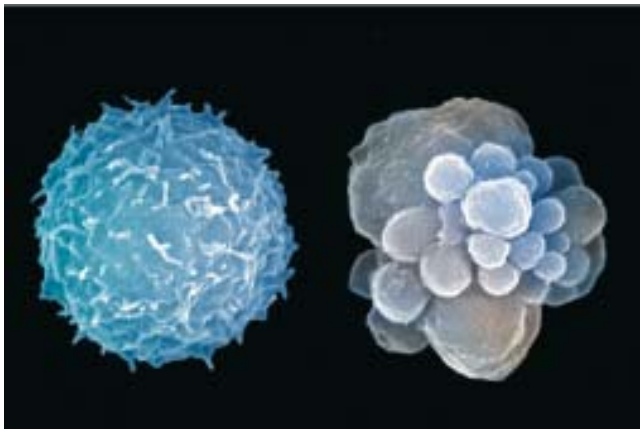


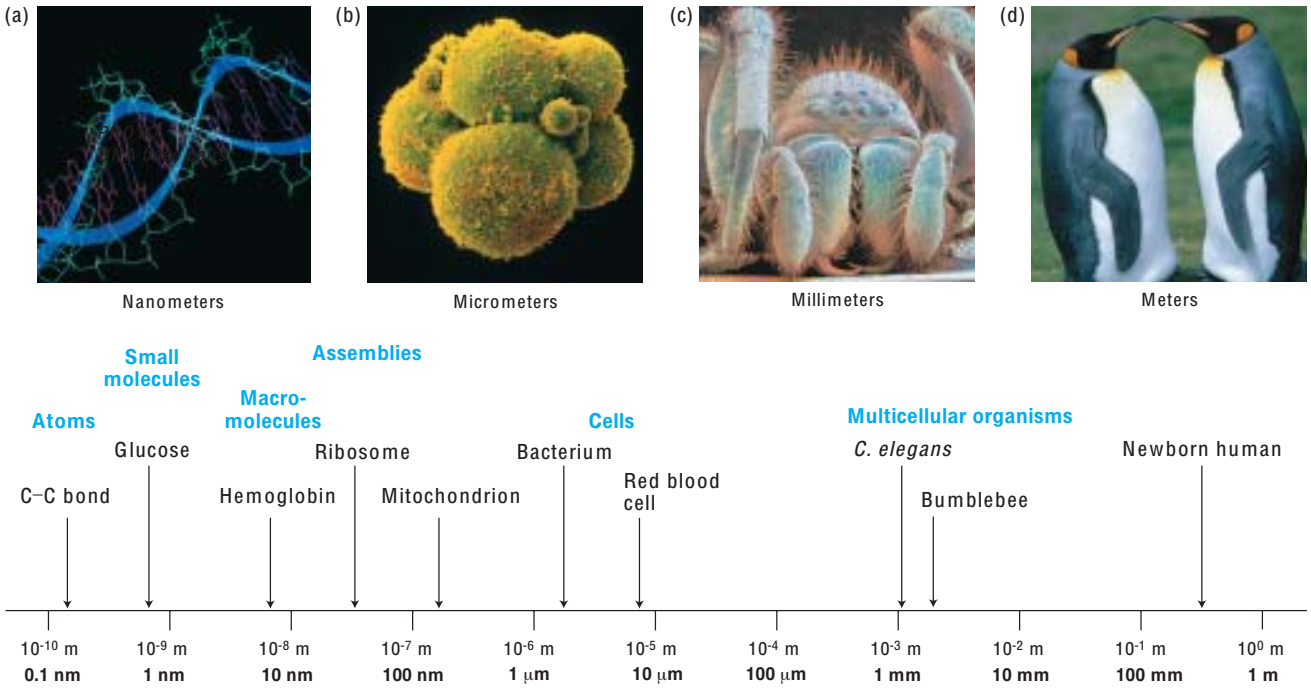
▲ FIGURE 1-19 Apoptotic cells break apart without spewing forth cell constituents that might harm neighboring cells. White blood cells normally look like the cell on the left. Cells undergoing programmed cell death (apoptosis), like the cell on the right, form numerous surface blebs that eventually are released. The cell is dying because it lacks certain growth signals. Apoptosis is important to eliminate virus-infected cells, remove cells where they are not needed (like the webbing that disappears as fingers develop), and to destroy immune system cells that would react with our own bodies. [Gopal Murti/Visuals Unlimited, Inc.]

fingers and thumb free to play the piano. Nerve cells in the brain soon die if they do not make proper or useful electrical connections with other cells. Some developing lymphocytes, the immune-system cells intended to recognize foreign proteins and polysaccharides, have the ability to react against our own tissues. Such self-reactive lymphocytes become programmed to die before they fully mature. If these cells are not weeded out before reaching maturity, they can cause autoimmune diseases in which our immune system destroys the very tissues it is meant to protect.

1.4 Investigating Cells and Their Parts

To build an integrated understanding of how the various molecular components that underlie cellular functions work together in a living cell, we must draw on various perspectives. Here, we look at how five disciplines—cell biology, biochemistry, genetics, genomics, and developmental biology—can contribute to our knowledge of cell structure and function. The experimental approaches of each field probe the cell's inner workings in different ways, allowing us to ask different types of questions about cells and what they do. Cell division provides a good example to illustrate the role of different perspectives in analyzing a complex cellular process.

The realm of biology ranges in scale more than a billionfold (Figure 1-20). Beyond that, it's ecology and earth science



▲ FIGURE 1-20 Biologists are interested in objects ranging in size from small molecules to the tallest trees. A sampling of biological objects aligned on a logarithmic scale. (a) The DNA double helix has a diameter of about 2 nm. (b) Eight-cell-stage human embryo three days after fertilization, about 200 μm across. (c) A wolf spider, about 15 mm across. (d) Emperor penguins are about 1 m tall. [Part (a) Will and Deni McIntyre. Part (b) Yorgas Nikas/Photo Researchers, Inc. Part (c) Gary Gaugler/Visuals Unlimited, Inc. Part (d) Hugh S. Rose/Visuals Unlimited, Inc.]

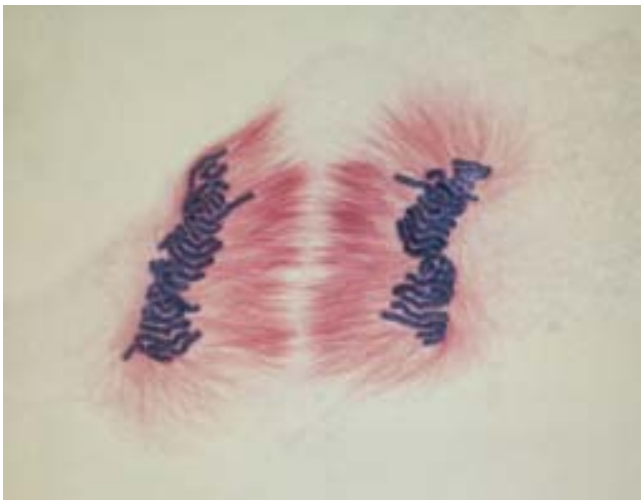
at the "macro" end, chemistry and physics at the "micro" end. The visible plants and animals that surround us are measured in meters (10^0–10^2 m). By looking closely, we can see a biological world of millimeters (1 mm = 10^{-3} m) and even tenths of millimeters (10^{-4} m). Setting aside oddities like chicken eggs, most cells are 1–100 micrometers (1 μm = 10^{-6} m) long and thus clearly visible only when magnified. To see the structures within cells, we must go farther down the size scale to 10–100 nanometers (1 nm = 10^{-9} m).

Cell Biology Reveals the Size, Shape, and Location of Cell Components

Actual observation of cells awaited development of the first, crude microscopes in the early 1600s. A compound microscope, the most useful type of light microscope, has two lenses. The total magnifying power is the product of the magnification by each lens. As better lenses were invented, the magnifying power and the ability to distinguish closely spaced objects, the **resolution**, increased greatly. Modern compound microscopes magnify the view about a thousandfold, so that a bacterium 1 micrometer (1 μm) long looks like it's a millimeter long. Objects about 0.2 μm apart can be discerned in these instruments.

Microscopy is most powerful when particular components of the cell are stained or labeled specifically, enabling them to be easily seen and located within the cell. A simple example is staining with dyes that bind specifically to DNA to visualize the chromosomes. Specific proteins can be detected by harnessing the binding specificity of **antibodies**, the proteins whose normal task is to help defend animals against infection and foreign substances. In general, each type of antibody binds to one protein or large polysaccharide and no other (Chapter 3). Purified antibodies can be chemically linked to a fluorescent molecule, which permits their detection in a special fluorescence microscope (Chapter 5). If a cell or tissue is treated with a detergent that partially dissolves cell membranes, fluorescent antibodies can drift in and bind to the specific protein they recognize. When the sample is viewed in the microscope, the bound fluorescent antibodies identify the location of the target protein (see Figure 1-15).

Better still is pinpointing proteins in living cells with intact membranes. One way of doing this is to introduce an engineered gene that codes for a hybrid protein: part of the hybrid protein is the cellular protein of interest; the other part is a protein that fluoresces when struck by ultraviolet light. A common fluorescent protein used for this purpose is *green fluorescent protein (GFP)*, a natural protein that makes some jellyfish colorful and fluorescent. GFP "tagging" could reveal, for instance, that a particular protein is first made on the endoplasmic reticulum and then is moved by the cell into the lysosomes. In this case, first the endoplasmic reticulum and later the lysosomes would glow in the dark.



