations of ribosomes in the [cytosol](#). [Membrane-bound ribosomes](#), attached to the cytosolic side of the [ER membrane](#), are engaged in the synthesis of proteins that are being concurrently translocated into the ER. [Free ribosomes](#), unattached to any membrane, synthesize all other proteins encoded by the nuclear [genome](#). Membrane-bound and free ribosomes are structurally and functionally identical. They differ only in the proteins they are making at any given time. When a [ribosome](#) happens to be making a [protein](#) with an [ER signal sequence](#), the signal directs the ribosome to the ER membrane.

Since many ribosomes can bind to a single [mRNA molecule](#), a [polyribosome](#) is usually formed, which becomes attached to the [ER membrane](#), directed there by the signal sequences on multiple growing [polypeptide](#) chains. The individual ribosomes associated with such an mRNA molecule can return to the [cytosol](#) when they finish translation near the 3' end of the mRNA molecule. The mRNA itself, however, remains attached to the ER membrane by a changing population of ribosomes, each transiently held at the membrane by the translocator. In contrast, if an mRNA molecule encodes a [protein](#) that lacks an [ER signal sequence](#), the polyribosome that forms remains free in the cytosol, and its protein product is discharged there. Therefore, only those mRNA molecules that encode proteins with an ER signal sequence bind to rough ER membranes; those mRNA molecules that encode all other proteins remain free in the cytosol. Individual ribosomal subunits are thought to move randomly between these two segregated populations of mRNA molecules.

Smooth ER Is Abundant in Some Specialized Cells

Regions of [ER](#) that lack bound ribosomes are called [smooth endoplasmic reticulum](#), or **smooth ER**. In the great majority of cells, such regions are scanty and are often partly smooth and partly rough. They are sometimes called *transitional ER* because they contain *ER exit sites* from which transport vesicles carrying newly synthesized proteins and lipids bud off for transport to the Golgi apparatus. In certain specialized cells, however, the smooth ER is abundant and has additional functions. In particular, it is usually prominent in cells that specialize in [lipid metabolism](#). Cells that synthesize [steroid](#) hormones from [cholesterol](#), for example, have an expanded smooth ER [compartment](#) to accommodate the enzymes needed to make cholesterol and to modify it to form the hormones.

The main cell type in the liver, the [hepatocyte](#), is another cell with an abundant smooth [ER](#). It is the principal site of production of lipoprotein particles, which carry lipids via the bloodstream to other parts of the body. The enzymes that synthesize the [lipid](#) components of lipoproteins are located in the [membrane](#) of the smooth ER, which also contains enzymes that catalyze a series of reactions to detoxify both lipid-soluble drugs and various harmful compounds produced by [metabolism](#). The most extensively studied of these *detoxification reactions* are carried out by the [cytochrome P450](#) family of enzymes, which catalyze a series of reactions in which water-insoluble drugs or metabolites that would otherwise accumulate to toxic levels in cell membranes are rendered sufficiently water-soluble to leave the cell and be excreted in the urine. Because the rough ER alone cannot house enough of these and other necessary enzymes, a major portion of the membrane in a hepatocyte normally consists of smooth ER.

When large quantities of certain compounds, such as the drug phenobarbital, enter the circulation, detoxification enzymes are synthesized in the liver in unusually large amounts, and the smooth [ER](#) doubles in surface area within a few days. Once the drug has disappeared, the excess smooth ER [membrane](#) is specifically and rapidly removed by [lysosome](#)-dependent process called *autophagocytosis* (discussed in Chapter 13). It is not known how these dramatic changes are regulated.

Another function of the [ER](#) in most eucaryotic cells is to sequester Ca^{2+} from the [cytosol](#). The release of Ca^{2+} into the cytosol from the ER, and its subsequent reuptake, is involved in many rapid responses to extracellular signals, as discussed in Chapter 15. The storage of Ca^{2+} in the [ER lumen](#) is facilitated by the high concentrations of Ca^{2+} -binding proteins there. In some cell types, and perhaps in most, specific regions of the ER are specialized for Ca^{2+} storage. Muscle cells, for example, have an abundant specialized smooth ER, called the [sarcoplasmic reticulum](#), which sequesters Ca^{2+} from the cytosol by means of a Ca^{2+} -[ATPase](#) that pumps in Ca^{2+} into its lumen. The release and reuptake of Ca^{2+} by the sarcoplasmic reticulum trigger the contraction and relaxation, respectively, of the myofibrils during each round of muscle contraction.

Rough and Smooth Regions of ER Can Be Separated by Centrifugation

To study the functions and biochemistry of the [ER](#), it is necessary to isolate the ER [membrane](#). This may seem like a hopeless task because the ER is intricately interleaved with other components of the [cytosol](#). Fortunately, when tissues or cells are disrupted by homogenization, the ER breaks into fragments and reseals into many small ($\sim 100\text{--}200$ nm in diameter) closed vesicles called [microsomes](#), which are relatively easy to purify. Microsomes derived from rough ER are studded with ribosomes and are called *rough microsomes*. The ribosomes are always found on the outside surface, so the interior of the [microsome](#) is biochemically equivalent to the luminal space of the ER. Because they can be readily purified in functional form, rough microsomes are especially useful for studying the many processes performed by the rough ER. To the biochemist they represent small authentic versions of the rough ER, still capable of [protein](#) synthesis, [protein glycosylation](#), Ca^{2+} uptake, and [lipid](#) synthesis.

Many vesicles of a size similar to that of rough microsomes, but lacking attached ribosomes, are also found in these homogenates. Such *smooth microsomes* are derived in part from smooth portions of the [ER](#) and in part from vesiculated fragments of the [plasma membrane](#), Golgi apparatus, endosomes, and mitochondria (the ratio depending on the tissue). Thus, whereas rough microsomes are derived from rough portions of ER, the origins of smooth microsomes cannot be as easily assigned. The microsomes of the liver are an exception. Because of the unusually large quantities of smooth ER in hepatocytes, most of the smooth microsomes in liver homogenates are derived from smooth ER.

The ribosomes attached to rough microsomes make them more dense than smooth microsomes. As a result, the rough and smooth microsomes can be separated from each other by [equilibrium](#) centrifugation. When the separated rough and smooth microsomes of liver are compared with regard to such properties as [enzyme](#) activity or [polypeptide](#) composition, they are very similar, although not identical: apparently most of the components of the [ER membrane](#) can diffuse freely between the rough and smooth regions, as would be expected for a continuous, fluid membrane. The rough microsomes, however, contain more than 20 proteins that are not present in smooth microsomes, showing that some separation mechanism must exist for a subset of ER membrane proteins. Some of the proteins in this

subset help to bind ribosomes to the rough ER, while others presumably produce the flattened shape of this part of the ER. It is not clear whether these membrane proteins are confined to the rough ER by forming large two-dimensional assemblies in the [lipid bilayer](#), or whether they are instead held in place by interactions with a network of structural proteins on one or the other face of the rough ER membrane.

Signal Sequences Were First Discovered in Proteins Imported into the Rough ER

Signal sequences (and the [signal sequence](#) strategy of [protein](#) sorting) were first discovered in the early 1970s in secreted proteins that are translocated across the [ER membrane](#) as a first step toward their eventual discharge from the cell. In the key experiment, the [mRNA](#) encoding a secreted protein was translated by ribosomes *in vitro*. When microsomes were omitted from this [cell-free system](#), the protein synthesized was slightly larger than the normal secreted protein, the extra length being the N-terminal leader peptide. In the presence of microsomes derived from the rough ER, however, a protein of the correct size was produced. These results were explained by the *signal hypothesis*, which postulated that the leader serves as an [ER signal sequence](#) that directs the secreted protein to the ER membrane and is then cleaved off by a [signal peptidase](#) in the ER membrane before the [polypeptide](#) chain has been completed.

According to the signal hypothesis, the secreted [protein](#) should be extruded into the [lumen](#) of the [microsome](#) during its synthesis *in vitro*. This can be demonstrated by treatment with a protease: a newly synthesized protein made in the absence of microsomes is degraded when the protease is added to the medium, whereas the same protein made in the presence of microsomes remains intact because it is protected by the microsomal [membrane](#). When proteins without [ER](#) signal sequences are similarly synthesized *in vitro*, they are not imported into microsomes and are therefore degraded by protease treatment.

The signal hypothesis has been thoroughly tested by genetic and biochemical experiments and is found to apply to both plant and animal cells, as well as to [protein translocation](#) across the bacterial [plasma membrane](#) and, as we have seen, the membranes of mitochondria, chloroplasts, and peroxisomes. N-terminal [ER](#) signal sequences guide not only soluble secreted proteins, but also the precursors of all other proteins made by ribosomes bound to the rough ER membrane, including membrane proteins. The signaling function of these peptides has been demonstrated directly by using [recombinant DNA](#) techniques to attach ER signal sequences to proteins that do not normally have them; the resulting fusion proteins are directed to the ER.

Cell-free systems in which proteins are imported into microsomes have provided powerful assay procedures for identifying, purifying, and studying the various components of the molecular machinery responsible for the [ER](#) import process.

A Signal-Recognition Particle (SRP) Directs ER Signal Sequences to a Specific Receptor in the Rough ER Membrane

The [ER signal sequence](#) is guided to the ER [membrane](#) by at least two components: a [signal-recognition particle \(SRP\)](#), which cycles between the ER membrane and the [cytosol](#) and binds to the signal sequence, and an [SRP receptor](#) in the ER membrane. The SRP is a [complex](#) particle consisting of six different [polypeptide](#) chains bound to a single small [RNA molecule](#) (Figure 12-41A). Homologs of the SRP and its receptor are found in all organisms that have been studied, indicating that this [protein](#)-targeting mechanism arose early in evolution and has been conserved.

[ER](#) signal sequences vary greatly in [amino acid](#) sequence, but each has eight or more nonpolar amino acids at its center. How can the [SRP](#) bind specifically to so many different sequences? The answer has come from the crystal structure of the SRP [protein](#), which shows that the signal-sequence-[binding site](#) is a large hydrophobic pocket lined by methionines. Because methionines have an unbranched, flexible side chains, the pocket is sufficiently plastic to accommodate hydrophobic signal sequences of different sequences and shapes.

The [SRP](#) binds to the [ER signal sequence](#) as soon as the peptide has emerged from the [ribosome](#). This causes a pause in [protein](#) synthesis, the pause presumably gives the ribosome enough time to bind to the ER [membrane](#) before the synthesis of the [polypeptide](#) chain is completed, thereby ensuring that the protein is not released into the [cytosol](#). This safety device may be especially important for secreted and lysosomal hydrolases that could wreak havoc in the cytosol; however, cells that secrete large amounts of hydrolases take the added precaution of having high concentrations of hydrolase inhibitors in their cytosol.

Once formed, the [SRP-ribosome complex](#) binds to the SRP [receptor](#), which is an [integral membrane protein](#) exposed only on the cytosolic surface of the rough [ER](#) membrane. This interaction brings the SRP-ribosome complex to a [protein translocator](#). The SRP and SRP receptor are then released, and the growing [polypeptide](#) chain is transferred across the membrane.

The Polypeptide Chain Passes Through an Aqueous Pore in the Translocator

It has long been debated whether [polypeptide](#) chains are transferred across the [ER membrane](#) in direct contact with the [lipid bilayer](#) or through a pore in a [protein translocator](#). The debate ended with the purification of the protein translocator, which was shown to form a water-filled pore in the membrane through which the polypeptide chain traverses the membrane. The translocator, called the **Sec61 complex**, consists of three or four protein complexes, each composed of three transmembrane proteins, that assemble into a donutlike structure.

When a [ribosome](#) binds, the central hole in the translocator lines up with a tunnel in the large ribosomal [subunit](#) through which the growing [polypeptide](#) chain exits from the ribosome. The bound ribosome forms a tight seal with the translocator, such that the space inside the ribosome is continuous with the [lumen](#) of the [ER](#) and no molecules can escape from the ER. The pore in the translocator cannot be open permanently, however; if it were, Ca^{2+} would leak

out of the ER when the ribosome detaches. It is thought that a luminal ER [protein](#) serves as a plug or that the translocator itself can rearrange to close the pore when no ribosome is bound. Thus, the pore is a dynamic structure that opens only transiently when a ribosome with a growing polypeptide chain attaches to the ER [membrane](#).

The [signal sequence](#) in the growing [polypeptide](#) chain is thought to trigger the opening of the pore: after the signal sequence is released from the [SRP](#) and the growing chain has reached a sufficient length, the signal sequence binds to a specific site inside the pore itself, thereby opening the pore. An [ER signal sequence](#) is therefore recognized twice: first, by an SRP in the [cytosol](#), and then by a [binding site](#) in the ER [protein translocator](#). This may help to ensure that only appropriate proteins enter the [lumen](#) of the ER.

Translocation Across the ER Membrane Does Not Always Require Ongoing Polypeptide Chain Elongation

As we have seen, [translocation](#) of proteins into mitochondria, chloroplasts, and peroxisomes occurs posttranslationally, after the [protein](#) has been made and released into the [cytosol](#), whereas translocation across the [ER membrane](#) usually occurs during translation (co-translationally). This explains why ribosomes are bound to the ER but usually not to other organelles.

Some proteins, however, are imported into the [ER](#) after their synthesis has been completed, demonstrating that [translocation](#) does not always require ongoing translation.

Posttranslational [protein](#) translocation is especially common across the ER [membrane](#) in [yeast](#) cells and across the bacterial [plasma membrane](#) (which is thought to be evolutionarily related to the ER; see [Figure 12-4](#)). To function in [posttranslational](#) translocation, the translocator needs accessory proteins that feed the [polypeptide](#) chain into the pore and drive translocation. In bacteria, a translocation [motor protein](#), the SecA [ATPase](#), attaches to the cytosolic side of the translocator, where it undergoes cyclic conformational changes driven by ATP hydrolysis. Each time an ATP is hydrolyzed, a portion of the SecA protein inserts into the pore of the translocator, pushing a short segment of the passenger protein with it. As a result of this ratchet mechanism, the SecA protein pushes the polypeptide chain of the transported protein across the membrane.

Eucaryotic cells use a different set of accessory proteins that associate with the Sec61 [complex](#). These proteins span the [ER membrane](#) and use a small [domain](#) on the luminal side of the ER membrane to deposit an hsp70-like chaperone [protein](#) (called *BiP*, for *binding protein*) onto the [polypeptide](#) chain as it emerges from the pore into the [ER lumen](#). Unidirectional [translocation](#) is driven by cycles of BiP binding and release, as described earlier for the mitochondrial hsp70 proteins that pull proteins across mitochondrial membranes.