▲ FIGURE 1-21 During the later stages of mitosis, microtubules (red) pull the replicated chromosomes (black) toward the ends of a dividing cell. This plant cell is stained with a DNA-binding dye (ethidium) to reveal chromosomes and with fluorescent-tagged antibodies specific for tubulin to reveal microtubules. At this stage in mitosis, the two copies of each replicated chromosome (called chromatids) have separated and are moving away from each other. [Courtesy of Andrew Bajer.]

Chromosomes are visible in the light microscope only during mitosis, when they become highly condensed. The extraordinary behavior of chromosomes during mitosis first was discovered using the improved compound microscopes of the late 1800s. About halfway through mitosis, the replicated chromosomes begin to move apart. Microtubules, one of the three types of cytoskeletal filaments, participate in this movement of chromosomes during mitosis. Fluorescent tagging of tubulin, the protein subunit that polymerizes to form microtubules, reveals structural details of cell division that otherwise could not be seen and allows observation of chromosome movement (Figure 1-21).

Electron microscopes use a focused beam of electrons instead of a beam of light. In transmission electron microscopy, specimens are cut into very thin sections and placed under a high vacuum, precluding examination of living cells. The resolution of transmission electron microscopes, about 0.1 nm, permits fine structural details to be distinguished, and their powerful magnification would make a 1-μm-long bacterial cell look like a soccer ball. Most of the organelles in eukaryotic cells and the double-layered structure of the plasma membrane were first observed with electron microscopes (Chapter 5). With new specialized electron microscopy techniques, three-dimensional models of organelles and large protein complexes can be constructed from multiple images. But to obtain a more detailed look at the individual macromolecules within cells, we must turn to techniques within the purview of biochemistry.

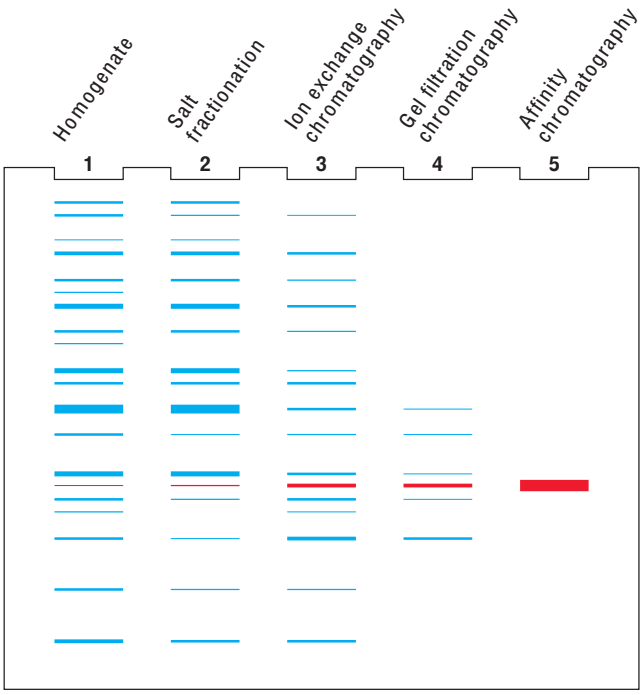
Biochemistry Reveals the Molecular Structure and Chemistry of Purified Cell Constituents

Biochemists extract the contents of cells and separate the constituents based on differences in their chemical or physical properties, a process called *fractionation*. Of particular interest are proteins, the workhorses of many cellular processes. A typical fractionation scheme involves use of various separation techniques in a sequential fashion. These separation techniques commonly are based on differences in the size of molecules or the electrical charge on their surface (Chapter 3). To purify a particular protein of interest, a purification scheme is designed so that each step yields a preparation with fewer and fewer contaminating proteins, until finally only the protein of interest remains (Figure 1-22).

The initial purification of a protein of interest from a cell extract often is a tedious, time-consuming task. Once a small amount of purified protein is obtained, antibodies to it can be produced by methods discussed in Chapter 6. For a biochemist, antibodies are near-perfect tools for isolating larger amounts of a protein of interest for further analysis. In effect, antibodies can "pluck out" the protein they specifically recognize and bind from a semipure sample containing numerous different proteins. An increasingly common alternative is to engineer a gene that encodes a protein of interest with a small attached protein "tag," which can be used to pull out the protein from whole cell extracts.

Purification of a protein is a necessary prelude to studies on how it catalyzes a chemical reaction or carries out other functions and how its activity is regulated. Some enzymes are made of multiple protein chains (subunits) with one chain catalyzing a chemical reaction and other chains regulating when and where that reaction occurs. The molecular machines that perform many critical cell processes constitute even larger assemblies of proteins. By separating the individual proteins composing such assemblies, their individual catalytic or other activities can be assessed. For example, purification and study of the activity of the individual proteins composing the DNA replication machine provided clues about how they work together to replicate DNA during cell division (Chapter 4).

The folded, three-dimensional structure, or conformation, of a protein is vital to its function. To understand the relation between the function of a protein and its form, we need to know both what it does and its detailed structure. The most widely used method for determining the complex structures of proteins, DNA, and RNA is x-ray crystallography. Computer-assisted analysis of the data often permits the location of every atom in a large, complex molecule to be determined. The double-helix structure of DNA, which is key to its role in heredity, was first proposed based on x-ray crystallographic studies. Throughout this book you will encounter numerous examples of protein structures as we zero in on how proteins work.



▲ **FIGURE 1-22 Biochemical purification of a protein from a cell extract often requires several separation techniques.** The purification can be followed by gel electrophoresis of the starting protein mixture and the fractions obtained from each purification step. In this procedure, a sample is applied to wells in the top of a gelatin-like slab and an electric field is applied. In the presence of appropriate salt and detergent concentrations, the proteins move through the fibers of the gel toward the anode, with larger proteins moving more slowly through the gel than smaller ones (see Figure 3-32). When the gel is stained, separated proteins are visible as distinct bands whose intensities are roughly proportional to the protein concentration. Shown here are schematic depictions of gels for the starting mixture of proteins (lane 1) and samples taken after each of several purification steps. In the first step, salt fractionation, proteins that precipitated with a certain amount of salt were re-dissolved; electrophoresis of this sample (lane 2) shows that it contains fewer proteins than the original mixture. The sample then was subjected in succession to three types of column chromatography that separate proteins by electrical charge, size, or binding affinity for a particular small molecule (see Figure 3-34). The final preparation is quite pure, as can be seen from the appearance of just one protein band in lane 5. [After J. Berg et al., 2002, *Biochemistry,* W. H. Freeman and Company, p. 87.]

Genetics Reveals the Consequences of Damaged Genes

Biochemical and crystallographic studies can tell us much about an individual protein, but they cannot prove that it is required for cell division or any other cell process. The importance of a protein is demonstrated most firmly if a mu-

tation that prevents its synthesis or makes it nonfunctional adversely affects the process under study.

We define the **genotype** of an organism as its composition of genes; the term also is commonly used in reference to different versions of a single gene or a small number of genes of interest in an individual organism. A diploid organism generally carries two versions (**alleles**) of each gene, one derived from each parent. There are important exceptions, such as the genes on the X and Y chromosomes in males of some species including our own. The **phenotype** is the visible outcome of a gene's action, like blue eyes versus brown eyes or the shapes of peas. In the early days of genetics, the location and chemical identity of genes were unknown; all that could be followed were the observable characteristics, the phenotypes. The concept that genes are like "beads" on a long "string," the chromosome, was proposed early in the 1900s based on genetic work with the fruit fly *Drosophila*.

In the classical genetics approach, mutants are isolated that lack the ability to do something a normal organism can do. Often large genetic "screens" are done, looking for many different mutant individuals (e.g., fruit flies, yeast cells) that are unable to complete a certain process, such as cell division or muscle formation. In experimental organisms or cultured cells, mutations usually are produced by treatment with a **mutagen**, a chemical or physical agent that promotes mutations in a largely random fashion. But how can we isolate and maintain mutant organisms or cells that are defective in some process, such as cell division, that is necessary for survival? One way is to look for **temperature-sensitive mutants**. These mutants are able to grow at one temperature, the *permissive* temperature, but not at another, usually higher temperature, the *nonpermissive* temperature. Normal cells can grow at either temperature. In most cases, a temperature-sensitive mutant produces an altered protein that works at the permissive temperature but unfolds and is nonfunctional at the nonpermissive temperature. Temperature-sensitive screens are readily done with viruses, bacteria, yeast, roundworms, and fruit flies.

By analyzing the effects of numerous different temperature-sensitive mutations that altered cell division, geneticists discovered all the genes necessary for cell division without knowing anything, initially, about which proteins they encode or how these proteins participate in the process. The great power of genetics is to reveal the existence and relevance of proteins without prior knowledge of their biochemical identity or molecular function. Eventually these "mutation-defined" genes were isolated and replicated (cloned) with **recombinant DNA** techniques discussed in Chapter 9. With the isolated genes in hand, the encoded proteins could be produced in the test tube or in engineered bacteria or cultured cells. Then the biochemists could investigate whether the proteins associate with other proteins or DNA or catalyze particular chemical reactions during cell division (Chapter 21).