Proteins that are transported into the [ER](#) by a [posttranslational](#) mechanism are first released into the [cytosol](#), where they are prevented from folding up by binding to chaperone proteins, as discussed earlier for proteins destined for mitochondria and chloroplasts. In all of these

cases where [translocation](#) occurs without a [ribosome](#) sealing the pore, it remains a mystery how the [polypeptide](#) chain can slide through the pore in the translocator without allowing ions and other molecules to pass through.

The ER Signal Sequence Is Removed from Most Soluble Proteins After Translocation

We have seen that in chloroplasts and mitochondria, the [signal sequence](#) is cleaved from precursor proteins once it has crossed the [membrane](#). Similarly, N-terminal [ER](#) signal sequences are removed by a [signal peptidase](#) on the luminal side of the ER membrane. The signal sequence by itself, however, is not sufficient for signal [cleavage](#) by the peptidase; this requires an adjacent cleavage site that is specifically recognized by the peptidase. We shall see below that ER signal sequences that occur within the [polypeptide](#) chain—rather than at the N-terminus—do not have these recognition sites and are never cleaved; instead, they can serve to retain transmembrane proteins in the [lipid bilayer](#) after the [translocation](#) process has been completed.

The N-terminal [ER signal sequence](#) of a soluble [protein](#) has two signaling functions. It directs the protein to the [ER membrane](#), and it serves as a [start-transfer signal](#) (or start-transfer peptide) that opens the pore. Even after it is cleaved off by [signal peptidase](#), the signal sequence is thought to remain bound to the translocator while the rest of the protein is threaded continuously through the membrane as a large loop. Once the C-terminus of the protein has passed through the membrane, the translocated protein is released into the [ER lumen](#). The signal sequence is released from the pore and rapidly degraded to amino acids by other proteases in the ER.

While bound in the [translocation](#) pore, signal sequences are in contact not only with the Sec61 [complex](#), which forms the walls of the pore, but also with the hydrophobic [lipid](#) core of the [membrane](#). This was shown in chemical cross-linking experiments in which signal sequences and the [hydrocarbon](#) chains of lipids could be covalently linked together. To release the [signal sequence](#) into the membrane, the translocator has to open laterally. The translocator is therefore gated in two directions: it can open to form a pore across the membrane to let the [hydrophilic](#) portions of proteins cross the [lipid bilayer](#), and it can open laterally within the membrane to let hydrophobic portions of proteins partition into the bilayer. This lateral gating mechanism is crucial for the insertion of transmembrane proteins into the lipid bilayer, as we discuss next.

In Single-Pass Transmembrane Proteins, a Single Internal ER Signal Sequence Remains in the Lipid Bilayer as a Membrane-spanning α Helix

The [translocation](#) process for proteins destined to remain in the [membrane](#) is more [complex](#) than it is for soluble proteins, as some parts of the [polypeptide](#) chain are translocated across the [lipid bilayer](#) whereas others are not. Nevertheless, all modes of insertion of membrane proteins can be considered as variants of the sequence of events just

described for transferring a soluble [protein](#) into the [lumen](#) of the [ER](#). We begin by describing the three ways in which [single-pass transmembrane proteins](#) become inserted into the ER.

In the simplest case, an N-terminal [signal sequence](#) initiates [translocation](#), just as for a soluble [protein](#), but an additional hydrophobic segment in the [polypeptide](#) chain stops the transfer process before the entire polypeptide chain is translocated. This [stop-transfer signal](#) anchors the protein in the [membrane](#) after the [ER signal sequence](#) (the [start-transfer signal](#)) has been released from the translocator and has been cleaved off. The stop-transfer sequence is transferred into the bilayer by the lateral gating mechanism, and it remains there as a single α -helical membrane-spanning segment, with the N-terminus of the protein on the luminal side of the membrane and the C-terminus on the cytosolic side.

In the other two cases, the [signal sequence](#) is internal, rather than at the N-terminal end of the [protein](#). Like the N-terminal [ER](#) signal sequences, the internal signal sequence is recognized by an [SRP](#), which brings the [ribosome](#) making the protein to the ER [membrane](#) and serves as a [start-transfer signal](#) that initiates the [translocation](#) of the protein. After release from the translocator, the internal start-transfer sequence remains in the [lipid bilayer](#) as a single membrane-spanning [\$\alpha\$ helix](#).

Internal start-transfer sequences, can bind to the [translocation](#) apparatus in either of two orientations, and the orientation of the inserted start-transfer sequence, in turn, determines which [protein](#) segment (the one preceding or the one following the start-transfer sequence) is moved across the [membrane](#) into the [ER lumen](#). In one case, the resulting [membrane protein](#) has its C-terminus on the luminal side, while in the other, it has its N-terminus on the luminal side. The orientation of the start-transfer sequence depends on the distribution of nearby charged amino acids, as described in the figure legend.

Combinations of Start-Transfer and Stop-Transfer Signals Determine the Topology of Multipass Transmembrane Proteins

In [multipass transmembrane proteins](#), the [polypeptide](#) chain passes back and forth repeatedly across the [lipid bilayer](#) (see [Figure 10-17](#)). It is thought that an internal [signal sequence](#) serves as a [start-transfer signal](#) in these proteins to initiate [translocation](#), which continues until a stop-transfer sequence is reached. In double-pass transmembrane proteins, for example, the polypeptide can then be released into the bilayer. In more [complex](#) multipass proteins, in which many hydrophobic α helices span the bilayer, a second start-transfer sequence reinitiates translocation further down the polypeptide chain until the next stop-transfer sequence causes polypeptide release, and so on for subsequent start-transfer and stop-transfer sequences.

Whether a given hydrophobic [signal sequence](#) functions as a start-transfer or stop-transfer sequence must depend on its location in a [polypeptide](#) chain, since its function can be switched by changing its location in the [protein](#) using [recombinant DNA](#) techniques. Thus, the distinction between start-transfer and stop-transfer sequences results mostly from their relative order in the growing polypeptide chain. It seems that the [SRP](#) begins scanning an

unfolded polypeptide chain for hydrophobic segments at its N-terminus and proceeds toward the C-terminus, in the direction that the protein is synthesized. By recognizing the first appropriate hydrophobic segment to emerge from the [ribosome](#), the SRP sets the “[reading frame](#)”: if [translocation](#) is initiated, the next appropriate hydrophobic segment is recognized as a stop-transfer sequence, causing the region of the polypeptide chain in between to be threaded across the [membrane](#). A similar scanning process continues until all of the hydrophobic regions in the protein have been inserted into the membrane.

Because [membrane](#) proteins are always inserted from the cytosolic side of the [ER](#) in this programmed manner, all copies of the same [polypeptide](#) chain will have the same orientation in the [lipid bilayer](#). This generates an asymmetrical ER membrane in which the [protein](#) domains exposed on one side are different from those domains exposed on the other. This asymmetry is maintained during the many membrane budding and fusion events that transport the proteins made in the ER to other cell membranes. Thus, the way in which a newly synthesized protein is inserted into the ER membrane determines the orientation of the protein in all of the other membranes as well.

When proteins are dissociated from a [membrane](#) and are then reconstituted into artificial [lipid](#) vesicles, a random mixture of right-side-out and inside-out [protein](#) orientations usually results. Thus, the protein asymmetry observed in cell membranes seems not to be an inherent property of the protein, but instead results solely from the process by which proteins are inserted into the [ER](#) membrane from the [cytosol](#).

Intracellular Vesicular Traffic

Every cell must eat, and it must communicate with the world around it. In a procaryotic cell, all the eating and communicating takes place across the [plasma membrane](#). The cell secretes digestive enzymes, for example, across the plasma membrane to the cell exterior. It then transports the small metabolites generated by digestion in the extracellular space across the same membrane into the [cytosol](#). Eucaryotic cells, by contrast, have evolved an elaborate [internal membrane](#) system that allows them to take up macromolecules by the process of [endocytosis](#) and deliver them to digestive enzymes stored in lysosomes inside the cell. As a consequence, metabolites generated by the digestion of macromolecules are delivered directly from the lysosomes to the cytosol as they are produced. In addition to allowing the ingestion of macromolecules by the *endocytic pathway*, the internal membrane system allows eucaryotic cells to regulate the delivery of newly synthesized proteins, carbohydrates, and lipids to the cell exterior. The *biosynthetic-secretory pathway* allows the cell to modify the molecules it produces in a series of steps, store them until needed, and then deliver them to the exterior through a specific cell-surface [domain](#) by a process called [exocytosis](#). An outline of the endocytic and biosynthetic-secretory pathways, which

ultimately connect the plasma membrane to the [endoplasmic reticulum](#) ([ER](#)) deep within the cell, is shown in.

The interior space, or [lumen](#), of each [membrane](#)-enclosed [compartment](#) along the biosynthetic-secretory and endocytic pathways is topologically equivalent to the lumen of every other compartment. Moreover, these compartments are in constant communication, with molecules being passed from a donor compartment to a target compartment by means of numerous membrane-enclosed *transport packages*. Some of these packages are small spherical vesicles, while others are larger irregular vesicles or fragments of the donor compartment. We shall use the term *transport vesicle* to apply to all forms of these packages.

Vesicles continually bud off from one [membrane](#) and fuse with another, carrying membrane components and soluble molecules referred to as *cargo*. This membrane **traffic** flows along highly organized, directional routes, which allows the cell to secrete and eat. The biosynthetic-secretory pathway leads outward from the [ER](#) toward the Golgi apparatus and cell surface, with a side route leading to lysosomes, while the endocytic pathway leads inward from the [plasma membrane](#). In each case, the flow of membrane between compartments is balanced, with retrieval pathways balancing the flow in the opposite direction, bringing membrane and selected proteins back to the [compartment](#) of origin.

To perform its function, each transport [vesicle](#) that buds from a [compartment](#) must be selective. It must take up only the appropriate proteins and must fuse only with the appropriate target [membrane](#). A vesicle carrying cargo from the Golgi apparatus to the [plasma membrane](#), for example, must exclude proteins that are to stay in the Golgi apparatus, and it must fuse only with the plasma membrane and not with any other [organelle](#).

We begin this chapter by considering the molecular mechanisms of budding and fusion that underlie all transport. We then discuss the fundamental problem of how, in the face of this transport, the differences between the compartments are maintained. Finally, we consider the function of the Golgi apparatus, lysosomes, secretory vesicles, and endosomes, as we trace the pathways that connect these organelles.

The Molecular Mechanisms of Membrane Transport and the Maintenance of Compartmental Diversity

Transport processes mediate a continual exchange of components between the ten or more chemically distinct, [membrane](#)-enclosed compartments that collectively comprise the biosynthetic-secretory and endocytic pathways. In the presence of this massive exchange, how can each [compartment](#) maintain its specialized character? To answer this question, we must first consider what defines the character of a compartment. Above all, it is the composition of the enclosing membrane: molecular markers displayed on the cytosolic surface of this membrane serve as guidance cues for incoming **traffic** and ensure that transport vesicles fuse only with the correct compartment, thereby dictating the pattern of **traffic** between one compartment and another. Many membrane markers, however, are

found on more than one [organelle](#), and thus it is the specific combination of marker molecules that gives each organelle its unique molecular address.

How are these [membrane](#) markers kept at high concentration on one [compartment](#) and at low concentration on another? To answer this question, we need to consider how patches of membrane, enriched or depleted in specific components, bud off from one compartment and transfer to another. In this [section](#) we describe how this is achieved.

We begin by discussing the sorting events that underlie the segregation of proteins into separate [membrane](#) domains. This sorting process depends on the assembly of a special [protein](#) coat on the cytosolic face of the donor membrane. We shall therefore consider how coats form, what they are made of, and how they enable specific components of a membrane to be extracted and delivered to another membrane. Finally, we discuss how transport vesicles dock at the appropriate target membrane and fuse with it to deliver the contents to their target [organelle](#).

There Are Various Types of Coated Vesicles

Most transport vesicles form from specialized, coated regions of membranes. They bud off as [coated vesicles](#) that have a distinctive cage of proteins covering their cytosolic surface. Before the [vesicle](#) fuses with a target [membrane](#), the coat is discarded, as is required to allow the two cytosolic membrane surfaces to interact directly and fuse.

The coat is thought to perform two principal functions. First, it concentrates specific [membrane](#) proteins in a specialized membrane patch that then gives rise to the [vesicle](#) membrane. It thus helps select the appropriate molecules for transport. Second, the assembly of the coat proteins into curved, basketlike lattices deforms the membrane patch and thereby molds the forming vesicles, which explains why vesicles with the same type of coat have a relatively uniform size.

There are three well-characterized types of coated vesicles, which differ in their coat proteins: [clathrin-coated](#), [COPI-coated](#), and [COPII-coated](#) vesicles. Each type is used for different transport steps in the cell. Clathrin-coated vesicles, for example, mediate transport from the Golgi apparatus and from the [plasma membrane](#), whereas COPI- and COPII-coated vesicles most commonly mediate transport from the [ER](#) and the Golgi. There is, however, much more variety than this short list suggests. As we discuss below, there are at least three types of clathrin-coated vesicles, each specialized for a different transport step, and the COPI-coated vesicles may be similarly diverse. Moreover, still other coats have been seen in the [electron microscope](#), whose molecular compositions and functions are not yet known.

The Assembly of a Clathrin Coat Drives Vesicle Formation

Clathrin-coated vesicles were the first coated vesicles discovered and have been the most thoroughly studied. They provide a good example of how vesicles form.

The major [protein](#) component of [clathrin](#)-coated vesicles is [clathrin](#) itself. Each clathrin [subunit](#) consists of three large and three small [polypeptide](#) chains that together form a three-legged structure called a *triskelion*. Clathrin triskelions assemble into a basketlike convex framework of hexagons and pentagons to form coated pits on the cytosolic surface of membranes ([Figure 13-6](#)). Under appropriate conditions, isolated triskelions spontaneously self-assemble into typical polyhedral cages in a test tube, even in the absence of the [membrane](#) vesicles that these baskets normally enclose. Thus, the geometry of the clathrin cage is [determined](#) by the clathrin triskelion alone.

A second major coat [protein](#) in [clathrin](#)-coated vesicles is a multisubunit [complex](#) called [adaptin](#). It is required both to bind the clathrin coat to the [membrane](#) and to trap various transmembrane proteins, including transmembrane receptors that capture soluble cargo molecules inside the [vesicle](#)—so-called *cargo receptors*. In this way, a selected set of membrane proteins and the soluble proteins that interact with them are packaged into each newly formed clathrin-coated transport vesicle.

There are at least four types of adaptins, each specific for a different set of cargo receptors. Clathrin-coated vesicles budding from different membranes use different adaptins and thus package different receptors and cargo molecules. The formation of a [clathrin-coated pit](#) is driven by forces generated by the successive assembly of adaptins and the clathrin coat on the cytosolic surface of the [membrane](#). The lateral interactions between adaptins and between clathrin molecules then aid in bud formation.

Both The Pinching-off and Uncoating of Coated Vesicles Are Regulated Processes

As a [clathrin](#)-coated bud grows, soluble cytoplasmic proteins, including [dynamitin](#), assemble as a ring around the neck of each bud. Dynamitin is a [GTPase](#), which regulates the rate with which vesicles pinch off from the [membrane](#). In the pinching-off process, the two noncytosolic leaflets of the membrane are brought into close proximity and fuse, sealing off the forming [vesicle](#). To perform this task, [dynamitin](#) recruits other proteins to the neck of the budding vesicle, which together with dynamitin help to bend the membrane, either by directly distorting the bilayer structure locally or by changing the [lipid](#) composition, or both. A local change in lipid composition may result from the action of lipid-modifying enzymes that are recruited into the dynamitin [complex](#).

Once the [vesicle](#) is released from the [membrane](#), the [clathrin](#) coat is rapidly lost. A chaperone [protein](#) of the hsp70 family functions as an uncoating [ATPase](#), using the energy of ATP hydrolysis to peel off the coat. Another protein called *auxillin*, which is attached to the vesicle, is believed to activate the ATPase. Because the coated bud persists much longer than the coat on the vesicle, additional control mechanisms must somehow prevent the coat from being removed before it has formed a vesicle.

Although there are many similarities in [vesicle](#) budding at various locations in the cell, each cell [membrane](#) poses its own special challenges. The [plasma membrane](#), for example, is

comparatively flat and stiff, owing to its [cholesterol](#)-rich [lipid](#) composition and underlying cortical [cytoskeleton](#). Thus, [clathrin](#) coats have to produce considerable force to introduce curvature, especially at the neck of the bud where [dynamin](#) and its associated proteins facilitate the sharp bends required for the pinching-off of the vesicle. In contrast, vesicle budding from many **intracellular** membranes occurs preferentially at regions where the membranes are already curved, such as the rims of Golgi cisternae or membrane tubules.

COPI-coated vesicles and **COPII-coated vesicles** transport material early in the secretory pathway: COPII-coated packages bud from the [ER](#), and COPI-coated packages bud from pre-Golgi compartments and Golgi cisternae (see [Figure 13-5](#)). The coats of COPI and COPII vesicles consist, in part, of large [protein](#) complexes that are composed of seven individual coat-protein subunits for COPI and four for COPII coats. Some COPI coat-protein subunits show sequence similarity to adaptins, suggesting a common evolutionary origin.

Not All Transport Vesicles are Spherical

Transport vesicles occur in various sizes and shapes. When living cells that have been genetically engineered to express fluorescent [membrane](#) components are observed under the microscope, endosomes and the *trans* Golgi network are seen to continually send out long tubules. Coat proteins assemble onto the tubules and help recruit specific cargo. The tubules then either withdraw, or they pinch off with the help of [dynamin](#)-like proteins and thus can serve as transport vesicles. Depending on the relative efficiencies of membrane tubulation and severing, differently sized portions of a donor [organelle](#) can pinch off.

Tubules have a much higher surface-to-volume ratio than the organelles from which they form. They are therefore relatively enriched in [membrane](#) proteins compared with soluble cargo proteins. As we discuss later, this property of tubules is used for sorting proteins in endosomes. Thus, **vesicular transport** does not necessarily occur only through uniformly sized spherical vesicles, but can involve larger portions of a donor [organelle](#).

Monomeric GTPases Control Coat Assembly

The **vesicular transport** performed by both [clathrin](#)-coated and COP-coated vesicles depends on a variety of GTP-binding proteins that control both the spatial and the temporal aspects of [membrane](#) exchange. As discussed in Chapter 3, large families of GTP-binding proteins regulate diverse processes within cells. These proteins act as molecular switches that flip between an active state with GTP bound and an inactive state with GDP bound. Two classes of proteins regulate the flipping: *guanine-nucleotide-exchange factors (GEFs)* activate the proteins by catalyzing the exchange of GDP for GTP, and [GTPase-activating proteins \(GAPs\)](#) inactivate the proteins by triggering the hydrolysis of the bound GTP to GDP (see [Figure 3-72](#)). Although both monomeric GTP-binding proteins (monomeric GTPases) and trimeric GTP-binding proteins ([G](#) proteins) have essential roles in **vesicular** transport, the roles of the monomeric GTPases are better understood, and we focus our discussion on them.

To ensure that [membrane traffic](#) to and from an [organelle](#) is balanced, coat proteins must assemble only when and where they are needed. *Coat-recruitment GTPases*, which are members of a family of monomeric GTPases, usually serve this function. They include the [ARF proteins](#), which are responsible for both COPI coat assembly and [clathrin](#) coat assembly at Golgi membranes, and the **Sar1 protein**, which is responsible for COPII coat assembly at the [ER membrane](#). Clathrin coat assembly at the [plasma membrane](#) is also thought to involve a [GTPase](#), but its identity is unknown.

Coat-recruitment GTPases are usually found in high concentration in the [cytosol](#) in an inactive, GDP-bound state. When a COPII-[coated vesicle](#) is to bud from the [ER membrane](#), a specific [GEF](#) embedded in the ER membrane binds to cytosolic Sar1, causing the Sar1 to release its GDP and bind GTP in its place (recall that GTP is present in much higher concentration in the cytosol than GDP and therefore will spontaneously bind after GDP is released). In its GTP-bound state, Sar1 exposes a hydrophobic tail, which inserts into the [lipid bilayer](#) of the ER membrane. The tightly bound Sar1 now recruits coat [protein](#) subunits to the ER membrane to initiate budding. Other GEFs and coat-recruitment GTPases operate in a similar way on other membranes.

Some coat [protein](#) subunits also interact, albeit more weakly, with the head groups of certain [lipid](#) molecules, in particular phosphatidic [acid](#) and phosphoinositides, as well as with the cytoplasmic tails of some of the [membrane](#) proteins they recruit into the bud. Activated coat-recruitment GTPases at sites of bud formation can locally activate phospholipase D, which converts some phospholipids to phosphatidic acid, thereby enhancing the binding of coat proteins. Together, these protein-protein and protein-lipid interactions tightly bind the coat to the membrane, causing the membrane to deform into a bud, which then pinches off as a [coated vesicle](#).

The coat-recruitment GTPases also have a role in coat disassembly. The hydrolysis of bound GTP to GDP causes the [GTPase](#) to change its [conformation](#) so that its hydrophobic tail pops out of the [membrane](#), causing the [vesicle](#)'s coat to disassemble. Although it is not known what triggers the GTP hydrolysis process, it has been proposed that the GTPases work like timers, which hydrolyze GTP at a slow but predictable rate. COPII coats, for example, accelerate GTP hydrolysis by Sar1, thereby triggering coat disassembly at a certain time after coat assembly has begun. Thus, a fully formed vesicle will be produced only when bud formation occurs faster than the timed disassembly process; otherwise, disassembly will be triggered before a vesicle pinches off, and the process will have to start again at a more appropriate time and place. Completion of coating or contact with the target membrane may also trigger coat disassembly.

SNARE Proteins and Targeting GTPases Guide Membrane Transport

To ensure that [membrane traffic](#) proceeds in an orderly way, transport vesicles must be highly selective in recognizing the correct target membrane with which to fuse. Because of the diversity of membrane systems, a [vesicle](#) is likely to encounter many potential target

membranes before it finds the correct one. Specificity in targeting is ensured because all transport vesicles display surface markers that identify them according to their origin and type of cargo, while target membranes display [complementary](#) receptors that recognize the appropriate markers. This crucial recognition step is thought to be controlled mainly by two classes of proteins: [SNAREs](#) and targeting GTPases called *Rabs*. SNARE proteins seem to have a central role both in providing specificity and in catalyzing the fusion of vesicles with the target membrane. *Rabs* seem to work together with other proteins to regulate the initial docking and tethering of the vesicle to the target membrane.

There are at least 20 different [SNAREs](#) in an animal cell, each associated with a particular [membrane](#)-enclosed [organelle](#) involved in the biosynthetic-secretory or endocytic pathway. These transmembrane proteins exist as [complementary](#) sets—[vesicle](#) membrane [SNAREs](#), called **v-SNAREs**, and target membrane SNAREs, called **t-SNAREs**, ([Figures 13-11](#) and [13-12](#)). v-SNAREs and t-SNAREs have characteristic helical domains. When a [v-SNARE](#) interacts with a [t-SNARE](#), the helical domains of one wrap around the helical domains of the other to form stable *trans-SNARE complexes*, which lock the two membranes together. We discuss later how the *trans-SNARE complex* is thought to contribute to membrane fusion. The specificity with which SNAREs interact determines the specificity of vesicle docking and fusion. In this way SNAREs specify [compartment](#) identity and govern the orderly transfer of material during [vesicular transport](#).

[SNAREs](#) have been best characterized in nerve cells, where they mediate the docking and fusion of synaptic vesicles at the nerve terminal [plasma membrane](#) (see [Figure 13-12](#)). The SNARE complexes at neuron terminals are the targets of powerful neurotoxins that are secreted by the bacteria that cause tetanus and botulism. These toxins are highly specific proteases that enter specific neurons, cleave SNARE proteins in the nerve terminals and thereby block synaptic transmissions, often fatally.

[Interacting SNAREs Need To Be Pried Apart Before They Can Function Again](#)

Most SNARE proteins in cells have already participated in multiple rounds of [membrane](#) targeting and are sometimes present in a membrane as stable complexes with one or two partner [SNAREs](#) (see [Figure 13-11](#)). The complexes have to be disassembled before the SNAREs can mediate new rounds of transport. A crucial [protein](#) called **NSF** cycles between membranes and the [cytosol](#) and catalyzes the disassembly process. It is an [ATPase](#) that structurally resembles a minor class of cytosolic chaperone proteins that use the energy of ATP hydrolysis to solubilize and help refold denatured proteins. Similarly, **NSF** uses ATP to unravel the [coiled-coil](#) interaction between the helical domains of SNARE proteins, using several adaptor proteins to bind to the SNAREs.

The requirement for SNARE [complex](#) disassembly may help explain why membranes do not fuse indiscriminately in cells. If the t-SNAREs in a target [membrane](#) were always active, then any membrane containing an appropriate [v-SNARE](#) would fuse whenever the two membranes made contact. The requirement for **NSF**-mediated reactivation of SNAREs

allows the cell to control when and where membranes fuse. In addition, t-SNAREs in target membranes are often associated with inhibitory proteins that must be released before the [t-SNARE](#) can function. This release step may be controlled by the targeting GTPases, as we discuss next.

Rab Proteins Help Ensure the Specificity of Vesicle Docking

[Rab proteins](#) make an important contribution to the specificity of [vesicular transport](#). They are monomeric GTPases, and with over 30 known members, they are the largest subfamily of these GTPases. Like the [SNAREs](#), each [Rab protein](#) has a characteristic distribution on cell membranes and every [organelle](#) has at least one Rab protein on its cytosolic. Rab proteins are thought to facilitate and regulate the rate of [vesicle](#) docking and the matching of v-SNAREs and t-SNAREs, as required for [membrane](#) fusion.

Like the coat-recruitment GTPases discussed earlier (see [Figure 13-10](#)), Rab proteins cycle between a [membrane](#) and the [cytosol](#). In their GDP-bound state they are inactive and in the cytosol, and in their GTP-bound state they are active and associated with the membrane of an [organelle](#) or transport [vesicle](#). Many transport vesicles only form if a proper complement of SNARE and Rab proteins are included in the membrane, so as to allow the vesicle to dock and fuse appropriately.

The [amino acid](#) sequences of Rab proteins are most dissimilar near their C-terminal tails. Tail-swapping experiments indicate that the tail determines the **intracellular** location of each family member, presumably by enabling the [protein](#) to bind to [complementary](#) proteins, including GEFs, on the surface of the appropriate [organelle](#). Once in its GTP-bound state and [membrane](#)-bound through a [lipid](#) anchor, a [Rab protein](#) is thought to bind to other proteins (called Rab effectors) that facilitate the docking process.

In contrast to the highly conserved structure of Rab proteins, the structures of **Rab effectors** vary greatly from one [Rab protein](#) to the next. One Rab effector, for example, is a large protein [complex](#) that serves to direct vesicles to specific sites on the [plasma membrane](#) for [exocytosis](#). Vesicle fusion is limited to the region where this complex resides, even though the required t-SNAREs are uniformly distributed in the membrane. Some Rab effectors are long, filamentous, tethering proteins, which may restrict the movement of vesicles between adjacent Golgi cisternae. Others bind to Rab proteins in their active GTP-bound state and prevent premature GTP hydrolysis. Yet others are motor proteins that propel vesicles along [actin](#) filaments or microtubules to their proper target.

Although the Rab proteins and their effectors use widely different molecular mechanisms to influence [vesicular transport](#), they have a common function. They help concentrate and tether vesicles near their target site and trigger the release of SNARE control proteins. In this way Rab proteins speed up the process by which appropriate SNARE proteins in two membranes find each other. Some Rab proteins function on the [vesicle](#), whereas others function on the target [membrane](#). The pairing of v-SNAREs and t-SNAREs then locks the docked vesicle onto the target membrane, readying it for fusion, which we discuss next. After fusion,

the [Rab protein](#) hydrolyzes its bound GTP and the inactive GDP-bound protein returns to the [cytosol](#) to participate in another cycle of transport.

SNAREs May Mediate Membrane Fusion

Once a transport [vesicle](#) has recognized its target [membrane](#) and docked there, it unloads its cargo by membrane fusion. Fusion does not always follow immediately, however. As we discuss later, in the process of regulated [exocytosis](#), fusion is delayed until it is triggered by a specific extracellular signal.

Thus, docking and fusion are two distinct and separable processes. Docking requires only that the two membranes come close enough for proteins protruding from the [lipid](#) bilayers to interact and adhere. Fusion requires a much closer approach, bringing the lipid bilayers to within 1.5 [nm](#) of each other so that they can join. When the membranes are in such close apposition, lipids can flow from one bilayer to the other. For this close approach, water must be displaced from the [hydrophilic](#) surface of the [membrane](#)—a process that is energetically highly unfavorable. It seems likely that all membrane fusions in cells are catalyzed by specialized fusion proteins that provide a way to overcome this energy barrier. We have already discussed the role of [dynamin](#) in a related task during [clathrin-coated vesicle](#) budding.