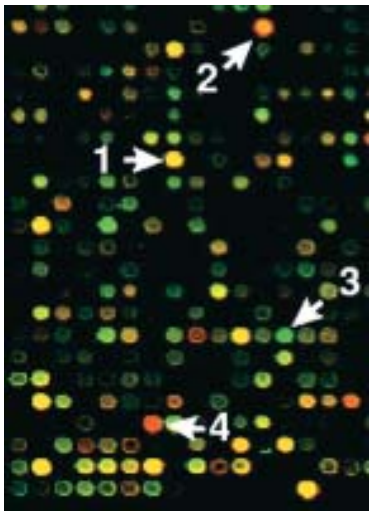
The analysis of genome sequences from various organisms during the past decade has identified many previously unknown DNA regions that are likely to encode proteins (i.e., protein-coding genes). The general function of the protein encoded by a sequence-identified gene may be deduced by analogy with known proteins of similar sequence. Rather than randomly isolating mutations in novel genes, several techniques are now available for inactivating specific genes by engineering mutations into them (Chapter 9). The effects of such deliberate gene-specific mutations provide information about the role of the encoded proteins in living organisms. This application of genetic techniques starts with a gene/protein sequence and ends up with a mutant phenotype; traditional genetics starts with a mutant phenotype and ends up with a gene/protein sequence.

Genomics Reveals Differences in the Structure and Expression of Entire Genomes

Biochemistry and genetics generally focus on one gene and its encoded protein at a time. While powerful, these traditional approaches do not give a comprehensive view of the structure and activity of an organism's genome, its entire set of genes. The field of **genomics** does just that, encompassing the molecular characterization of whole genomes and the determination of global patterns of gene expression. The recent completion of the genome sequences for more than 80 species of bacteria and several eukaryotes now permits comparisons of entire genomes from different species. The results provide overwhelming evidence of the molecular unity of life and the evolutionary processes that made us what we are (see Section 1.5). Genomics-based methods for comparing thousands of pieces of DNA from different individuals all at the same time are proving useful in tracing the history and migrations of plants and animals and in following the inheritance of diseases in human families.

New methods using **DNA microarrays** can simultaneously detect all the mRNAs present in a cell, thereby indicating which genes are being transcribed. Such global patterns of gene expression clearly show that liver cells transcribe a quite different set of genes than do white blood cells or skin cells. Changes in gene expression also can be monitored during a disease process, in response to drugs or other external signals, and during development. For instance, the recent identification of all the mRNAs present in cultured fibroblasts before, during, and after they divide has given us an overall view of transcriptional changes that occur during cell division (Figure 1-23). Cancer diagnosis is being transformed because previously indistinguishable cancer cells have distinct gene expression patterns and prognoses (Chapter 23). Similar studies with different organisms and cell types are revealing what is universal about the genes involved in cell division and what is specific to particular organisms.

The entire complement of proteins in a cell, its **proteome**, is controlled in part by changes in gene transcription. The regulated synthesis, processing, localization, and degradation of specific proteins also play roles in determining the proteome of a particular cell, and the association of certain proteins with one another is critical to the functional abilities



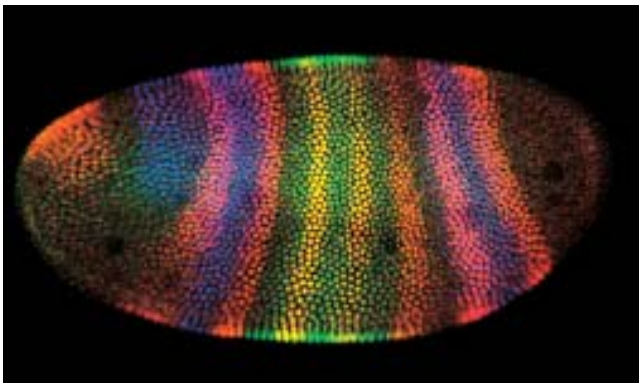
▲ FIGURE 1-23 DNA microarray analysis gives a global view of changes in transcription following addition of serum to cultured human cells. Serum contains growth factors that stimulate nondividing cells to begin growing and dividing. DNA microarray analysis can detect the relative transcription of genes in two different cell populations (see Figure 9-35). The microarray consists of tiny spots of DNA attached to a microscope slide. Each spot contains many copies of a DNA sequence from a single human gene. One preparation of RNA, containing all the different types of RNA being made in nongrowing cells cultured without serum, is labeled with green fluorescent molecules. Another RNA population from growing, serum-treated, cells is labeled with red. The two are mixed and hybridized to the slide, where they "zipper up" with their corresponding genes. Green spots (e.g., spot 3) therefore indicate genes that are transcribed in nondividing (serum-deprived) cells; red spots (e.g., spot 4) indicate genes that are transcribed in dividing cells, and yellow spots (e.g., spots 1 and 2) indicate genes that are transcribed equally in dividing and nondividing cells. [From V. R. Iyer et al., 1999, *Science* **283**:83.]

of cells. New techniques for monitoring the presence and interactions of numerous proteins simultaneously, called **proteomics**, are one way of assembling a comprehensive view of the proteins and molecular machines important for cell functioning. The field of proteomics will advance dramatically once high-throughput x-ray crystallography, currently under development, permits researchers to rapidly determine the structures of hundreds or thousands of proteins.

Developmental Biology Reveals Changes in the Properties of Cells as They Specialize

Another approach to viewing cells comes from studying how they change during development of a complex organism. Bacteria, algae, and unicellular eukaryotes (protozoans, yeasts) often, but by no means always, can work solo. The concerted actions of the trillions of cells that compose our bodies require an enormous amount of communication and division of labor. During the development of multicellular organisms, differentiation processes form hundreds of cell types, each specialized for a particular task: transmission of electrical signals by neurons, transport of oxygen by red blood cells, destruction of infecting bacteria by macrophages, contraction by muscle cells, chemical processing by liver cells.

Many of the differences among differentiated cells are due to production of specific sets of proteins needed to carry out the unique functions of each cell type. That is, only a subset of an organism's genes is transcribed at any given time or in any given cell. Such **differential gene expression** at different times or in different cell types occurs in bacteria, fungi, plants, animals, and even viruses. Differential gene expression is readily apparent in an early fly embryo in which all the cells look alike until they are stained to detect the proteins encoded by particular genes (Figure 1-24). Transcription can change within one cell type in response to an external signal or in accordance with a biological clock; some genes, for instance, undergo a daily cycle between low and high transcription rates.



▲ FIGURE 1-24 Differential gene expression can be detected in early fly embryos before cells are morphologically different. An early *Drosophila* embryo has about 6000 cells covering its surface, most of which are indistinguishable by simple light microscopy. If the embryo is made permeable to antibodies with a detergent that partially dissolves membranes, the antibodies can find and bind to the proteins they recognize. In this embryo we see antibodies tagged with a fluorescent label bound to proteins that are in the nuclei; each small sphere corresponds to one nucleus. Three different antibodies were used, each specific for a different protein and each giving a distinct color (yellow, green, or blue) in a fluorescence microscope. The red color is added to highlight overlaps between the yellow and blue stains. The locations of the different proteins show that the cells are in fact different at this early stage, with particular genes turned on in specific stripes of cells. These genes control the subdivision of the body into repeating segments, like the black and yellow stripes of a hornet. [Courtesy of Sean Carroll, University of Wisconsin.]

Producing different kinds of cells is not enough to make an organism, any more than collecting all the parts of a truck in one pile gives you a truck. The various cell types must be organized and assembled into all the tissues and organs. Even more remarkable, these body parts must work almost immediately after their formation and continue working during the growth process. For instance, the human heart begins to beat when it is less than 3 mm long, when we are mere 23-day-old embryos, and continues beating as it grows into a fist-size muscle. From a few hundred cells to billions, and still ticking.

In the developing organism, cells grow and divide at some times and not others, they assemble and communicate, they prevent or repair errors in the developmental process, and they coordinate each tissue with others. In the adult organism, cell division largely stops in most organs. If part of an organ such as the liver is damaged or removed, cell division resumes until the organ is regenerated. The legend goes that Zeus punished Prometheus for giving humans fire by chaining him to a rock and having an eagle eat his liver. The punishment was eternal because, as the Greeks evidently knew, the liver regenerates.

Developmental studies involve watching where, when, and how different kinds of cells form, discovering which signals trigger and coordinate developmental events, and understanding the differential gene action that underlies differentiation (Chapters 15 and 22). During development we can see cells change in their normal context of other cells. Cell biology, biochemistry, cell biology, genetics, and genomics approaches are all employed in studying cells during development.

Choosing the Right Experimental Organism for the Job

Our current understanding of the molecular functioning of cells rests on studies with viruses, bacteria, yeast, protozoa, slime molds, plants, frogs, sea urchins, worms, insects, fish, chickens, mice, and humans. For various reasons, some organisms are more appropriate than others for answering particular questions. Because of the evolutionary conservation of genes, proteins, organelles, cell types, and so forth, discoveries about biological structures and functions obtained with one experimental organism often apply to others. Thus researchers generally conduct studies with the organism that is most suitable for rapidly and completely answering the question being posed, knowing that the results obtained in one organism are likely to be broadly applicable. Figure 1-25 summarizes the typical experimental uses of various organisms whose genomes have been sequenced completely or nearly so. The availability of the genome sequences for these organisms makes them particularly useful for genetics and genomics studies.

Bacteria have several advantages as experimental organisms: They grow rapidly, possess elegant mechanisms for controlling gene activity, and have powerful genetics. This latter property relates to the small size of bacterial genomes, the ease of obtaining mutants, the availability of techniques for transferring genes into bacteria, an enormous wealth of knowledge about bacterial gene control and protein functions, and the relative simplicity of mapping genes relative to one another in the genome. Single-celled yeasts not only have some of the same advantages as bacteria, but also possess the cell organization, marked by the presence of a nucleus and organelles, that is characteristic of all eukaryotes.

Studies of cells in specialized tissues make use of animal and plant "models," that is, experimental organisms with attributes typical of many others. Nerve cells and muscle cells, for instance, traditionally were studied in mammals or in creatures with especially large or accessible cells, such as the giant neural cells of the squid and sea hare or the flight muscles of birds. More recently, muscle and nerve development have been extensively studied in fruit flies *(Drosophila melanogaster)*, roundworms *(Caenorhabditis elegans)*, and zebrafish in which mutants can be readily isolated. Organisms with large-celled embryos that develop outside the

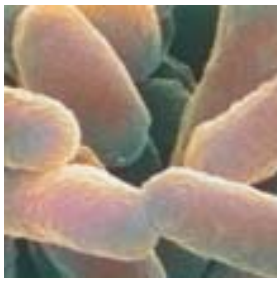
▶ FIGURE 1-25 Each experimental organism used in cell biology has advantages for certain types of studies. Viruses and bacteria have small genomes amenable to genetic dissection. Many insights into gene control initially came from studies with these organisms. The yeast *Saccharomyces cerevisiae* has the cellular organization of a eukaryote but is a relatively simple single-celled organism that is easy to grow and to manipulate genetically. In the nematode worm *Caenorhabditis elegans*, which has a small number of cells arranged in a nearly identical way in every worm, the formation of each individual cell can be traced. The fruit fly *Drosophila melanogaster*, first used to discover the properties of chromosomes, has been especially valuable in identifying genes that control embryonic development. Many of these genes are evolutionarily conserved in humans. The zebrafish *Danio rerio* is used for rapid genetic screens to identify genes that control development and organogenesis. Of the experimental animal systems, mice (*Mus musculus*) are evolutionarily the closest to humans and have provided models for studying numerous human genetic and infectious diseases. The mustard-family weed *Arabidopsis thaliana*, sometimes described as the *Drosophila* of the plant kingdom, has been used for genetic screens to identify genes involved in nearly every aspect of plant life. Genome sequencing is completed for many viruses and bacterial species, the yeast *Saccharomyces cerevisiae*, the roundworm *C. elegans*, the fruit fly *D. melanogaster*, humans, and the plant *Arabidopsis thaliana*. It is mostly completed for mice and in progress for zebrafish. Other organisms, particularly frogs, sea urchins, chickens, and slime molds, continue to be immensely valuable for cell biology research. Increasingly, a wide variety of other species are used, especially for studies of evolution of cells and mechanisms. [Part (a) Visuals Unlimited, Inc. Part (b) Kari Lountmaa/Science Photo Library/Photo Researchers, Inc. Part (c) Scimat/Photo Researchers, Inc. Part (d) Photo Researchers, Inc. Part (e) Darwin Dale/Photo Researchers, Inc. Part (f) Inge Spence/Visuals Unlimited, Inc. Part (g) J. M. Labat/Jancana/Visuals Unlimited, Inc. Part (h) Darwin Dale/Photo Researchers, Inc.]

(a)

Viruses

Proteins involved in DNA, RNA, protein synthesis
Gene regulation
Cancer and control of cell proliferation
Transport of proteins and organelles inside cells
Infection and immunity
Possible gene therapy approaches

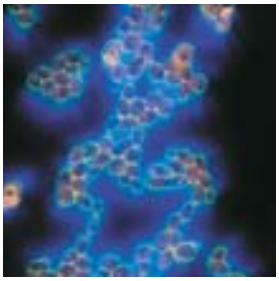
(b)



Bacteria

Proteins involved in DNA, RNA, protein synthesis, metabolism
Gene regulation
Targets for new antibiotics
Cell cycle
Signaling

(c)



Yeast (*Saccharomyces cerevisiae*)

Control of cell cycle and cell division
Protein secretion and membrane biogenesis
Function of the cytoskeleton
Cell differentiation
Aging
Gene regulation and chromosome structure

(d)

Roundworm (*Caenorhabditis elegans*)

Development of the body plan
Cell lineage
Formation and function of the nervous system
Control of programmed cell death
Cell proliferation and cancer genes
Aging
Behavior
Gene regulation and chromosome structure

(e)

Fruit fly (*Drosophila melanogaster*)

Development of the body plan
Generation of differentiated cell lineages
Formation of the nervous system, heart, and musculature
Programmed cell death
Genetic control of behavior
Cancer genes and control of cell proliferation
Control of cell polarization
Effects of drugs, alcohol, pesticides

(f)

Zebrafish

Development of vertebrate body tissues
Formation and function of brain and nervous system
Birth defects
Cancer

(g)

Mice, including cultured cells

Development of body tissues
Function of mammalian immune system
Formation and function of brain and nervous system
Models of cancers and other human diseases
Gene regulation and inheritance
Infectious disease

(h)

Plant (*Arabidopsis thaliana*)

Development and patterning of tissues
Genetics of cell biology
Agricultural applications
Physiology
Gene regulation
Immunity
Infectious disease

mother (e.g., frogs, sea urchins, fish, and chickens) are extremely useful for tracing the fates of cells as they form different tissues and for making extracts for biochemical studies. For instance, a key protein in regulating mitosis was first identified in studies with frog and sea urchin embryos and subsequently purified from extracts (Chapter 21).

Using recombinant DNA techniques researchers can engineer specific genes to contain mutations that inactivate or increase production of their encoded proteins. Such genes can be introduced into the embryos of worms, flies, frogs, sea urchins, chickens, mice, a variety of plants, and other organisms, permitting the effects of activating a gene abnormally or inhibiting a normal gene function to be assessed. This approach is being used extensively to produce mouse versions of human genetic diseases. New techniques specifically for inactivating particular genes by injecting short pieces of RNA are making quick tests of gene functions possible in many organisms.

Mice have one enormous advantage over other experimental organisms: they are the closest to humans of any animal for which powerful genetic approaches are feasible. Engineered mouse genes carrying mutations similar to those associated with a particular inherited disease in humans can be introduced into mouse embryonic stem (ES) cells. These cells can be injected into an early embryo, which is then implanted into a pseudopregnant female mouse (Chapter 9). If the mice that develop from the injected ES cells exhibit diseases similar to the human disease, then the link between the disease and mutations in a particular gene or genes is supported. Once mouse models of a human disease are available, further studies on the molecular defects causing the disease can be done and new treatments can be tested, thereby minimizing human exposure to untested treatments.

A continuous unplanned genetic screen has been performed on human populations for millennia. Thousands of inherited traits have been identified and, more recently, mapped to locations on the chromosomes. Some of these traits are inherited propensities to get a disease; others are eye color or other minor characteristics. Genetic variations in virtually every aspect of cell biology can be found in human populations, allowing studies of normal and disease states and of variant cells in culture.

Less-common experimental organisms offer possibilities for exploring unique or exotic properties of cells and for studying standard properties of cells that are exaggerated in a useful fashion in a particular animal. For example, the ends of chromosomes, the **telomeres**, are extremely dilute in most cells. Human cells typically contain 92 telomeres (46 chromosomes $\times$ 2 ends per chromosome). In contrast, some protozoa with unusual "fragmented" chromosomes contain millions of telomeres per cell. Recent discoveries about telomere structure have benefited greatly from using this natural variation for experimental advantage.

1.5 A Genome Perspective on Evolution

Comprehensive studies of genes and proteins from many organisms are giving us an extraordinary documentation of the history of life. We share with other eukaryotes thousands of individual proteins, hundreds of macromolecular machines, and most of our organelles, all as a result of our shared evolutionary history. New insights into molecular cell biology arising from genomics are leading to a fuller appreciation of the elegant molecular machines that arose during billions of years of genetic tinkering and evolutionary selection for the most efficient, precise designs. Despite all that we currently know about cells, many new proteins, new macromolecular assemblies, and new activities of known ones remain to be discovered. Once a more complete description of cells is in hand, we will be ready to fully investigate the rippling, flowing dynamics of living systems.