[SNAREs](#) are thought to have a central role in [membrane](#) fusion. The formation of the SNARE [complex](#) may work like a winch, using the energy that is freed when the interacting helices wrap around each other to pull the membrane faces together, while simultaneously squeezing out water molecules from the interface ([Figure 13-15](#)). When liposomes containing purified v-SNAREs are mixed with liposomes containing matching t-SNAREs, their membranes fuse, albeit slowly. In the cell, other proteins recruited to the fusion site presumably cooperate with SNAREs to initiate fusion. Moreover, inhibitory proteins may have to be released to allow the complete zipping-up of SNARE pairs. In some cases, such as in regulated [exocytosis](#), a localized influx of Ca^{2+} triggers the fusion process.

Viral Fusion Proteins and SNAREs May Use Similar Strategies

Membrane fusion is important in other processes beside [vesicular transport](#). Examples are the fusion of the plasma membranes of sperm and [egg](#) that occurs at [fertilization](#) (discussed in Chapter 20) and the fusion of myoblasts during muscle cell [development](#). All cell [membrane](#) fusions require special proteins and are subject to tight controls, which ensure that only appropriate membranes fuse. The controls are crucial for maintaining both the identity of cells and the individuality of each type of [intracellular compartment](#).

The [membrane](#) fusions catalyzed by viral fusion proteins are the best understood. These proteins have a crucial role in permitting the entry of enveloped viruses (which have a [lipid](#)-bilayer-based membrane coat) into the cells that they. For example, viruses—such as human immunodeficiency [virus](#) ([HIV](#)), which causes AIDS—bind to cell-surface receptors, and then the viral and plasma membranes fuse. This fusion event allows the viral [nucleic acid](#) to enter the [cytosol](#), where it replicates. Other viruses, such as influenza virus, first enter the cell

by [receptor-mediated endocytosis](#) (discussed later) and are delivered to endosomes. In this case, the low [pH](#) in endosomes activates a fusion [protein](#) in the viral envelope that catalyzes the fusion of the viral and endosomal membranes. This likewise releases the viral nucleic acid into the cytosol.

The three-dimensional structures of the fusion proteins of influenza [virus](#) and [HIV](#) provide valuable insights into the molecular mechanism of the [membrane](#) fusion catalyzed by these proteins. An exposure of the influenza fusion [protein](#) to low [pH](#), or an exposure of HIV fusion protein to receptors on the target cell membrane, uncovers previously buried hydrophobic regions. These regions, called fusion peptides, are observed to then insert directly into the hydrophobic core of [lipid bilayer](#) of the target membrane. Thus, the viral fusion proteins are, for a moment, integral membrane proteins in two separate lipid bilayers. Structural rearrangements in the fusion proteins then bring the two lipid bilayers into very close apposition and destabilize them so that the bilayers fuse. For viral fusion, the fusion proteins are the only components required, supporting the possibility that [SNAREs](#) are also the central players in the process of bilayer fusion in cells.

Summary

The differences between the many different [membrane](#)-enclosed compartments in a eucaryotic cell are maintained by directed, selective transport of particular membrane components from one [compartment](#) to another. Transport vesicles, which can be spherical or tubular, bud from specialized coated regions of the donor membrane. The assembly of the coat helps to collect specific membrane and soluble cargo molecules for transport and to drive the formation of the [vesicle](#).

Of the various types of coated vesicles, the best characterized are [clathrin](#)-coated vesicles, which mediate transport from the [plasma membrane](#) and the [trans Golgi network](#), and COPI- and COPII-coated vesicles, which mediate transport between the [ER](#) and the Golgi apparatus and between Golgi cisternae. In clathrin-coated vesicles, adaptins link the clathrin to the [vesicle](#) membrane and also trap specific cargo molecules for packaging into the vesicle. The coat is shed rapidly after budding, which is necessary for a vesicle to fuse with its appropriate target membrane.

Monomeric GTPases help regulate various steps in [vesicular transport](#), including both [vesicle](#) budding and docking. The coat-recruitment GTPases, including Sar1 and the ARF proteins, regulate coat assembly and disassembly. A family of Rab proteins functions as vesicle targeting GTPases. Being incorporated with v-[SNAREs](#) into budding transport vesicles, the Rab proteins help ensure that the vesicles deliver their contents only to the appropriate [membrane](#)-enclosed [compartment](#): the one that displays [complementary t-SNARE](#) proteins. Complementary v-[SNARE](#) and t-SNARE proteins form stable trans-SNARE complexes, thereby bringing their membrane bilayers into close apposition for fusion.

Transport from the ER through the Golgi Apparatus

Newly synthesized proteins enter the biosynthetic- secretory pathway in the [ER](#) by crossing the ER [membrane](#) from the [cytosol](#). During their subsequent transport, from the ER to the Golgi apparatus and from the Golgi apparatus to the cell surface and elsewhere, these proteins pass through a series of compartments, where they are successively modified. Transfer from one [compartment](#) to the next involves a delicate balance between forward and backward (retrieval) transport pathways. Some transport vesicles select cargo molecules and move them to the next compartment in the pathway, while others retrieve escaped proteins and return them to a previous compartment where they normally function. Thus, the pathway from the ER to the cell surface involves many sorting steps, which continually select membrane and soluble lumenal proteins for packaging and transport—in vesicles or [organelle](#) fragments that bud from the ER and Golgi apparatus.

In this [section](#) we focus mainly on the [Golgi apparatus](#) (also called the **Golgi [complex](#)**). It is a major site of [carbohydrate](#) synthesis, as well as a sorting and dispatching station for the products of the [ER](#). Many of the cell's polysaccharides are made in the Golgi apparatus, including the pectin and hemicellulose of the [cell wall](#) in plants and most of the glycosaminoglycans of the [extracellular matrix](#) in animals. But the Golgi apparatus also lies on the exit route from the ER, and a large proportion of the carbohydrates that it makes are attached as [oligosaccharide](#) side chains to the many proteins and lipids that the ER sends to it. A subset of these oligosaccharide groups serve as tags to direct specific proteins into vesicles that then transport them to lysosomes. But most proteins and lipids, once they have acquired their appropriate oligosaccharides in the Golgi apparatus, are recognized in other ways for targeting into the transport vesicles going to other destinations.

Proteins Leave the ER in COPII-coated Transport Vesicles

To initiate their journey along the biosynthetic-secretory pathway, proteins that have entered the [ER](#) and are destined for the Golgi apparatus or beyond are first packaged into small COPII-coated transport vesicles. These transport vesicles bud from specialized regions of the ER called *ER exit sites*, whose [membrane](#) lacks bound ribosomes. In most animal cells, ER exit sites seem to be randomly dispersed throughout the ER network.

Originally it was thought that all proteins that are not tethered in the [ER](#) enter transport vesicles by default. However, it is now clear that packaging into vesicles that leave the ER can also be a selective process. Some cargo proteins are actively recruited into such vesicles, where they become concentrated. It is thought that these cargo proteins display exit (transport) signals on their surface that are recognized by [complementary receptor](#) proteins that become trapped in the budding [vesicle](#) by interacting with components of the COPII coat. At a much lower rate, proteins without such exit signals can also get packaged in vesicles, so that even proteins that normally function in the ER (so-called *ER resident proteins*) slowly leak out of the ER. Similarly, secretory proteins that are made in high concentrations may leave the ER without the help of sorting receptors.

The exit signals that direct proteins out of the [ER](#) for transport to the Golgi and beyond are mostly not understood. There is one exception, however. The ERGIC53 [protein](#) seems to serve as a [receptor](#) for packaging some secretory proteins into COPII-coated vesicles. Its role in protein transport was identified because humans who lack it owing to an inherited [mutation](#) have lowered serum levels of two secreted blood-clotting factors (Factor V and Factor VIII) and therefore bleed excessively. The ERGIC53 protein is a [lectin](#) that binds mannose and is thought to recognize this [sugar](#) on Factor V and Factor VIII proteins, thereby packaging the proteins into transport vesicles in the ER.

Only Proteins That Are Properly Folded and Assembled Can Leave the ER

To exit from the [ER](#), proteins must be properly folded and, if they are subunits of multimeric [protein](#) complexes, they may need to be completely assembled. Those that are misfolded or incompletely assembled are retained in the ER, where they are bound to chaperone proteins (see Chapter 6), such as *BiP* or *calnexin*. The chaperones may cover up the exit signals or somehow anchor the proteins in the ER ([Figure 13-18](#)). Such failed proteins are eventually transported back into the [cytosol](#) where they are degraded by proteasomes (discussed in Chapter 12). This quality-control step is important, as misfolded or misassembled proteins could potentially interfere with the functions of normal proteins if they were transported onward. The amount of corrective action is surprisingly large. More than 90% of the newly synthesized subunits of the T cell [receptor](#) and of the [acetylcholine receptor](#) (discussed in Chapter 11), for example, are normally degraded in the cell without ever reaching the cell surface, where they function. Thus, cells must make a large excess of many protein molecules from which to select the few that fold and assemble properly.

Sometimes, however, this quality-control mechanism is detrimental. The predominant mutations that cause cystic fibrosis, a common inherited disease, produce a [plasma membrane protein](#) important for Cl⁻ transport that is only slightly misfolded. Although the [mutant](#) protein would function perfectly normally if it reached the plasma membrane, it is retained in the [ER](#). The devastating disease thus results not because the [mutation](#) inactivates the protein, but because the active protein is discarded before it reaches the plasma membrane.

Transport from the ER to the Golgi Apparatus Is Mediated by Vesicular Tubular Clusters

After transport vesicles have budded from an [ER](#) exit site and have shed their coat, they begin to fuse with one another. This fusion of membranes from the same [compartment](#) is called *homotypic fusion*, to distinguish it from *heterotypic fusion*, in which a [membrane](#) from one compartment fuses with the membrane of a different compartment. As with heterotypic fusion, homotypic fusion requires a set of matching [SNAREs](#). In this case, however, the interaction is symmetrical, with v-SNAREs and t-SNAREs contributed by both membranes.

The structures formed when [ER](#)-derived vesicles fuse with one another are called **vesicular tubular clusters**, on the basis of their convoluted appearance in the [electron](#)

[microscope](#) (Figure 13-20A). These clusters constitute a new [compartment](#) that is separate from the ER and lacks many of the proteins that function in the ER. They are generated continually and function as transport packages that bring material from the ER to the Golgi apparatus. The clusters are relatively short-lived because they quickly move along microtubules to the Golgi apparatus, where they fuse and deliver their contents.

As soon as **vesicular** tubular clusters form, they begin budding off vesicles of their own. Unlike the COPII-coated vesicles that bud from the [ER](#), these vesicles are COPI-coated. They carry back to the ER resident proteins that have escaped, as well as proteins that participated in the ER budding [reaction](#) and are being returned. This retrieval process demonstrates the exquisite control mechanisms that regulate coat assembly reactions. The COPI coat assembly begins only seconds after the COPII coats have been shed. It remains a mystery how this switchover in coat assembly is controlled.

The *retrieval* (or *retrograde*) *transport* continues as the **vesicular** tubular clusters move to the Golgi apparatus. Thus, the clusters continuously mature, gradually changing their composition as selected proteins are returned to the [ER](#). A similar retrieval process continues from the Golgi apparatus, after the **vesicular** tubular clusters have delivered their cargo.

The Retrieval Pathway to the ER Uses Sorting Signals

The retrieval pathway for returning escaped proteins back to the [ER](#) depends on *ER retrieval signals*. Resident ER [membrane](#) proteins, for example, contain signals that bind directly to COPI coats and are thus packaged into COPI-coated transport vesicles for retrograde delivery to the ER. The best-characterized signal of this type consists of two lysines, followed by any two other amino acids, at the extreme C-terminal end of the ER [membrane protein](#). It is called a *aKKXX sequence*, based on the single-letter [amino acid](#) code.

Soluble [ER](#) resident proteins, such as BiP, also contain a short retrieval signal at their C-terminal end, but it is different: it consists of a Lys-Asp-Glu-Leu or similar sequence. If this signal (called the *KDEL sequence*) is removed from BiP by genetic engineering, the [protein](#) is slowly secreted from the cell. If the signal is transferred to a protein that is normally secreted, the protein is now efficiently returned to the ER, where it accumulates.

Unlike the retrieval signals on [ER membrane](#) proteins that can interact directly with the COPI coat, soluble ER resident proteins must bind to specialized [receptor](#) proteins such as the *KDEL receptor*—a [multipass transmembrane protein](#) that binds to the KDEL sequence and packages any protein displaying it into COPI-coated retrograde transport vesicles. To accomplish this task, the KDEL receptor itself must cycle between the ER and the Golgi apparatus, and its affinity for the KDEL sequence must be different in these two compartments. The receptor must have a high affinity for the KDEL sequence in **vesicular** tubular clusters and the Golgi apparatus, so as to capture escaped ER resident proteins that are present there at low concentration. It must have a low affinity for the KDEL sequence in the ER, however, to unload its cargo in spite of the very high concentration of KDEL-containing resident proteins in the ER.

How can the affinity of the KDEL [receptor](#) change depending on the [compartment](#) in which it resides? The answer may be related to the different [pH](#) values established in the different compartments, regulated by H^+ pumps in their [membrane](#). The KDEL receptor could bind the KDEL sequence under the slightly acidic conditions in **vesicular** tubular clusters and the Golgi compartment but release it at the neutral pH in the [ER](#). As we discuss later, such pH-sensitive [protein](#)-protein interactions form the basis for many of the sorting steps in the cell.

Most [membrane](#) proteins that function at the interface between the [ER](#) and Golgi apparatus, including v- and t-SNARES and some cargo receptors, enter the retrieval pathway to the ER. Whereas the recycling of some of these proteins is signal-mediated as just described, for others no specific signal seems to be required. Thus, while retrieval signals increase the efficiency of the retrieval process, some proteins—including some Golgi enzymes—randomly enter budding vesicles destined for the ER and are returned to the ER at a slower rate. Such Golgi enzymes cycle constantly between the ER and the Golgi, but their rate of return to the ER is slow enough for most of the [protein](#) to be found in the Golgi apparatus.

Many Proteins are Selectively Retained in the Compartments in which they Function

The KDEL retrieval pathway only partly explains how [ER](#) resident proteins are maintained in the ER. As expected, cells that express genetically modified ER resident proteins, from which the KDEL sequence has been experimentally removed, secrete these proteins. But secretion occurs at a much slower rate than for a normal secretory [protein](#). It seems that ER resident proteins are anchored in the ER by a mechanism that is independent of their KDEL signal and that only those proteins that escape retention are captured and returned via the KDEL [receptor](#). A suggested mechanism of retention is that ER resident proteins bind to one another, thus forming complexes that are too big to enter transport vesicles. Because ER resident proteins are present in the ER at very high concentrations (estimated to be millimolar), relatively low-affinity interactions would suffice to have most of the proteins tied up in such complexes.