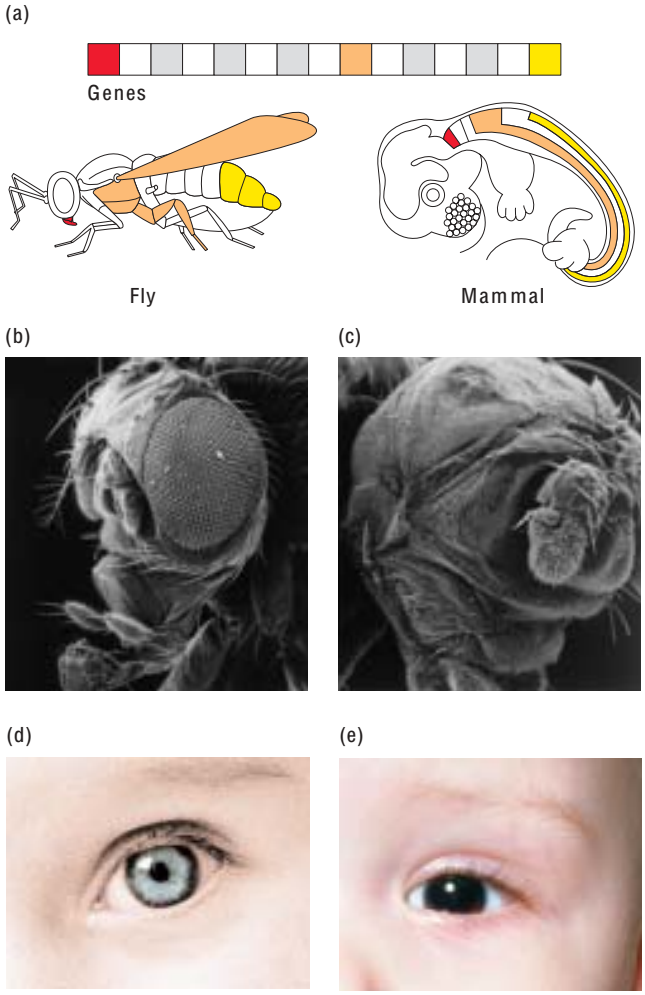
Metabolic Proteins, the Genetic Code, and Organelle Structures Are Nearly Universal

Even organisms that look incredibly different share many biochemical properties. For instance, the enzymes that catalyze degradation of sugars and many other simple chemical reactions in cells have similar structures and mechanisms in most living things. The genetic code whereby the nucleotide sequences of mRNA specifies the amino acid sequences of proteins can be read equally well by a bacterial cell and a human cell. Because of the universal nature of the genetic code, bacterial "factories" can be designed to manufacture growth factors, insulin, clotting factors, and other human proteins with therapeutic uses. The biochemical similarities among organisms also extend to the organelles found in eukaryotic cells. The basic structures and functions of these subcellular components are largely conserved in all eukaryotes.

Computer analysis of DNA sequence data, now available for numerous bacterial species and several eukaryotes, can locate protein-coding genes within genomes. With the aid of the genetic code, the amino acid sequences of proteins can be deduced from the corresponding gene sequences. Although simple conceptually, "finding" genes and deducing the amino acid sequences of their encoded proteins is complicated in practice because of the many noncoding regions in eukaryotic DNA (Chapter 9). Despite the difficulties and occasional ambiguities in analyzing DNA sequences, comparisons of the genomes from a wide range of organisms provide stunning, compelling evidence for the conservation of the molecular mechanisms that build and change organisms and for the common evolutionary history of all species.

Many Genes Controlling Development Are Remarkably Similar in Humans and Other Animals

As humans, we probably have a biased and somewhat exaggerated view of our status in the animal kingdom. Pride in our swollen forebrain and its associated mental capabilities may blind us to the remarkably sophisticated abilities of other species: navigation by birds, the sonar system of bats, homing by salmon, or the flight of a fly.



◀ FIGURE 1-26 Similar genes, conserved during evolution, regulate many developmental processes in diverse animals. Insects and mammals are estimated to have had a common ancestor about half a billion years ago. They share genes that control similar processes, such as growth of heart and eyes and organization of the body plan, indicating conservation of function from ancient times. (a) *Hox* genes are found in clusters on the chromosomes of most or all animals. *Hox* genes encode related proteins that control the activities of other genes. *Hox* genes direct the development of different segments along the head-to-tail axis of many animals as indicated by corresponding colors. Each gene is activated (transcriptually) in a specific region along the head-to-toe axis and controls the growth of tissues there. For example, in mice the *Hox* genes are responsible for the distinctive shapes of vertebrae. Mutations affecting *Hox* genes in flies cause body parts to form in the wrong locations, such as legs in lieu of antennae on the head. These genes provide a head-to-tail address and serve to direct formation of the right structures in the right places. (b) Development of the large compound eyes in fruit flies requires a gene called *eyeless* (named for the mutant phenotype). (c) Flies with inactivated *eyeless* genes lack eyes. (d) Normal human eyes require the human gene, called *Pax6*, that corresponds to *eyeless*. (e) People lacking adequate *Pax6* function have the genetic disease *aniridia*, a lack of irises in the eyes. *Pax6* and *eyeless* encode highly related proteins that regulate the activities of other genes, and are descended from the same ancestral gene. [Parts (a) and (b) Andreas Hefti, Interdepartmental Electron Microscopy (IEM) Biocenter, University of Basel. Part (d) © Simon Fraser/Photo Researchers, Inc.]

Despite all the evidence for evolutionary unity at the cellular and physiological levels, everyone expected that genes regulating animal development would differ greatly from one phylum to the next. After all, insects and sea urchins and mammals look *so* different. We must have many unique proteins to create a brain like ours . . . or must we? The fruits of research in developmental genetics during the past two decades reveal that insects and mammals, which have a common ancestor about half a billion years ago, possess many similar development-regulating genes (Figure 1-26). Indeed, a large number of these genes appear to be conserved in many and perhaps all animals. Remarkably, the developmental functions of the proteins encoded by these genes are also often preserved. For instance, certain proteins involved in eye development in insects are related to protein regulators of eye development in mammals. Same for development of the heart, gut, lungs, and capillaries and for placement of body parts along the head-to-tail and back-to-front body axes (Chapter 15).

This is not to say that all genes or proteins are evolutionarily conserved. Many striking examples exist of proteins that, as far as we can tell, are utterly absent from certain lineages of animals. Plants, not surprisingly, exhibit many such differences from animals after a billion-year separation in their evolution. Yet certain DNA-binding proteins differ between peas and cows at only two amino acids out of 102!

Darwin's Ideas About the Evolution of Whole Animals Are Relevant to Genes

Darwin did not know that genes exist or how they change, but we do: the DNA replication machine makes an error, or a mutagen causes replacement of one nucleotide with another or breakage of a chromosome. Some changes in the genome are innocuous, some mildly harmful, some deadly; a very few are beneficial. Mutations can change the sequence of a gene in a way that modifies the activity of the encoded protein or alters when, where, and in what amounts the protein is produced in the body.

Gene-sequence changes that are harmful will be lost from a population of organisms because the affected individuals cannot survive as well as their relatives. This selection process is exactly what Darwin described without knowing the underlying mechanisms that cause organisms to vary. Thus the selection of whole organisms for survival is really a selection of genes, or more accurately sets of genes. A population of organisms often contains many variants that are

all roughly equally well-suited to the prevailing conditions. When conditions change—a fire, a flood, loss of preferred food supply, climate shift—variants that are better able to adapt will survive, and those less suited to the new conditions will begin to die out. In this way, the genetic composition of a population of organisms can change over time.

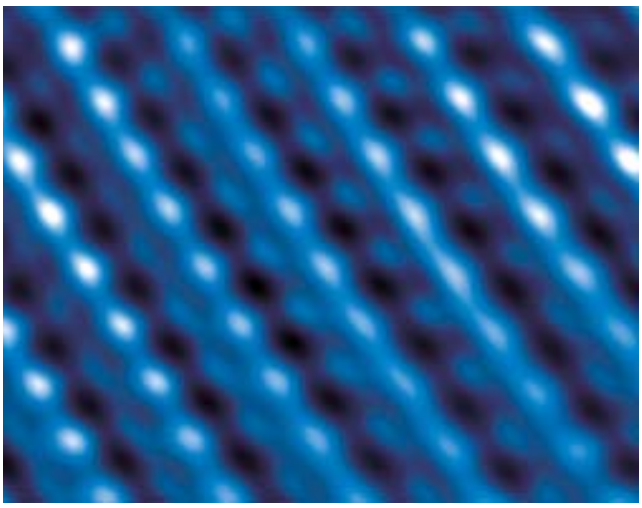
Human Medicine Is Informed by Research on Other Organisms

Mutations that occur in certain genes during the course of our lives contribute to formation of various human cancers. The normal, wild-type forms of such "cancer-causing" genes generally encode proteins that help regulate cell proliferation or death (Chapter 23). We also can inherit from our parents mutant copies of genes that cause all manner of genetic diseases, such as cystic fibrosis, muscular dystrophy, sickle cell anemia, and Huntington's disease. Happily we can also inherit genes that make us robustly resist disease. A remarkable number of genes associated with cancer and other human diseases are present in evolutionarily distant animals. For example, a recent study shows that more than three-quarters of the known human disease genes are related to genes found in the fruit fly *Drosophila*.

With the identification of human disease genes in other organisms, experimental studies in experimentally tractable organisms should lead to rapid progress in understanding the normal functions of the disease-related genes and what occurs when things go awry. Conversely, the disease states themselves constitute a genetic analysis with well-studied phenotypes. All the genes that can be altered to cause a certain disease may encode a group of functionally related proteins. Thus clues about the normal cellular functions of proteins come from human diseases and can be used to guide initial research into mechanism. For instance, genes initially identified because of their link to cancer in humans can be studied in the context of normal development in various model organisms, providing further insight about the functions of their protein products.

2

CHEMICAL FOUNDATIONS



Polysaccharide chains on the surface of cellulose visualized by atomic force microscopy. [Courtesy of M. Miles from A. A. Baker et al., 2000, *Biophys J.* **79**:1139–1145.]