Aggregation of proteins that function in the same [compartment](#)—called *kin recognition*—is a general mechanism that compartments use to organize and retain their resident proteins. Golgi enzymes that function together, for example, also bind to each other and are thereby restrained from entering transport vesicles.

The Length of the Transmembrane Region of Golgi Enzymes Determines their Location in The Cell

Vesicles that leave the Golgi apparatus of animal cells destined for the [plasma membrane](#) are rich in [cholesterol](#). The cholesterol fills the space between the kinked [hydrocarbon](#) chains of the lipids in the bilayer, forcing them into tighter alignment and increasing the separation between the [lipid](#) head groups of the two leaflets of the bilayer. Thus the [lipid bilayer](#) of the cholesterol-derived vesicles is thicker than that of the Golgi membrane itself. Transmembrane proteins must have sufficiently long transmembrane segments to span this thickness if they

are to enter the cholesterol-rich transport [vesicle](#) budding from the Golgi apparatus destined for the plasma membrane. Proteins with shorter transmembrane segments are excluded.

This exclusion is thought to explain why [membrane](#) proteins that normally reside in the Golgi and the [ER](#) have shorter transmembrane segments (around 15 amino acids) than do [plasma membrane](#) proteins (around 20–25 amino acids). When the transmembrane segments of Golgi proteins are extended by [recombinant DNA](#) techniques, the proteins are no longer efficiently retained in the Golgi apparatus and are transported to the plasma membrane instead. Thus, at least some Golgi proteins seem to be retained in the Golgi apparatus mainly because they cannot enter transport vesicles heading for the plasma membrane.

The Golgi Apparatus Consists of an Ordered Series of Compartments

Because of its large and regular structure, the Golgi apparatus was one of the first organelles described by early light microscopists. It consists of a collection of flattened, [membrane](#)-enclosed *cisternae*, somewhat resembling a stack of pancakes. Each of these Golgi stacks usually consists of four to six cisternae, although some unicellular flagellates can have up to 60. In animal cells, many stacks are linked by tubular connections between corresponding cisternae, thus forming a single [complex](#), which is usually located near the cell [nucleus](#) and close to the [centrosome](#). This localization depends on microtubules. If microtubules are experimentally depolymerized, the Golgi apparatus reorganizes into individual stacks that are found throughout the [cytoplasm](#), adjacent to [ER](#) exit sites. In some cells, including most plant cells, hundreds of individual Golgi stacks are normally dispersed throughout the [cytoplasm](#).

During their passage through the Golgi apparatus, transported molecules undergo an ordered series of covalent modifications. Each Golgi stack has two distinct faces: a ***cis* face** (or entry face) and a ***trans* face** (or exit face). Both *cis* and *trans* faces are closely associated with special compartments, each composed of a network of interconnected tubular and cisternal structures: the ***cis* Golgi network (CGN)** (also called the *intermediate compartment*) and the ***trans* Golgi network (TGN)**, respectively. Proteins and lipids enter the *cis* Golgi network in **vesicular** tubular clusters arriving from the [ER](#) and exit from the *trans* Golgi network bound for the cell surface or another compartment. Both networks are thought to be important for [protein](#) sorting. As we have seen, proteins entering the CGN can either move onward in the Golgi apparatus or be returned to the ER. Similarly, proteins exiting from the TGN can either move onward and be sorted according to whether they are destined for lysosomes, secretory vesicles, or the cell surface, or be returned to an earlier compartment.

The Golgi apparatus is especially prominent in cells that are specialized for secretion, such as the goblet cells of the intestinal epithelium, which secrete large amounts of [polysaccharide](#)-rich mucus into the gut. In such cells, unusually large vesicles are found on the *trans* side of the Golgi apparatus, which faces the [plasma membrane domain](#) where secretion occurs.

Oligosaccharide Chains Are Processed in the Golgi Apparatus

As described in Chapter 12, a single species of *N*-linked **oligosaccharide** is attached *en bloc* to many proteins in the **ER** and then trimmed while the **protein** is still in the ER. Further modifications and additions occur in the Golgi apparatus, depending on the protein. The outcome is that two broad classes of *N*-linked oligosaccharides, the **complex oligosaccharides** and the **high-mannose oligosaccharides**, are found attached to mammalian glycoproteins (**Figure 13-25**). Sometimes both types are attached (in different places) to the same **polypeptide** chain.

Complex oligosaccharides are generated by a combination of trimming the original *N*-linked **oligosaccharide** added in the **ER** and the addition of further sugars. By contrast, **high-mannose oligosaccharides** have no new sugars added to them in the Golgi apparatus. They contain just two *N*-acetylglucosamines and many mannose residues, often approaching the number originally present in the **lipid**-linked oligosaccharide precursor added in the ER. Complex oligosaccharides can contain more than the original two *N*-acetylglucosamines as well as a variable number of galactose and sialic **acid** residues and, in some cases, fucose. Sialic acid is of special importance because it is the only **sugar** in glycoproteins that bears a negative charge. Whether a given oligosaccharide remains high-mannose or is processed is **determined** largely by its position on the **protein**. If the oligosaccharide is accessible to the processing enzymes in the Golgi apparatus, it is likely to be converted to a **complex** form; if it is inaccessible because its sugars are tightly held to the protein's surface, it is likely to remain in a high-mannose form. The processing that generates **complex oligosaccharide** chains follows the highly ordered pathway shown.

Proteoglycans Are Assembled in the Golgi Apparatus

It is not only the *N*-linked **oligosaccharide** chains on proteins that are altered as the proteins pass through the Golgi cisternae *en route* from the **ER** to their final destinations; many proteins are also modified in other ways. Some proteins have sugars added to the **OH** groups of selected serine or threonine side chains. This **O-linked glycosylation**, like the extension of *N*-linked oligosaccharide chains, is catalyzed by a series of glycosyl transferase enzymes that use the **sugar** nucleotides in the **lumen** of the Golgi apparatus to add sugar residues to a **protein** one at a time. Usually, *N*-acetylgalactosamine is added first, followed by a variable number of additional sugar residues, ranging from just a few to 10 or more.

The Golgi apparatus confers the heaviest **glycosylation** of all on **proteoglycan core proteins**, which it modifies to produce **proteoglycans**. As discussed in Chapter 19, this process involves the polymerization of one or more **glycosaminoglycan chains** (long unbranched polymers composed of repeating **disaccharide** units) via a xylose link onto serines on the core **protein**. Many proteoglycans are secreted and become components of the **extracellular matrix**, while others remain anchored to the **plasma membrane**. Still others form a major component of slimy materials, such as the mucus that is secreted to form a protective coating over many epithelia.

The sugars incorporated into glycosaminoglycans are heavily sulfated in the Golgi apparatus immediately after these polymers are made, thus adding a significant portion of their characteristically large negative charge. Some tyrosine residues in proteins also become sulfated shortly before they exit from the Golgi apparatus. In both cases, the sulfation depends on the sulfate donor 3'-phosphoadenosine-5'-phosphosulfate, or PAPS, that is transported from the [cytosol](#) into the [lumen](#) of the *trans* Golgi network.

What Is the Purpose of Glycosylation?

There is an important difference between the construction of an [oligosaccharide](#) and the synthesis of other macromolecules such as [DNA](#), [RNA](#), and [protein](#). Whereas nucleic acids and proteins are copied from a [template](#) in a repeated series of identical steps using the same [enzyme](#) or set of enzymes, [complex](#) carbohydrates require a different enzyme at each step, each product being recognized as the exclusive [substrate](#) for the next enzyme in the series. Given the complicated pathways that have evolved to synthesize them, it seems likely that the oligosaccharides on glycoproteins and glycosphingolipids have important functions, but for the most part these functions are not known.

N-linked [glycosylation](#), for example, is prevalent in all eucaryotes, including yeasts, but is absent from procaryotes. Because one or more *N*-linked oligosaccharides are present on most proteins transported through the [ER](#) and Golgi apparatus—a pathway that is unique to eucaryotic cells—one might suspect that they function to aid folding and the transport process. We have already discussed a number of instances for which this is so—the use of a [carbohydrate](#) as a marker during [protein](#) folding in the ER, for example, and the use of carbohydrate-binding lectins in guiding ER-to-Golgi transport. As we discuss later, lectins also participate in protein sorting in the *trans* Golgi network.

Because chains of sugars have limited flexibility, even a small *N*-linked [oligosaccharide](#) protrudes from the surface of a [glycoprotein](#) and can thus limit the approach of other macromolecules to the [protein](#) surface. In this way, for example, the presence of oligosaccharides tends to make a glycoprotein more resistant to digestion by proteases. It may be that the oligosaccharides on cell-surface proteins originally provided an ancestral eucaryotic cell with a protective coat that, unlike the rigid bacterial [cell wall](#), left the cell with the freedom to change shape and move. But if so, these [sugar](#) chains have since become modified to serve other purposes as well. The oligosaccharides attached to some cell-surface proteins, for example, are recognized by transmembrane lectins called *selectins*, which function in cell-cell adhesion processes.

Glycosylation can also have important regulatory roles. Signaling through the cell-surface signaling [receptor Notch](#), for example, is important for proper [cell fate](#) determination in [development](#). Notch is a [transmembrane protein](#) that is *O*-glycosylated by addition of a single fucose to some serines, threonines, and hydroxylysines. Some cell types express an additional glycosyltransferase that adds an *N*-acetylglucosamine to each of these fucoses in the Golgi apparatus. This addition sensitizes the Notch receptor, and thus allows these cells to

respond selectively to activating stimuli. In this way, [glycosylation](#) has become important to the establishment of spatial boundaries in developing tissues.

The Golgi Cisternae Are Organized as a Series of Processing Compartments

Proteins exported from the [ER](#) enter the first of the Golgi processing compartments (the *cis* Golgi [compartment](#)), after having passed through the *cis* Golgi network. They then move to the next compartment (the *medial* compartment, consisting of the central cisternae of the stack) and finally to the *trans* compartment, where [glycosylation](#) is completed.

The [lumen](#) of the *trans* compartment is thought to be continuous with the *trans* Golgi network, where proteins are segregated into different transport packages and dispatched to their final destinations—the [plasma membrane](#), lysosomes, or secretory vesicles.

The [oligosaccharide](#) processing steps occur in a correspondingly organized sequence in the Golgi stack, with each cisterna containing a characteristic abundance of processing enzymes. Proteins are modified in successive stages as they move from cisterna to cisterna across the stack, so that the stack forms a multistage processing unit. This compartmentalization might seem unnecessary, since each oligosaccharide processing [enzyme](#) can accept a [glycoprotein](#) as a [substrate](#) only after it has been properly processed by the preceding enzyme. Nonetheless, it is clear that processing occurs in a spatial as well as a biochemical sequence: enzymes catalyzing early processing steps are concentrated in the cisternae toward the *cis* face of the Golgi stack, whereas enzymes catalyzing later processing steps are concentrated in the cisternae toward the *trans* face.

The functional differences between the *cis*, *medial*, and *trans* subdivisions of the Golgi apparatus were discovered by localizing the enzymes involved in processing *N*-linked oligosaccharides in distinct regions of the [organelle](#), both by physical fractionation of the organelle and by labeling the enzymes in [electron microscope](#) sections with antibodies. The removal of mannose residues and the addition of *N*-acetylglucosamine, for example, were shown to occur in the *medial* [compartment](#), while the addition of galactose and sialic [acid](#) was found to occur in the *trans* compartment and the *trans* Golgi network.

The functional and structural divisions of the Golgi stack pose two important questions. How are molecules transported from one Golgi cisterna to the next, and how are Golgi resident proteins retained in their appropriate places?

Transport Through the Golgi Apparatus May Occur by Vesicular Transport or Cisternal Maturation

It is still uncertain how the Golgi apparatus achieves and maintains its polarized structure and how molecules move from one cisterna to another. Functional evidence from [in vitro](#) transport assays and the finding of abundant transport vesicles in the vicinity of Golgi cisternae initially led to the view that these vesicles transport proteins between the cisternae, budding from one cisterna and fusing with the next. According to this [vesicular transport model](#), the Golgi apparatus is a relatively static structure, with its

enzymes held in place, while the molecules in transit are moved through the cisternae in sequence, carried by transport vesicles. Retrograde flow retrieves escaped [ER](#) and Golgi proteins and returns them to preceding compartments. Directional flow is achieved as forward-moving cargo molecules are selectively packaged into forward-moving vesicles, whereas proteins to be retrieved are selectively packaged into retrograde vesicles. Although both types of vesicles are likely to be COPI-coated, the coats may contain different adaptor proteins to confer selectivity on the packaging of cargo molecules. Alternatively, transport vesicles that shuttle between Golgi cisternae may not be directional at all, transporting cargo material randomly back and forth; directional flow would then occur because of the continual input at the *cis* cisterna and output at the *trans* cisterna. In either case, the movement of vesicles from each cisterna to an adjacent one is helped by a neat trick: the budding vesicles remain tethered by filamentous proteins that restrict their movement, so that their fusion with the correct target [membrane](#) is facilitated.

According to an alternative hypothesis, called the **cisternal maturation model**, the Golgi is viewed as a dynamic structure in which the cisternae themselves move through the Golgi stack. The **vesicular** tubular clusters that arrive from the [ER](#) fuse with one another to become a *cis* Golgi network, and this network then progressively matures to become *acis* cisterna, then a *medial* cisterna, and so on. Thus, at the *cis* face of a Golgi stack, new *cis* cisternae would continually form and then migrate through the stack as they mature. This model is supported by microscopic observations demonstrating that large structures such as [collagen](#) rods in fibroblasts and scales in certain algae—which are much too large to fit into classical transport vesicles—move progressively through the Golgi stack.

In the maturation model, the characteristic distribution of Golgi enzymes is explained by retrograde flow. Everything moves continuously forward with the maturing cisterna, including the processing enzymes that belong in the early Golgi apparatus. But budding COPI-coated vesicles continually collect the appropriate enzymes, almost all of which are [membrane](#) proteins, and carry them back to the earlier cisterna where they function. A newly formed *cis* cisterna would therefore receive its normal complement of resident enzymes primarily from the cisterna just ahead of it and would later pass them back to the next *cis* cisterna that forms.

As we discuss later, when a cisterna finally moves up to become part of the *trans* Golgi network, various types of coated vesicles bud off of it until this network disappears, to be replaced by a maturing cisterna just behind. At the same time, other transport vesicles are continually retrieving [membrane](#) from post-Golgi compartments and returning this membrane to the *trans* Golgi network.

The [vesicular transport](#) and the cisternal maturation model are not mutually exclusive. Indeed, evidence suggests that transport may occur by a combination of the two mechanisms, in which some cargo is moved forward rapidly in transport vesicles, whereas other cargo is moved forward more slowly as the Golgi apparatus constantly renews itself through cisternal maturation.

Matrix Proteins Form a Dynamic Scaffold That Helps Organize the Apparatus

The unique architecture of the Golgi apparatus depends on both the [microtubule cytoskeleton](#), as already discussed, and cytoplasmic Golgi matrix proteins, which form a scaffold between adjacent cisternae and give the Golgi stack its structural integrity. Some of the matrix proteins form long, filamentous tethers that are thought to help retain Golgi transport vesicles close to the [organelle](#). When the cell prepares to divide, mitotic [protein](#) kinases phosphorylate the Golgi matrix proteins, causing the Golgi apparatus to fragment and disperse throughout the [cytosol](#). During disassembly, Golgi enzymes are returned in vesicles to the [ER](#), while other Golgi fragments are distributed to the two daughter cells. There, the matrix proteins are dephosphorylated, leading to the reassembly of the Golgi apparatus.

Remarkably, the Golgi matrix proteins can assemble into appropriately localized stacks near the [centrosome](#) even when Golgi [membrane](#) proteins are experimentally prevented from leaving the [ER](#). This observation suggests that the matrix proteins are largely responsible for both the structure and location of the Golgi apparatus.

Summary

Correctly folded and assembled proteins in the [ER](#) are packaged into COPII-coated transport vesicles that pinch off from the ER [membrane](#). Shortly thereafter the coat is shed and the vesicles fuse with one another to form **vesicular** tubular clusters, which move on [microtubule](#) tracks to the Golgi apparatus. Many resident ER proteins slowly escape, but they are returned to the ER from the **vesicular** tubular clusters and the Golgi apparatus by retrograde transport in COPI-coated vesicles.

The Golgi apparatus, unlike the [ER](#), contains many [sugar](#) nucleotides, which are used by a variety of glycosyl transferase enzymes to perform [glycosylation](#) reactions on [lipid](#) and [protein](#) molecules as they pass through the Golgi apparatus. The N-linked oligosaccharides that are added to proteins in the ER are often initially trimmed by the removal of mannoses, and further sugars are added. Moreover, the Golgi is the site where O-linked glycosylation occurs and where glycosaminoglycan chains are added to core proteins to form proteoglycans. Sulfation of the sugars in proteoglycans and of selected tyrosines on proteins also occurs in a late Golgi [compartment](#).

The Golgi apparatus distributes the many proteins and lipids that it receives from the [ER](#) and then modifies the [plasma membrane](#), lysosomes, and secretory vesicles. It is a polarized structure consisting of one or more stacks of disc-shaped cisternae, each stack organized as a series of at least three functionally distinct compartments, termed cis, medial, and trans cisternae. The cis and trans cisternae are both connected to special sorting stations, called the [cis Golgi network](#) and the [trans Golgi network](#), respectively. Proteins and lipids move through the Golgi stack in the cis-to-trans direction. This movement may occur by [vesicular transport](#), by progressive maturation of the cis cisternae that migrate continuously through the stack, or by a combination of these two mechanisms. The enzymes that function in each particular region of the stack are thought to be kept there by continual

retrograde **vesicular** transport from more distal cisternae. The finished new proteins end up in the trans Golgi network, which packages them in transport vesicles and dispatches them to their specific destinations in the cell.

Transport into the Cell from the Plasma Membrane: Endocytosis

The routes that lead inward from the cell surface to lysosomes start with the process of [endocytosis](#), by which cells take up macromolecules, particulate substances, and, in specialized cases, even other cells. In this process, the material to be ingested is progressively enclosed by a small portion of the [plasma membrane](#), which first invaginates and then pinches off to form an *endocytic vesicle* containing the ingested substance or particle. Two main types of [endocytosis](#) are distinguished on the basis of the size of the endocytic vesicles formed. One type is called [phagocytosis](#) (“cellular eating”), which involves the ingestion of large particles, such as microorganisms or dead cells via large vesicles called *phagosomes* (generally >250 [nm](#) in diameter). The other type is [pinocytosis](#) (“cellular drinking”), which involves the ingestion of fluid and solutes via small pinocytic vesicles (about 100 nm in diameter). Most eucaryotic cells are continually ingesting fluid and solutes by pinocytosis; large particles are most efficiently ingested by specialized phagocytic cells.

Specialized Phagocytic Cells Can Ingest Large Particles

[Phagocytosis](#) is a special form of [endocytosis](#) in which large particles such as microorganisms and dead cells are ingested via large endocytic vesicles called [phagosomes](#). In [protozoa](#), [phagocytosis](#) is a form of feeding: large particles taken up into phagosomes end up in lysosomes, and the products of the subsequent digestive processes pass into the [cytosol](#) to be utilized as food. However, few cells in multicellular organisms are able to ingest such large particles efficiently. In the gut of animals, for example, the particles of food are broken down extracellularly and their hydrolysis products are imported into cells.

Phagocytosis is important in most animals for purposes other than nutrition, and it is mainly carried out by specialized cells—so-called *professional phagocytes*. In mammals, three classes of white blood cells act as professional phagocytes—[macrophages](#), [neutrophils](#), and [dendritic cells](#). These cells all develop from hemopoietic stem cells (discussed in Chapter 22), and they defend us against infection by ingesting invading microorganisms. Macrophages also have an important role in scavenging senescent cells and cells that have died by [apoptosis](#). In quantitative terms, the latter function is by far the most important: our macrophages phagocytose more than 10^{11} senescent red blood cells in each of us every day, for example.

Whereas the endocytic vesicles involved in [pinocytosis](#) are small and uniform, [phagosomes](#) have diameters that are [determined](#) by the size of the ingested particle, and they can be almost as large as the phagocytic cell itself. The phagosomes fuse with lysosomes inside the cell, and the ingested material is then degraded. Any indigestible substances will remain in lysosomes, forming *residual bodies*. Some of the

internalized [plasma membrane](#) components never reach the [lysosome](#), because they are retrieved from the [phagosome](#) in transport vesicles and returned to the plasma membrane.

To be phagocytosed, particles must first bind to the surface of the [phagocyte](#). However, not all particles that bind are ingested. Phagocytes have a variety of specialized surface receptors that are functionally linked to the phagocytic machinery of the cell. Unlike [pinocytosis](#), which is a [constitutive](#) process that occurs continuously, [phagocytosis](#) is a triggered process, requiring that receptors be activated that transmit signals to the cell interior and initiate the response. The best-characterized triggers are antibodies, which protect us by binding to the surface of infectious microorganisms to form a coat in which the tail region of each antibody [molecule](#), called the Fc region, is exposed on the exterior (discussed in Chapter 24). This antibody coat is recognized by specific *Fc receptors* on the surface of macrophages and neutrophils, whose binding induces the phagocytic cell to extend pseudopods that engulf the particle and fuse at their tips to form a [phagosome](#).

Several other classes of receptors that promote [phagocytosis](#) have been characterized. Some recognize *complement* components, which collaborate with antibodies in targeting microbes for destruction (discussed in Chapter 25). Others directly recognize oligosaccharides on the surface of certain microorganisms. Still others recognize cells that have died by [apoptosis](#). Apoptotic cells lose the asymmetric distribution of phospholipids in their [plasma membrane](#). As a consequence, negatively charged phosphatidylserine, which is normally confined to the cytosolic leaflet of the [lipid bilayer](#), is now exposed on the outside of the cell, where it triggers the phagocytosis of the dead cell.

Remarkably, macrophages will also phagocytose a variety of inanimate particles—such as glass, latex beads, or asbestos fibers—yet they do not phagocytose live animal cells. It seems that living animal cells display “don’t-eat-me” signals in the form of cell-surface proteins that bind to inhibiting receptors on the surface of macrophages. The inhibitory receptors recruit tyrosine phosphatases that antagonize the **intracellular** signaling events required to initiate [phagocytosis](#), thereby locally inhibiting the phagocytic process. Thus phagocytosis, like many other cell processes, depends on a balance between positive signals that activate the process and negative signals that inhibit it.

Pinocytic Vesicles Form from Coated Pits in the Plasma Membrane

Virtually all eucaryotic cells continually ingest bits of their [plasma membrane](#) in the form of small pinocytic (endocytic) vesicles, which are later returned to the cell surface. The rate at which plasma membrane is internalized in this process of [pinocytosis](#) varies between cell types, but it is usually surprisingly large. A [macrophage](#), for example, ingests 25% of its own volume of fluid each hour. This means that it must ingest 3% of its plasma membrane each minute, or 100% in about half an hour. Fibroblasts endocytose at a somewhat lower rate (1% per minute), whereas some amoebae ingest their plasma membrane even more rapidly. Since a cell's surface area and volume remain unchanged during this process, it is clear that the same amount of membrane that is being removed by [endocytosis](#) is being added to the cell

surface by [exocytosis](#), the converse process, as we discuss later. In this sense, endocytosis and exocytosis are linked processes that can be considered to constitute an [endocytic-exocytic cycle](#).

The endocytic part of the cycle often begins at [clathrin-coated pits](#). These specialized regions typically occupy about 2% of the total [plasma membrane](#) area. The lifetime of a [clathrin-coated pit](#) is short: within a minute or so of being formed, it invaginates into the cell and pinches off to form a [clathrin-coated vesicle](#) (Figure 13-41). It has been estimated that about 2500 clathrin-coated vesicles leave the plasma membrane of a cultured [fibroblast](#) every minute. The coated vesicles are even more transient than the coated pits: within seconds of being formed, they shed their coat and are able to fuse with early endosomes. Since extracellular fluid is trapped in clathrin-coated pits as they invaginate to form coated vesicles, any substance dissolved in the extracellular fluid is internalized—a process called [fluid-phase endocytosis](#).

Not All Pinocytic Vesicles Are Clathrin-coated

In addition to [clathrin](#)-coated pits and vesicles, there are other, less well-understood mechanisms by which cells can form pinocytic vesicles. One of these pathways initiates at **caveolae** (from the Latin for “little cavities”), originally recognized by their ability to transport molecules across endothelial cells, which form the inner lining of blood vessels. Caveolae are present in the [plasma membrane](#) of most cell types, and in some of these they are seen as deeply invaginated flasks in the [electron microscope](#). They are thought to form from [lipid rafts](#), which are patches of the plasma membrane that are especially rich in [cholesterol](#), glycosphingolipids, and GPI-anchored membrane proteins. The major structural [protein](#) in caveolae is **caveolin**, a multipass [integral membrane protein](#) that is a member of a heterogeneous protein family.

In contrast to [clathrin](#)-coated and COPI- or COPII-coated vesicles, caveolae are thought to invaginate and collect cargo proteins by virtue of the [lipid](#) composition of the caveolar [membrane](#), rather than by the assembly of a cytosolic [protein](#) coat. Caveolae pinch off from the [plasma membrane](#) and can deliver their contents either to [endosome](#)-like compartments or (in a process called [transcytosis](#), which is discussed later) to the plasma membrane on the opposite side of a polarized cell. Some animal viruses also enter cells in vesicles derived from caveolae. The viruses are first delivered to an endosome-like [compartment](#), from where they are moved to the [ER](#). In the ER, they extrude their [genome](#) into the [cytosol](#) to start their infectious cycle. It remains a mystery how material endocytosed in caveolae-derived vesicles can end up in so many different locations in the cell.

Cells Import Selected Extracellular Macromolecules by Receptor-mediated Endocytosis

In most animal cells, [clathrin](#)-coated pits and vesicles provide an efficient pathway for taking up specific macromolecules from the extracellular fluid. In this process, called [receptor-](#)

mediated endocytosis, the macromolecules bind to complementary transmembrane receptor proteins, accumulate in coated pits, and then enter the cell as receptor-macromolecule complexes in clathrin-coated vesicles. Receptor-mediated endocytosis provides a selective concentrating mechanism that increases the efficiency of internalization of particular ligands more than a hundredfold, so that even minor components of the extracellular fluid can be specifically taken up in large amounts without taking in a correspondingly large volume of extracellular fluid. A particularly well-understood and physiologically important example is the process whereby mammalian cells take up cholesterol.

Many animals cells take up cholesterol through receptor-mediated endocytosis and, in this way, acquire most of the cholesterol they require to make new membrane. If the uptake is blocked, cholesterol accumulates in the blood and can contribute to the formation in blood vessel walls of *atherosclerotic plaques*, deposits of lipid and fibrous tissue that can cause strokes and heart attacks by blocking blood flow. In fact, it was through a study of humans with a strong genetic predisposition for *atherosclerosis* that the mechanism of receptor-mediated endocytosis was first clearly revealed.

Most cholesterol is transported in the blood as cholesteryl esters in the form of lipid-protein particles known as **low-density lipoproteins (LDL)**. When a cell needs cholesterol for membrane synthesis, it makes transmembrane receptor proteins for LDL and inserts them into its plasma membrane. Once in the plasma membrane, the LDL receptors diffuse until they associate with clathrin-coated pits that are in the process of forming. Since coated pits constantly pinch off to form coated vesicles, any LDL particles bound to LDL receptors in the coated pits are rapidly internalized in coated vesicles. After shedding their clathrin coats, the vesicles deliver their contents to early endosomes, which are located near the cell periphery. Once the LDL and LDL receptors encounter the low pH in the endosomes, LDL is released from its receptor and is delivered via late endosomes to lysosomes. There the cholesteryl esters in the LDL particles are hydrolyzed to free cholesterol, which is now available to the cell for new membrane synthesis. If too much free cholesterol accumulates in a cell, the cell shuts off both its own cholesterol synthesis and the synthesis of LDL receptor proteins, so that it ceases either to make or to take up cholesterol.

This regulated pathway for the uptake of cholesterol is disrupted in individuals who inherit defective genes encoding LDL receptor proteins. The resulting high levels of blood cholesterol predispose these individuals to develop atherosclerosis prematurely, and many die at an early age of heart attacks resulting from coronary artery disease. In some cases, the receptor is lacking altogether. In others, the receptors are defective—in either the extracellular binding site for LDL or the **intracellular** binding site that attaches the receptor to the coat of a clathrin-coated pit. In the latter case, normal numbers of LDL-binding receptor proteins are present, but they fail to become localized in the clathrin-coated regions of the plasma membrane. Although LDL binds to the surface of these mutant cells, it is not

internalized, directly demonstrating the importance of clathrin-coated pits in the [receptor-mediated endocytosis](#) of cholesterol.