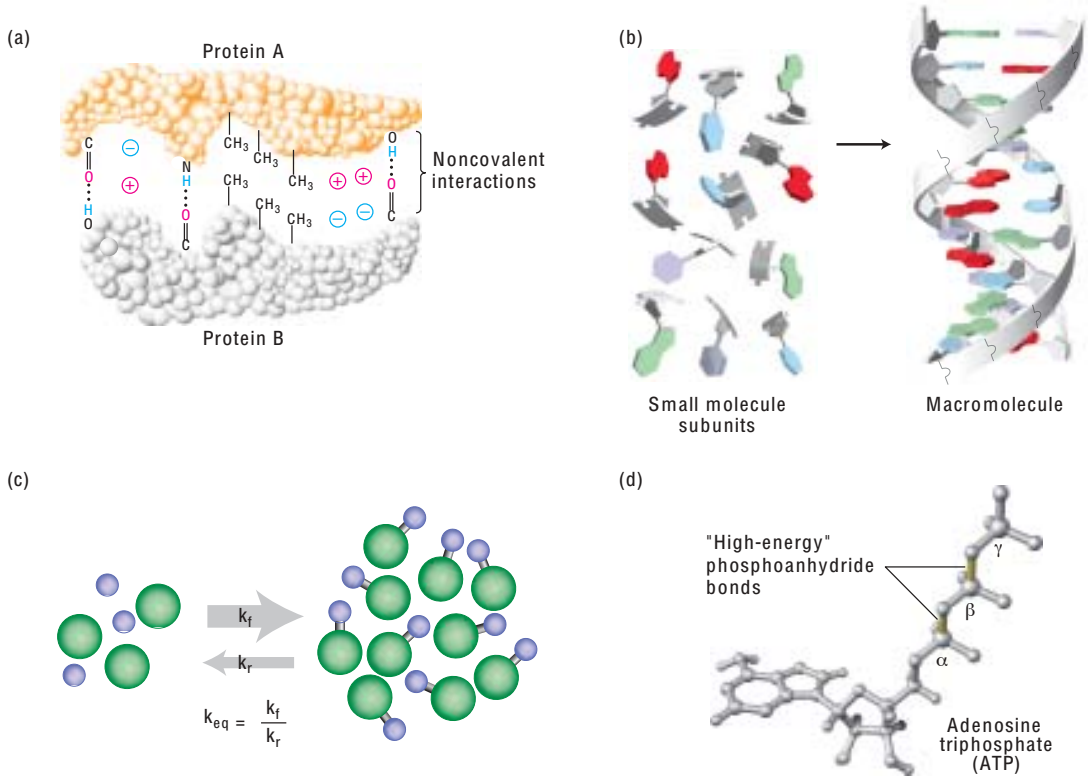
The life of a cell depends on thousands of chemical interactions and reactions exquisitely coordinated with one another in time and space and under the influence of the cell's genetic instructions and its environment. How does a cell extract critical nutrients and information from its environment? How does a cell convert the energy stored in nutrients into work (movement, synthesis of critical components)? How does a cell transform nutrients into the fundamental structures required for its survival (cell wall, nucleus, nucleic acids, proteins, cytoskeleton)? How does a cell link itself to other cells to form a tissue? How do cells communicate with one another so that the organism as a whole can function? One of the goals of molecular cell biology is to answer such questions about the structure and function of cells and organisms in terms of the properties of individual molecules and ions.

Life first arose in a watery environment, and the properties of this ubiquitous substance have a profound influence on the chemistry of life. Constituting 70–80 percent by weight of most cells, water is the most abundant molecule in biological systems. About 7 percent of the weight of living matter is composed of inorganic ions and small molecules such as amino acids (the building blocks of proteins), nucleotides (the building blocks of DNA and RNA), lipids (the building blocks of biomembranes), and sugars (the building blocks of starches and cellulose), the remainder being the macromolecules and macromolecular aggregates composed of these building blocks.

Many biomolecules (e.g., sugars) readily dissolve in water; these water-liking molecules are described as **hydrophilic**. Other biomolecules (e.g., fats like triacylglycerols) shun water; these are said to be **hydrophobic** (water-fearing). Still other biomolecules (e.g., phospholipids), referred to as **amphipathic**, are a bit schizophrenic, containing both hydrophilic and hydrophobic regions. These are used to build the membranes that surround cells and their internal organelles (Chapter 5). The smooth functioning of cells, tissues, and organisms depends on all these molecules, from the smallest to the largest. Indeed, the chemistry of the simple proton (H^+) with a mass of 1 dalton (Da) can be as important to the survival of a human cell as that of each gigantic DNA molecule with a mass as large as 8.6×10^{10} Da (single strand of DNA from human chromosome 1).

A relatively small number of principles and facts of chemistry are essential for understanding cellular processes at the molecular level (Figure 2-1). In this chapter we review some of these key principles and facts, beginning with the covalent bonds that connect atoms into a molecule and the noncovalent forces that stabilize groups of atoms within and between molecules. We then consider the key properties of the basic building blocks of cellular structures. After reviewing those aspects of chemical equilibrium that are most relevant to biological systems, we end the chapter with basic

OUTLINE



▲ FIGURE 2-1 Chemistry of life: key concepts. (a) Covalent and noncovalent interactions lie at the heart of all biomolecules, as when two proteins with complementary shapes and chemical properties come together to form a tightly bound complex. In addition to the covalent bonds that hold the atoms of an amino acid together and link amino acids together, noncovlent interactions help define the structure of each individual protein and serve to help hold the complementary structures together. (b) Small molecules serve as building blocks for larger structures. For example, to generate the information-carrying macromolecule DNA, the four small nucleotide building blocks deoxyadenylate (A), deoxythymidylate (T), deoxyguanylate (G), and deoxycytidylate (C) are covalently linked together into long strings (polymers), which then dimerize into the double helix. (c) Chemical reactions are reversible, and the distribution of the chemicals between starting compounds (*left*) and the products of the reactions (*right*) depends on the rate constants of the forward (k_f, upper arrow) and reverse (k_r, lower arrow) reactions. In the reaction shown, the forward reaction rate constant is faster than the reverse reaction, indicated by the thickness of the arrows. The ratio of these K_{eq}, provides an informative measure of the relative amounts of products and reactants that will be present at equilibrium. (d) In many cases, the source of energy for chemical reactions in cells is the hydrolysis of the molecule ATP. This energy is released when a high-energy phosphoanhydride bond linking the α and β or the β and γ phosphates in the ATP molecule (yellow) is broken by the addition of a water molecule. Proteins can efficiently transfer the energy of ATP hydrolysis to other chemicals, thus fueling other chemical reactions, or to other biomolecules for physical work.

principles of biochemical energetics, including the central role of ATP (adenosine triphosphate) in capturing and transferring energy in cellular metabolism.

2.1 Atomic Bonds and Molecular Interactions

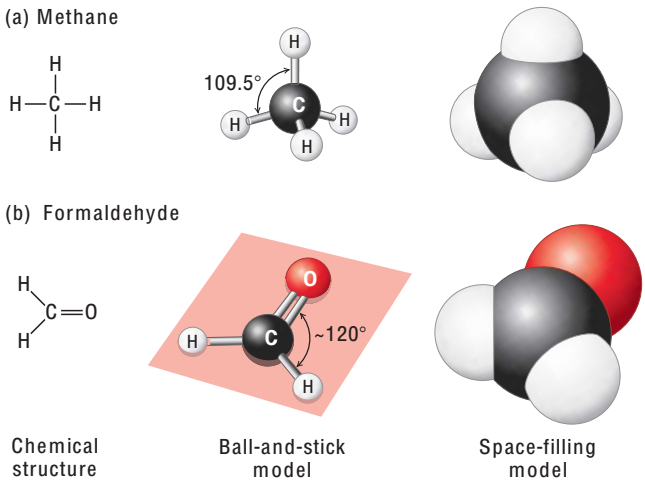
Strong and weak attractive forces between atoms are the glue that holds them together in individual molecules and permits interactions between different biological molecules. Strong forces form a covalent bond when two atoms share one pair of electrons ("single" bond) or multiple pairs of electrons ("double" bond, "triple" bond, etc.). The weak attractive forces of noncovalent interactions are equally important in determining the properties and functions of biomolecules such as proteins, nucleic acids, carbohydrates, and lipids. There are four major types of noncovalent interactions: ionic interactions, hydrogen bonds, van der Waals interactions, and the hydrophobic effect.

Each Atom Has a Defined Number and Geometry of Covalent Bonds

Hydrogen, oxygen, carbon, nitrogen, phosphorus, and sulfur are the most abundant elements found in biological molecules. These atoms, which rarely exist as isolated entities, readily form covalent bonds with other atoms, using electrons that reside in the outermost electron orbitals surrounding their nuclei. As a rule, each type of atom forms a

characteristic number of covalent bonds with other atoms, with a well-defined geometry determined by the atom's size and by both the distribution of electrons around the nucleus and the number of electrons that it can share. In some cases (e.g., carbon), the number of stable covalent bonds formed is fixed; in other cases (e.g., sulfur), different numbers of stable covalent bonds are possible.

All the biological building blocks are organized around the carbon atom, which normally forms four covalent bonds with two to four other atoms. As illustrated by the methane (CH_4) molecule, when carbon is bonded to four other atoms, the angle between any two bonds is 109.5° and the positions of bonded atoms define the four points of a tetrahedron (Figure 2-2a). This geometry helps define the structures of many biomolecules. A carbon (or any other) atom bonded to four dissimilar atoms or groups in a nonplanar configuration is said to be asymmetric. The tetrahedral orientation of bonds formed by an asymmetric carbon atom can be arranged in three-dimensional space in two different ways, producing molecules that are mirror images of each other, a property called *chirality*. Such molecules are called *optical isomers*, or stereoisomers. Many molecules in cells contain at least one asymmetric carbon atom, often called a chiral carbon atom. The different stereoisomers of a molecule usually have completely different biological activities because the arrangement of atoms within their structures differs, yielding their unique abilities to interact and chemically react with other molecules.



▲ FIGURE 2-2 Geometry of bonds when carbon is covalently linked to four or three other atoms. (a) If a carbon atom forms four single bonds, as in methane (CH_4), the bonded atoms (all H in this case) are oriented in space in the form of a tetrahedron. The letter representation on the left clearly indicates the atomic composition of the molecule and the bonding pattern. The ball-and-stick model in the center illustrates the geometric arrangement of the atoms and bonds, but the diameters of the balls representing the atoms and their nonbonding electrons are unrealistically small compared with the bond lengths. The sizes of the electron clouds in the space-filling model on the right more accurately represent the structure in three dimensions. (b) A carbon atom also can be bonded to three, rather than four, other atoms, as in formaldehyde (CH_2O). In this case, the carbon bonding electrons participate in two single bonds and one double bond, which all lie in the same plane. Unlike atoms connected by a single bond, which usually can rotate freely about the bond axis, those connected by a double bond cannot.

TABLE 2-1 Bonding Properties of Atoms Most Abundant in Biomolecules

Atom and Outer Electrons	Usual Number of Covalent Bonds	Bond Geometry
H	1	H
O	2	O
S	2, 4, or 6	S
N	3 or 4	N
P	5	P
C	4	C

Carbon can also bond to three other atoms in which all atoms are in a common plane. In this case, the carbon atom forms two typical single bonds with two atoms and a double bond (two shared electron pairs) with the third atom (Figure 2-2b). In the absence of other constraints, atoms joined by a single bond generally can rotate freely about the bond axis, while those connected by a double bond cannot. The rigid planarity imposed by double bonds has enormous significance for the shapes and flexibility of large biological molecules such as proteins and nucleic acids.

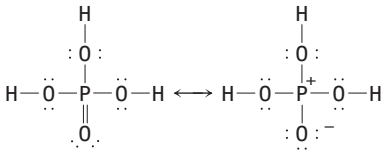
The number of covalent bonds formed by other common atoms is shown in Table 2-1. A hydrogen atom forms only one bond. An atom of oxygen usually forms only two covalent bonds, but has two additional pairs of electrons that can participate in noncovalent interactions. Sulfur forms two covalent bonds in hydrogen sulfide (H_2S), but also can accommodate six covalent bonds, as in sulfuric acid (H_2SO_4) and its sulfate derivatives. Nitrogen and phosphorus each have five electrons to share. In ammonia (NH_3), the nitrogen atom forms three covalent bonds; the pair of electrons around the atom not involved in a covalent bond can take part in noncovalent interactions. In the ammonium ion (NH_4^+), nitrogen forms four covalent bonds, which have a tetrahedral geometry. Phosphorus commonly forms five covalent bonds, as in phosphoric acid (H_3PO_4) and its phosphate derivatives, which form the backbone of nucleic acids. Phosphate groups attached to proteins play a key role in regulating the activity of many proteins (Chapter 3), and the central molecule in cellular energetics, ATP, contains three phosphate groups (see Section 2.4).

Electrons Are Shared Unequally in Polar Covalent Bonds

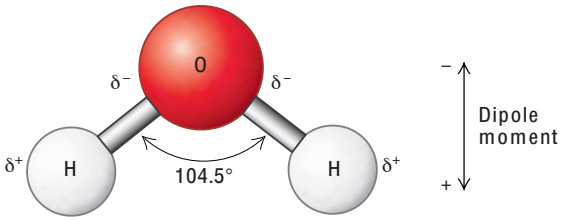
In many molecules, the bonded atoms exert different attractions for the electrons of the covalent bond, resulting in unequal sharing of the electrons. The extent of an atom's ability to attract an electron is called its **electronegativity**. A bond between atoms with identical or similar electronegativities is said to be **nonpolar**. In a nonpolar bond, the bonding electrons are essentially shared equally between the two atoms, as is the case for most C—C and C—H bonds. However, if two atoms differ in their electronegativities, the bond between them is said to be **polar**.

One end of a polar bond has a partial negative charge (δ^-), and the other end has a partial positive charge (δ^+). In an O—H bond, for example, the greater electronegativity of the oxygen atom relative to hydrogen results in the electrons spending more time around the oxygen atom than the hydrogen. Thus the O—H bond possesses an **electric dipole**, a positive charge separated from an equal but opposite negative charge. We can think of the oxygen atom of the O—H bond as having, on average, a charge of 25 percent of an electron, with the H atom having an equivalent positive charge. Because of its two O—H bonds, water molecules (H_2O) are dipoles that form electrostatic, noncovalent interactions with one another and with other molecules (Figure 2-3). These interactions play a critical role in almost every biochemical interaction and are thus fundamental to cell biology.

The polarity of the O═P double bond in H_3PO_4 results in a "resonance hybrid," a structure between the two forms shown below in which nonbonding electrons are shown as pairs of dots:



In the resonance hybrid on the right, one of the electrons from the P═O double bond has accumulated around the O atom, giving it a negative charge and leaving the P atom with a positive charge. These charges are important in noncovalent interactions.



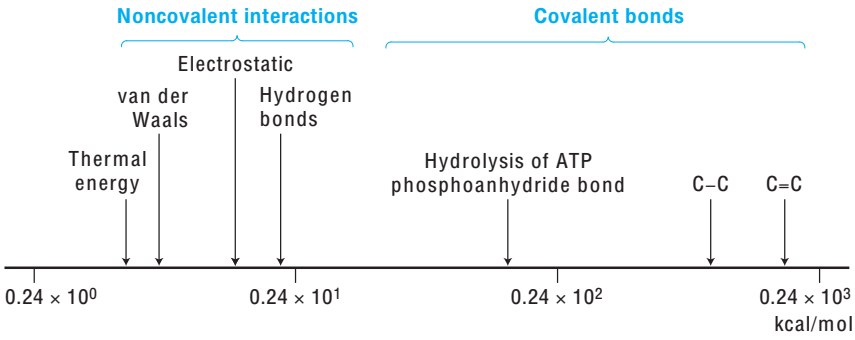
▲ **FIGURE 2-3 The dipole nature of a water molecule.** The symbol δ represents a partial charge (a weaker charge than the one on an electron or a proton). Because of the difference in the electronegativities of H and O, each of the polar H—O bonds in water has a dipole moment. The sizes and directions of the dipole moments of each of the bonds determine the net dipole moment of the molecule.

Covalent Bonds Are Much Stronger and More Stable Than Noncovalent Interactions

Covalent bonds are very stable because the energies required to break them are much greater than the thermal energy available at room temperature (25 °C) or body temperature (37 °C). For example, the thermal energy at 25 °C is approximately 0.6 kilocalorie per mole (kcal/mol), whereas the energy required to break the carbon-carbon single bond (C—C) in ethane is about 140 times larger (Figure 2-4). Consequently at room temperature (25 °C), fewer than 1 in 10^{12} ethane molecules is broken into a pair of $\cdot CH_3$ radicals, each containing an unpaired, nonbonding electron.

Covalent single bonds in biological molecules have energies similar to that of the C—C bond in ethane. Because more electrons are shared between atoms in double bonds, they require more energy to break than single bonds. For instance, it takes 84 kcal/mol to break a single C—O bond, but 170 kcal/mol to break a C═O double bond. The most common double bonds in biological molecules are C═O, C═N, C═C, and P═O.

The energy required to break noncovalent interactions is only 1–5 kcal/mol, much less than the bond energies of covalent bonds (see Figure 2-4). Indeed, noncovalent interactions are weak enough that they are constantly being



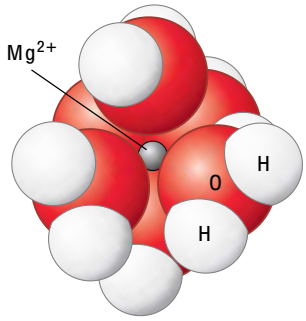
► **FIGURE 2-4 Relative energies of covalent bonds and noncovalent interactions.** Bond energies are determined as the energy required to break a particular type of linkage. Covalent bonds are one to two powers of 10 stronger than noncovalent interactions. The latter are somewhat greater than the thermal energy of the environment at normal room temperature (25 °C). Many biological processes are coupled to the energy released during hydrolysis of a phosphoanhydride bond in ATP.

formed and broken at room temperature. Although these interactions are weak and have a transient existence at physiological temperatures (25–37 °C), multiple noncovalent interactions can act together to produce highly stable and specific associations between different parts of a large molecule or between different macromolecules. We first review the four main types of noncovalent interactions and then consider their role in the binding of biomolecules to one another and to other molecules.

Ionic Interactions Are Attractions Between Oppositely Charged Ions

Ionic interactions result from the attraction of a positively charged ion—a **cation**—for a negatively charged ion—an **anion**. In sodium chloride (NaCl), for example, the bonding electron contributed by the sodium atom is completely transferred to the chlorine atom. Unlike covalent bonds, ionic interactions do not have fixed or specific geometric orientations, because the electrostatic field around an ion—its attraction for an opposite charge—is uniform in all directions.