More than 25 different receptors are known to participate in [receptor-mediated endocytosis](#) of different types of molecules, and they all apparently use the same [clathrin](#)-coated-pit pathway. Many of these receptors, like the [LDL](#) receptor, enter coated pits irrespective of whether they have bound their specific ligands. Others enter preferentially when bound to a specific [ligand](#), suggesting that a ligand-induced conformational change is required for them to activate the [signal sequence](#) that guides them into the pits. Since most [plasma membrane](#) proteins fail to become concentrated in clathrin-coated pits, the pits must function as molecular filters, preferentially collecting certain plasma membrane proteins (receptors) over others.

Signal peptides guide transmembrane proteins into [clathrin](#)-coated pits by binding to the adaptins. Despite a common function, their [amino acid](#) sequences vary. A common [endocytosis](#) signal consists of only four amino acids Y-X-X-Ψ, where Y is tyrosine, X any [polar](#) amino acid, and Ψ a hydrophobic amino acid. This short peptide, which is shared by many receptors, binds directly to one of the adaptins in clathrin-coated pits. By contrast, the cytosolic tail of the [LDL receptor](#) contains a unique signal (Asn-Pro-Val-Tyr) that apparently binds to the same [adaptin protein](#).

Electron-microscope studies of cultured cells exposed simultaneously to different labeled ligands demonstrate that many kinds of receptors can cluster in the same coated pit. The [plasma membrane](#) of one [clathrin-coated pit](#) can probably accommodate up to 1000 receptors of assorted varieties. Although all of the [receptor-ligand](#) complexes that use this endocytic pathway are apparently delivered to the same endosomal [compartment](#), the subsequent fates of the endocytosed molecules vary, as we discuss next.

Endocytosed Materials That Are Not Retrieved From Endosomes End Up in Lysosomes

The endosomal compartments of a cell can be [complex](#). They can be made visible in the [electron microscope](#) by adding a readily detectable [tracer molecule](#), such as the [enzyme](#) peroxidase, to the extracellular medium and leaving the cells for various lengths of time to take it up by [endocytosis](#). The distribution of the molecule after its uptake reveals the endosomal compartments as a set of heterogeneous, [membrane](#)-enclosed tubes extending from the periphery of the cell to the perinuclear region, where it is often close to the Golgi apparatus. Two sequential sets of endosomes can be readily distinguished in such labeling experiments. The tracer molecule appears within a minute or so in early endosomes, just beneath the [plasma membrane](#). After 5–15 minutes, it moves to late endosomes, close to the Golgi apparatus and near the [nucleus](#). Early and late endosomes differ in their [protein](#) compositions; they are associated with different Rab proteins, for example.

As mentioned earlier, the interior of the endosomal [compartment](#) is kept acidic ([pH](#) ~6) by a vacuolar [H⁺ ATPase](#) in the endosomal [membrane](#) that pumps H⁺ into the [lumen](#) from

the [cytosol](#). In general, later endosomes are more acidic than early endosomes. This acidic environment has a crucial role in the function of these organelles.

We have already seen how endocytosed materials that reach the late endosomes become mixed with newly synthesized [acid](#) hydrolases and end up being degraded in lysosomes. Many molecules, however, are specifically diverted from this journey to destruction. They are recycled instead from the early endosomes back to the [plasma membrane](#) via transport vesicles. Only molecules that are not retrieved from endosomes in this way are delivered to lysosomes for degradation.

Specific Proteins Are Removed from Early Endosomes and Returned to the Plasma Membrane

The **early endosomes** form a [compartment](#) that acts as the main sorting station in the endocytic pathway, just as the *cis* and *trans* Golgi networks serve this function in the biosynthetic-secretory pathway. In the acidic environment of the early [endosome](#), many internalized [receptor](#) proteins change their [conformation](#) and release their [ligand](#), just as the [M6P](#) receptors unload their cargo of [acid](#) hydrolases in the even more acidic late endosomes. Those endocytosed ligands that dissociate from their receptors in the early endosome are usually doomed to destruction in lysosomes, along with the other soluble contents of the endosome. Some other endocytosed ligands, however, remain bound to their receptors, and thereby share the fate of the receptors.

The fates of the [receptor](#) proteins—and of any ligands remaining bound to them—vary according to the specific type of receptor. (1) Most receptors are recycled and return to the same [plasma membrane domain](#) from which they came; (2) some proceed to a different domain of the plasma membrane, thereby mediating a process called [transcytosis](#); and (3) some progress to lysosomes, where they are degraded.

The [LDL receptor](#) follows the first pathway. It dissociates from its [ligand](#) LDL in the early [endosome](#) and is recycled to the [plasma membrane](#) for reuse, leaving the discharged LDL to be carried to lysosomes. The recycling vesicles bud from long, narrow tubules that extend from the early endosomes. It is likely that the geometry of these tubules helps the sorting process. Because tubules have a large membrane area enclosing a small volume, membrane proteins tend to accumulate there. Transport vesicles that return material to the plasma membrane begin budding from the tubules, but tubular portions of the early endosome also pinch off and fuse with one another to form [recycling endosomes](#), a way-station for the **traffic** between the early endosomes and the plasma membrane. During this process, the tubules and then the recycling endosome continuously shed vesicles that return to the plasma membrane.

The **transferrin receptor** follows a similar recycling pathway, but it also recycles its [ligand](#). Transferrin is a soluble [protein](#) that carries iron in the blood. Cell-surface transferrin receptors deliver transferrin with its bound iron to early endosomes by [receptor-mediated endocytosis](#). The low [pH](#) in the [endosome](#) induces transferrin to release its bound iron, but the iron-free

transferrin itself (called apotransferrin) remains bound to its receptor. The receptor-apotransferrin [complex](#) enters the tubular extensions of the early endosome and from there is recycled back to the [plasma membrane](#). When the apotransferrin returns to the neutral pH of the extracellular fluid, it dissociates from the receptor and is thereby freed to pick up more iron and begin the cycle again. Thus, transferrin shuttles back and forth between the extracellular fluid and the endosomal [compartment](#), avoiding lysosomes and delivering the iron that cells need to grow to the cell interior.

The second pathway that endocytosed receptors can follow from endosomes is taken both by opioid receptors (see [Figure 13-47](#)) and by the [receptor](#) that binds *epidermal growth factor* (EGF). EGF is a small, extracellular signal [protein](#) that stimulates epidermal and various other cells to divide. Unlike [LDL](#) receptors, EGF receptors accumulate in [clathrin](#)-coated pits only after binding EGF, and most of them do not recycle but are degraded in lysosomes, along with the ingested EGF. EGF binding therefore first activates **intracellular** signaling pathways and then leads to a decrease in the concentration of EGF receptors on the cell surface, a process called *receptor down-regulation* that reduces the cell's subsequent sensitivity to EGF (discussed in Chapter 15).

Multivesicular Bodies Form on the Pathway to Late Endosomes

It is still uncertain how endocytosed molecules move from the early to the late endosomal [compartment](#) so as to end up in lysosomes. A current view is that portions of the early endosomes migrate slowly along microtubules toward the cell interior, shedding tubules of material to be recycled to the [plasma membrane](#). The migrating endosomes enclose large amounts of invaginated membrane and internally pinched-off vesicles and are therefore called **multivesicular bodies**. It is unknown whether multivesicular bodies eventually fuse with a late endosomal compartment or if they fuse instead with each other to become late endosomes. At the end of this pathway, the late endosomes are converted to lysosomes as a result of their fusion with hydrolase-bearing transport vesicles from the *trans* Golgi network and their increased acidification.

The multivesicular bodies carry specific endocytosed [membrane](#) proteins that are to be degraded but exclude others that are to be recycled. As part of the [protein](#)-sorting process, specific proteins—for example, the occupied EGF [receptor](#) described previously—selectively partition to the invaginating membrane of the multivesicular bodies ([Figure 13-50](#)). In this way, the receptors, as well as any signaling proteins strongly bound to them, are rendered fully accessible to the digestive enzymes that will degrade them (see [Figure 13-50](#)).

Membrane proteins that are sorted into the [internal membrane](#) vesicles of a multivesicular body are first covalently modified with the small [protein ubiquitin](#). Unlike multi-ubiquitylation which typically targets [substrate](#) proteins for degradation in proteasomes (discussed in Chapter 6), ubiquitin tagging for sorting into the internal membrane vesicles of a multivesicular body requires the addition of only a single ubiquitin [molecule](#) that is added to activated receptors while still at the [plasma membrane](#). The ubiquitin tag facilitates the

uptake of the receptors into endocytic vesicles and is then recognized again by proteins that mediate the sorting process into the internal membrane vesicles of multivesicular bodies. In addition, membrane invagination in multivesicular bodies is regulated by a [lipid](#) kinase that phosphorylates [phosphatidylinositol](#). The phosphorylated head groups of these lipids are thought to serve as docking sites for the proteins that mediate the invagination process. Local modification of lipid molecules is thus another way in which specific membrane patches can be induced to change shape and destiny.

In addition to endocytosed [membrane](#) proteins, multivesicular bodies also contain most of the soluble content of early endosomes destined for digestion in lysosomes.

Macromolecules Can Be Transferred Across Epithelial Cell Sheets by Transcytosis

Some receptors on the surface of polarized epithelial cells transfer specific macromolecules from one extracellular space to another by [transcytosis](#). These receptors are endocytosed and then follow a pathway from endosomes to a different [plasma membrane](#). A newborn rat, for example, obtains antibodies from its mother's milk (which help protect it against infection) by transporting them across the epithelium of its gut. The [lumen](#) of the gut is acidic, and, at this low [pH](#), the antibodies in the milk bind to specific receptors on the [apical](#) (absorptive) surface of the gut epithelial cells. The [receptor](#)-antibody complexes are internalized via [clathrin](#)-coated pits and vesicles and are delivered to early endosomes. The complexes remain intact and are retrieved in transport vesicles that bud from the early [endosome](#) and subsequently fuse with the basolateral domain of the plasma membrane. On exposure to the neutral pH of the extracellular fluid that bathes the basolateral surface of the cells, the antibodies dissociate from their receptors and eventually enter the newborn's bloodstream.

The transcytotic pathway from the early [endosome](#) to the [plasma membrane](#) is not direct. The receptors first move from the early endosome to an intermediate endosomal [compartment](#), the **recycling endosome** described previously. The variety of pathways that different receptors follow from endosomes implies that, in addition to binding sites for their ligands and binding sites for coated pits, many receptors also possess sorting signals that guide them into the appropriate type of transport [vesicle](#) leaving the endosome and thereby to the appropriate target membrane in the cell.

A unique property of a [recycling endosomes](#) is that the exit of [membrane](#) proteins from the [compartment](#) can be regulated. Thus, cells can adjust the flux of proteins through the transcytotic pathway according to need. Although the mechanism of regulation is uncertain, it allows recycling endosomes an important role in adjusting the concentration of specific [plasma membrane](#) proteins. Fat cells and muscle cells, for example, contain large **intracellular** pools of the [glucose](#) transporters that are responsible for the uptake of glucose across the plasma membrane. These proteins are stored in specialized recycling endosomes until the cell is stimulated by the [hormone insulin](#) to increase its rate of glucose uptake. Then transport vesicles bud from the recycling [endosome](#) and deliver large numbers

of glucose transporters to the plasma membrane, thereby greatly increasing the rate of glucose uptake into the cell.

Epithelial Cells Have Two Distinct Early Endosomal Compartments But a Common Late Endosomal Compartment

In polarized epithelial cells, [endocytosis](#) occurs from both the *basolateral domain* and the *apical domain* of the [plasma membrane](#). Material endocytosed from either domain first enters an early endosomal [compartment](#) that is unique to that domain. This arrangement allows endocytosed receptors to be recycled back to their original membrane domain, unless they contain signals that mark them for [transcytosis](#) to the other domain. Molecules endocytosed from either plasma membrane domain that are not retrieved from the early endosomes end up in a common late endosomal compartment near the cell center and are eventually degraded in lysosomes.

Whether cells contain a few connected or many unconnected endosomal compartments seems to depend on the cell type and the physiological state of the cell. Like many other [membrane](#)-enclosed organelles, endosomes of the same type can readily fuse with one another (an example of homotypic fusion, discussed earlier) to create large continuous endosomes.

Summary

Cells ingest fluid, molecules, and particles by [endocytosis](#), in which localized regions of the [plasma membrane](#) invaginate and pinch off to form endocytic vesicles. Many of the endocytosed molecules and particles end up in lysosomes, where they are degraded. Endocytosis occurs both constitutively and as a triggered response to extracellular signals. Endocytosis is so extensive in many cells that a large fraction of the plasma membrane is internalized every hour. To make this possible, most of the plasma membrane components (proteins and [lipid](#)) that are endocytosed are continually returned to the cell surface by [exocytosis](#). This large-scale [endocytic-exocytic cycle](#) is mediated largely by [clathrin](#)-coated pits and vesicles.

Many cell-surface receptors that bind specific extracellular macromolecules become localized in [clathrin](#)-coated pits. As a result, they and their ligands are efficiently internalized in clathrin-coated vesicles, a process called [receptor-mediated endocytosis](#). The coated endocytic vesicles rapidly shed their clathrin coats and fuse with early endosomes.

Most of the ligands dissociate from their receptors in the acidic environment of the [endosome](#) and eventually end up in lysosomes, while most of the receptors are recycled via transport vesicles back to the cell surface for reuse. But [receptor-ligand](#) complexes can follow other pathways from the endosomal [compartment](#). In some cases, both the receptor and the ligand end up being degraded in lysosomes, resulting in receptor down-regulation. In other cases, both are transferred to a different [plasma membrane domain](#), and the ligand is thereby released by [exocytosis](#) at a surface of the cell different from that where it originated,

a process called [transcytosis](#). The transcytosis pathway includes [recycling endosomes](#), where endocytosed plasma membrane proteins can be stored until they are needed.

Tissue Replacement

A Natural Phenomenon

Bones regenerate every 10 years

Skin regenerates every two weeks

Macrophages scavenge senescent RBCs (10^{11} /day)

Tissue repair in response to minor injury

All organs have limited ability to regenerate (1 cm)

Every 30 Seconds a Person Dies from a Disease that can be Treated with Tissue Replacement

Human Intervention History

Blood Transfusion

- 1492: Pope Innocent VIII
- 1901: Human blood groups discovery Austrian Karl Landsteiner (1939 Rh)
- 1925: First transfusion center Alexander Bogdanov Moscow

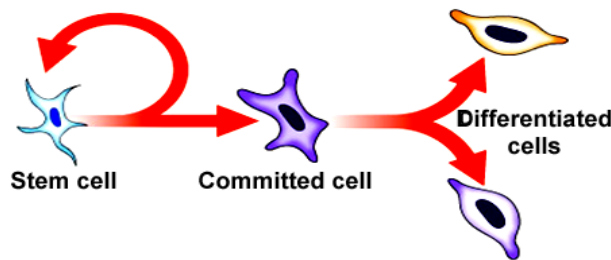
Solid Tissue

- 1905: Cornea transplant Eduard Zirm
- 1954: Kidney transplant Joseph Murray
- 1998: Hand transplant France
- 2005: Partial face transplant France

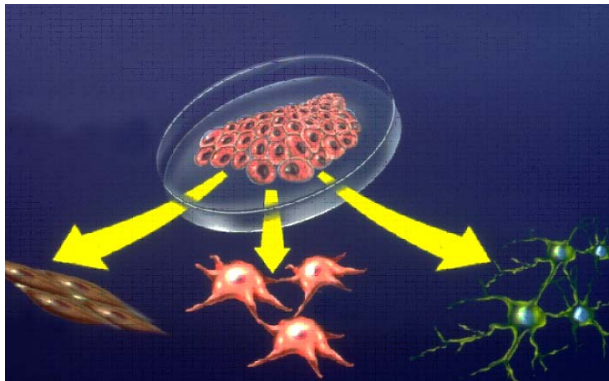
Current Approaches

- i. **Organ Repair via Stem Cells**
- ii. **Cytokine Induce Stem Cell Proliferation & Differentiation**
- iii. **In-vitro Organ/Tissue Fabrication**

i. Organ Repair via Stem Cells



Stem cells have two remarkable qualities that make them a good choice for repairing tissue damage



1. **Stem cells are unspecialized cells that renew themselves.**

2. **Under certain physiologic or experimental conditions, Stem cells can be induced to become cells with special functions.**

There are two types of stem cells Embryonic Stem Cells and Adult Stem Cells.

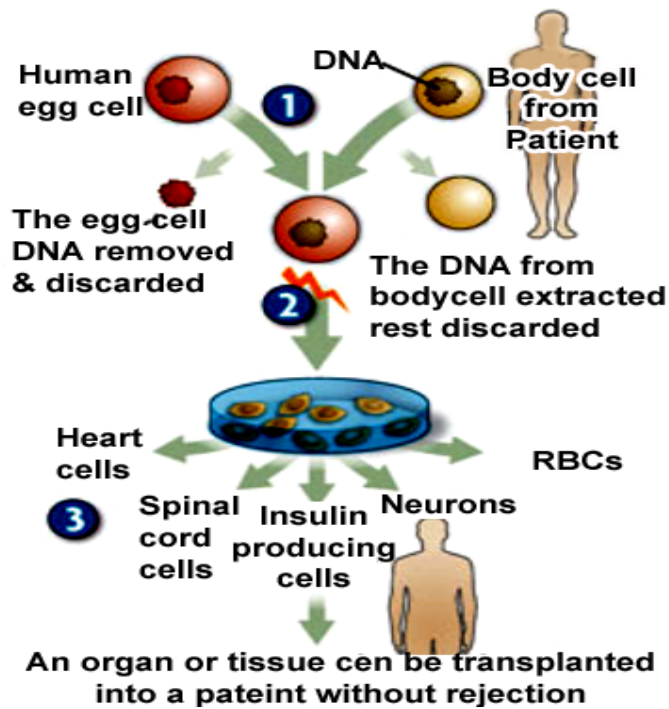
Embryonic Stem Cells

Are derived from the embryos (ethical dilemma: embryos destroyed when embryonic stem cells are isolated from the embryos) Blastomere Biopsy can be used to circumvent this issue.

Have the ability to differentiate into any cell type in the body.

Have unlimited ability to renew themselves.

Being foreign in origin, when transplanted can cause immune rejection. This issue can be resolved by generating customized embryonic stem cells (genetic makeup identical to the recipient) by using nuclear transfer technique.



Adult Stem cells

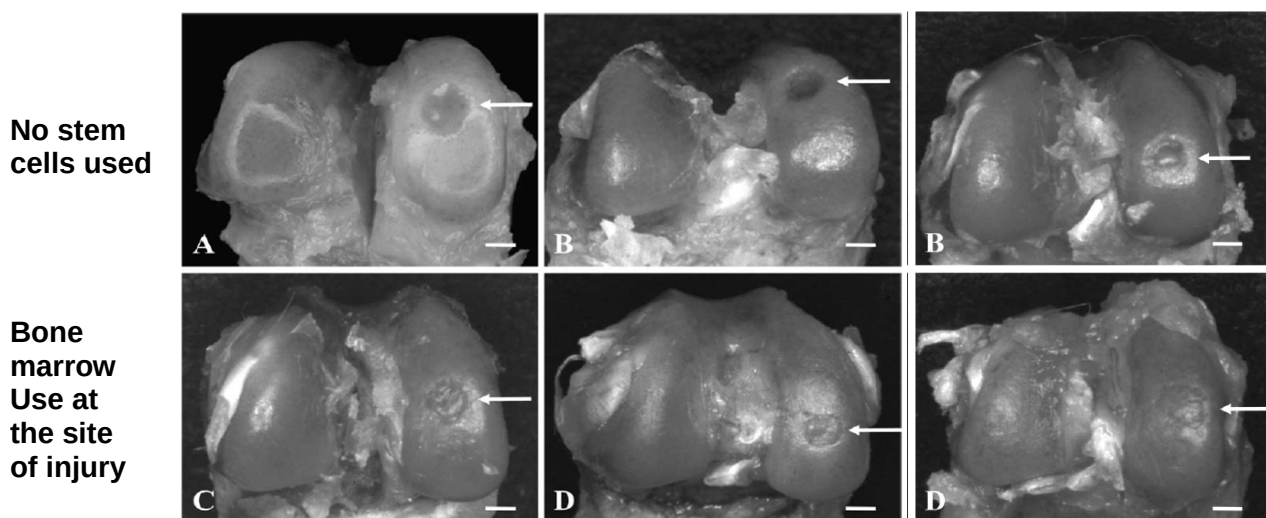
Are harvested from their niches from a (postnatal) individual.

No ethical dilemma for their use in research, since no potential person is destroyed in obtaining these cells.

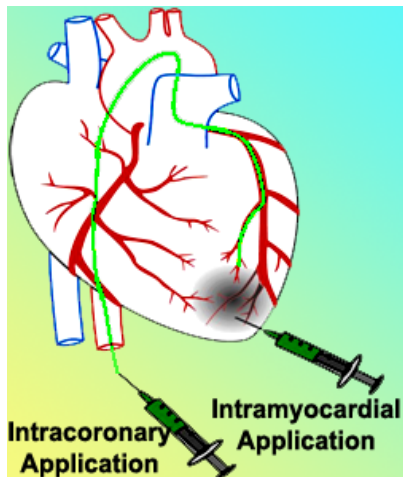
Have a limited ability to renew themselves and differentiate into different cell types.

Do not cause an immune rejection (if donor and the recipient are the same person).

Adult stem cells used to repair damaged rat bone

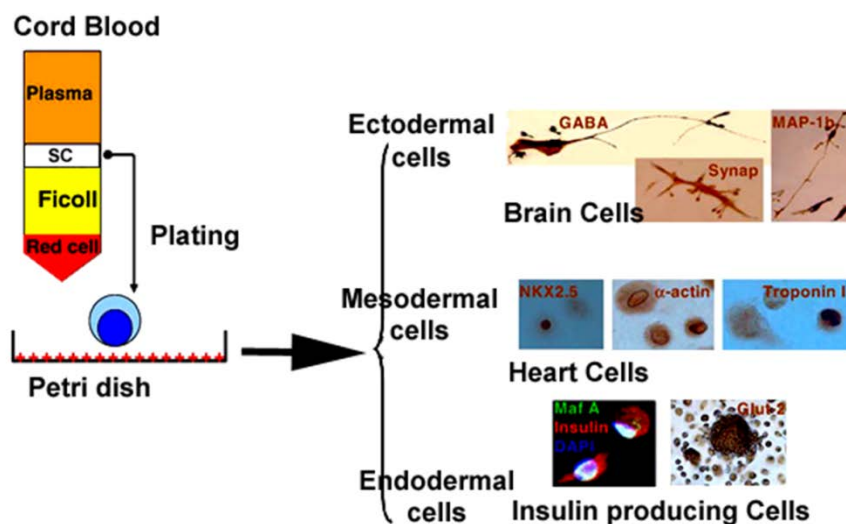


Repair of Damaged Heart Tissue after Heart Attack



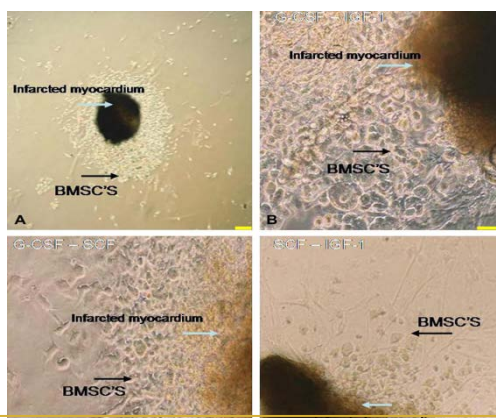
Düsseldorf group performed clinical stem cell application for the first time in 2001, treating acute myocardial infarction by intracoronary cell transfer. Currently, the optimal time point for cardiac stem cell application in an ischemic setting seems to lie within the first month after myocardial infarction. "There is growing evidence that bone marrow-derived stem/progenitor cells improve cardiac function" (Leri 2009).

Use of the Cord Blood cells may offer the best of both worlds (embryonic and Adult)



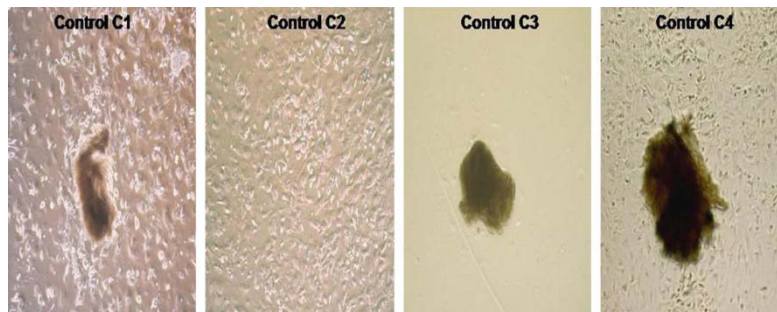
Cytokine Induced Stem Cell Proliferation & Differentiation

Almost all the tissues have a reservoir of stem cells which if appropriated stimulated can also be employed to repair damaged tissues.



IGF-1 and G-CSF complement each other in BMSC migration towards infarcted myocardium in a novel in vitro model

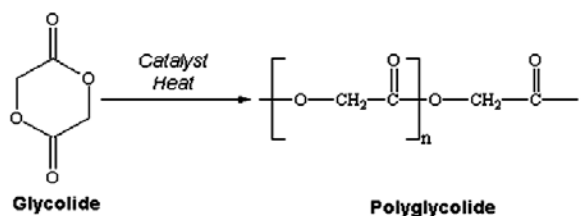
Mohsin Khan, Sobia Manzoor, Sadia Mohsin, Shaheen N. Khan¹, Fridoon Jawad Ahmad^{*,1}



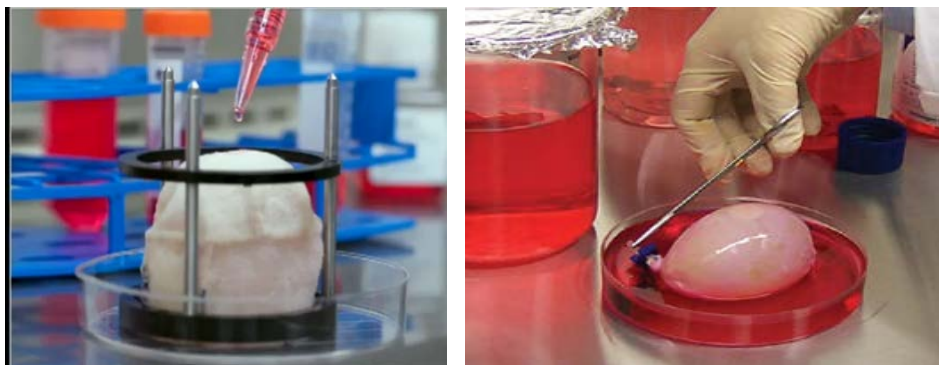
In-vitro Organ/Tissue Fabrication

Artificial Scaffolds

Scaffolds are required to fabricate solid organs since these organs have a specific shape that is important for their proper functioning.



can be used to make different scaffolds that can be seeded with cells in order to make complex three dimensional solid organs.

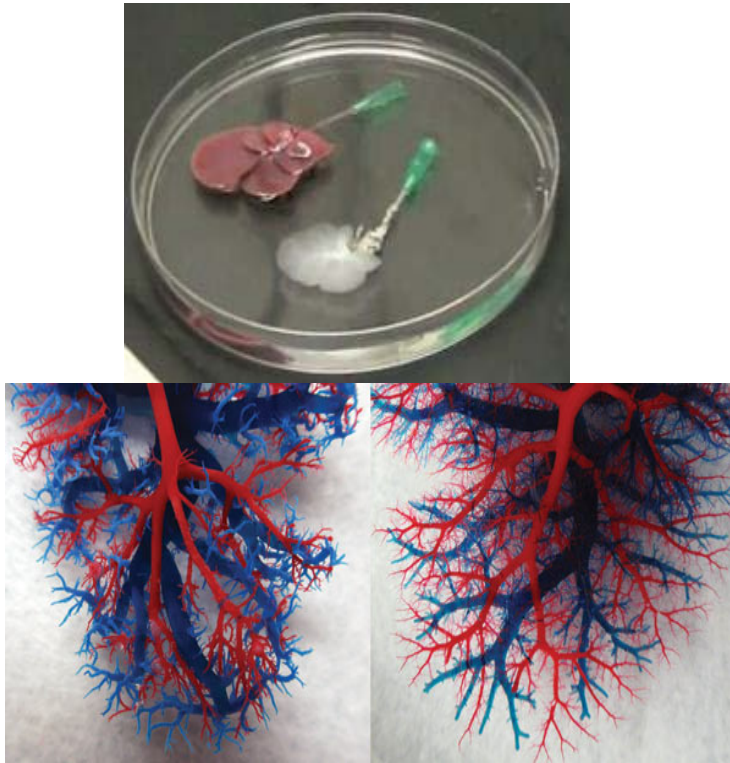


Scaffolds are biodegradable and once they are degraded only the cellular tissue remains.

Natural Scaffolds



De-cellularized amniotic membranes can be used to make two dimensional structures like cornea.



Solid organ with complex structures/vasculature (e.g. liver) can be de-cellularized and their extracellular matrix can be used to make such organs.