In aqueous solutions, simple ions of biological significance, such as Na^+, K^+, Ca^{2+}, Mg^{2+}, and Cl^-, do not exist as free, isolated entities. Instead, each is hydrated, surrounded by a stable shell of water molecules, which are held in place by ionic interactions between the central ion and the oppositely charged end of the water dipole (Figure 2-5). Most ionic compounds dissolve readily in water because the energy of hydration, the energy released when ions tightly bind water molecules, is greater than the lattice energy that stabilizes the crystal structure. Parts or all of the aqueous hydration shell must be removed from ions when they directly interact with proteins. For example, water of hydration is lost when ions pass through protein pores in the cell membrane during nerve conduction (Chapter 7).



▲ FIGURE 2-5 Electrostatic interaction between water and a magnesium ion (Mg^{2+}). Water molecules are held in place by electrostatic interactions between the two positive charges on the ion and the partial negative charge on the oxygen of each water molecule. In aqueous solutions, all ions are surrounded by a similar hydration shell.

The relative strength of the interaction between two ions, A^- and C^+, depends on the concentration of other ions in a solution. The higher the concentration of other ions (e.g., Na^+ and Cl^-), the more opportunities A^- and C^+ have to interact ionically with these other ions, and thus the lower the energy required to break the interaction between A^- and C^+. As a result, increasing the concentrations of salts such as NaCl in a solution of biological molecules can weaken and even disrupt the ionic interactions holding the biomolecules together.

Hydrogen Bonds Determine Water Solubility of Uncharged Molecules

A **hydrogen bond** is the interaction of a partially positively charged hydrogen atom in a molecular dipole (e.g., water) with unpaired electrons from another atom, either in the same (intramolecular) or in a different (intermolecular) molecule. Normally, a hydrogen atom forms a covalent bond with only one other atom. However, a hydrogen atom covalently bonded to an electronegative donor atom D may form an additional weak association, the hydrogen bond, with an acceptor atom A, which must have a nonbonding pair of electrons available for the interaction:

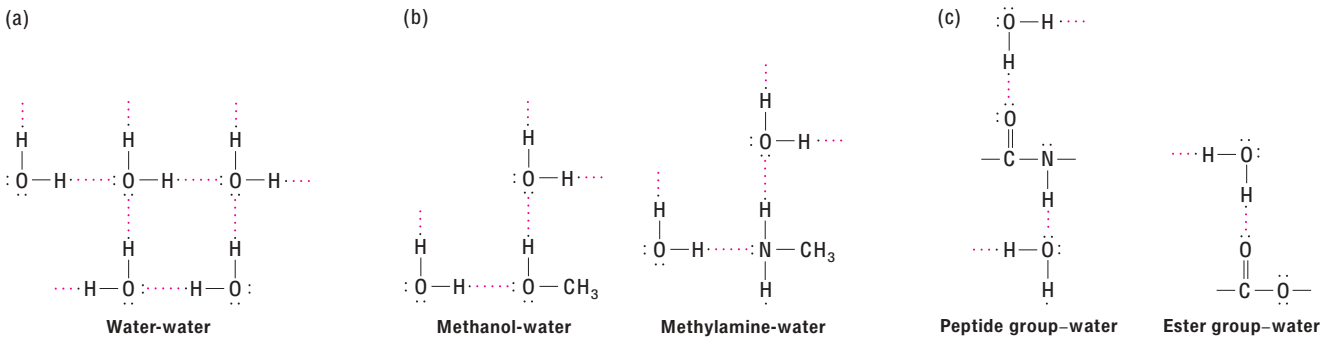
$$D^{\delta-}—H^{\delta+} + :A^{\delta-} \rightleftharpoons D^{\delta-}—H^{\delta+}\cdots\cdots:A^{\delta-}$$

Hydrogen bond

The length of the covalent D—H bond is a bit longer than it would be if there were no hydrogen bond, because the acceptor "pulls" the hydrogen away from the donor. An important feature of all hydrogen bonds is directionality. In the strongest hydrogen bonds, the donor atom, the hydrogen atom, and the acceptor atom all lie in a straight line. Nonlinear hydrogen bonds are weaker than linear ones; still, multiple nonlinear hydrogen bonds help to stabilize the three-dimensional structures of many proteins.

Hydrogen bonds are both longer and weaker than covalent bonds between the same atoms. In water, for example, the distance between the nuclei of the hydrogen and oxygen atoms of adjacent, hydrogen-bonded molecules is about 0.27 nm, about twice the length of the covalent O—H bonds within a single water molecule (Figure 2-6a). The strength of a hydrogen bond between water molecules (approximately 5 kcal/mol) is much weaker than a covalent O—H bond (roughly 110 kcal/mol), although it is greater than that for many other hydrogen bonds in biological molecules (1–2 kcal/mol). The extensive hydrogen bonding between water molecules accounts for many of the key properties of this compound, including its unusually high melting and boiling points and its ability to interact with many other molecules.

The solubility of uncharged substances in an aqueous environment depends largely on their ability to form hydrogen bonds with water. For instance, the hydroxyl group (—OH) in methanol (CH_3OH) and the amino group (—NH_2) in methylamine (CH_3NH_2) can form several hydrogen bonds with water, enabling these molecules to dissolve in water to



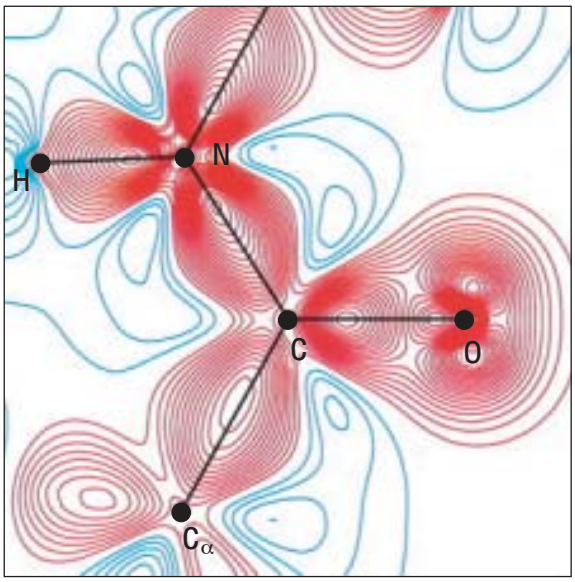
▲ FIGURE 2-6 Hydrogen bonding of water with itself and with other compounds. Each pair of nonbonding outer electrons in an oxygen or nitrogen atom can accept a hydrogen atom in a hydrogen bond. The hydroxyl and the amino groups can also form hydrogen bonds with water. (a) In liquid water, each water molecule apparently forms transient hydrogen bonds with several others, creating a dynamic network of hydrogen-bonded molecules. (b) Water also can form hydrogen bonds with methanol and methylamine, accounting for the high solubility of these compounds. (c) The peptide group and ester group, which are present in many biomolecules, commonly participate in hydrogen bonds with water or polar groups in other molecules.

high concentrations (Figure 2-6b). In general, molecules with polar bonds that easily form hydrogen bonds with water can readily dissolve in water; that is, they are hydrophilic. Many biological molecules contain, in addition to hydroxyl and amino groups, peptide and ester groups, which form hydrogen bonds with water (Figure 2-6c). X-ray crystallography combined with computational analysis permits an accurate depiction of the distribution of electrons in covalent bonds and the outermost unbonded electrons of atoms, as illustrated in Figure 2-7. These unbonded electrons can form hydrogen bonds with donor hydrogens.

Van der Waals Interactions Are Caused by Transient Dipoles

When any two atoms approach each other closely, they create a weak, nonspecific attractive force called a van der Waals interaction. These nonspecific interactions result from the momentary random fluctuations in the distribution of the electrons of any atom, which give rise to a transient unequal distribution of electrons. If two noncovalently bonded atoms are close enough together, electrons of one atom will perturb the electrons of the other. This perturbation generates a transient dipole in the second atom, and the two dipoles will attract each other weakly (Figure 2-8). Similarly, a polar covalent bond in one molecule will attract an oppositely oriented dipole in another.

Van der Waals interactions, involving either transiently induced or permanent electric dipoles, occur in all types of molecules, both polar and nonpolar. In particular, van der Waals interactions are responsible for the cohesion between molecules of nonpolar liquids and solids, such as heptane, $CH_3—(CH_2)_5—CH_3$, that cannot form hydrogen bonds or ionic interactions with other molecules. The strength of van der Waals interactions decreases rapidly with increasing distance; thus these noncovalent bonds can form only when



▲ FIGURE 2-7 Distribution of bonding and outer nonbonding electrons in the peptide group. Shown here is one amino acid within a protein called crambin. The black lines diagrammatically represent the covalent bonds between atoms. The red (negative) and blue (positive) lines represent contours of charge. The greater the number of contour lines, the higher the charge. The high density of red contour lines between atoms represents the covalent bonds (shared electron pairs). The two sets of red contour lines emanating from the oxygen (O) and not falling on a covalent bond (black line) represent the two pairs of nonbonded electrons on the oxygen that are available to participate in hydrogen bonding. The high density of blue contour lines near the hydrogen (H) bonded to nitrogen (N) represents a partial positive charge, indicating that this H can act as a donor in hydrogen bonding. [From C. Jelsch et al., 2000, *Proc. Nat'l. Acad. Sci.* USA **97:**3171. Courtesy of M. M. Teeter.]