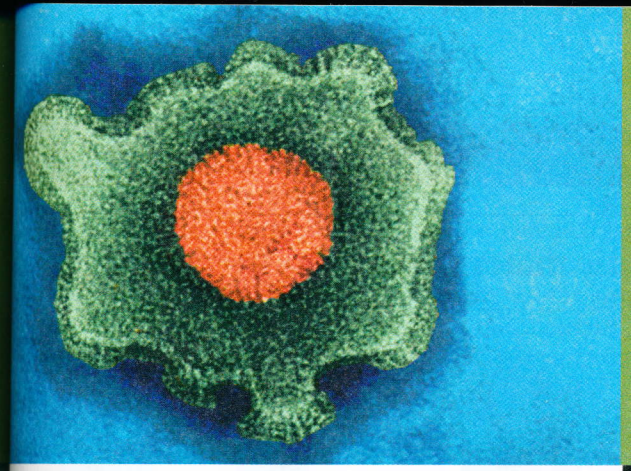

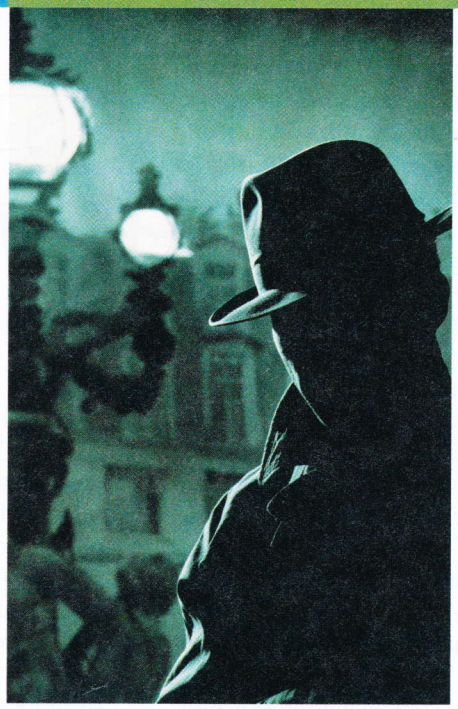

Sabotage Inside Our Cells

A saboteur drifts stealthily toward its target: a vital factory. Stopped at the perimeter by a guard, the intruder presents counterfeit identification to gain entry. Once inside, it surveys the scene and makes a quick decision: The time is not yet ripe for sabotage. So it lies low and waits silently, undetected, until it receives a go signal. The intruder now acts quickly, hijacking the factory machinery and diverting production to its own diabolical ends. Controlled by the intruder, the factory now manufactures replicas of the saboteur! When these duplicates are ready, they break their way out, destroying the factory as they exit. With the ruins behind them, they move off silently in search of new targets.

The scenario just described is played out millions of times each year. What is it? Industrial sabotage? Military espionage? In fact, the saboteur in this story is a herpesvirus, the type of virus that causes cold sores, genital herpes, chicken pox, and a number of other diseases. The colorized transmission electron microscope image above (at a magnification of 250,000×) and the computer model to the left show the structure of the herpesvirus.

Viruses share some of the characteristics of living organisms, such as genetic material in the form of nucleic acid packaged within a highly organized structure. A virus is generally not considered alive, however, because it is not cellular and cannot reproduce on its own. A **virus** is simply nucleic acid wrapped in a coat of protein and, for herpesviruses and some other animal viruses, a membranous envelope. Although a herpesvirus is fairly large as viruses go—about 200 nm across—its diameter is less than $\frac{1}{100}$ that of a typical human cell. Just about all a herpesvirus or any other virus can do is infect a host. It is the host that provides most of the tools and raw materials needed to duplicate the virus.

Once in the body, a herpesvirus tumbles along until it finds a suitable target cell, recognized when protein molecules on the outside of the virus fit into protein receptor molecules on the surface of the cell. Not recognizing the threat, the cell takes in the virus. Once inside the cell, the DNA of the herpesvirus enters the nucleus. In the nuclei of certain nerve cells, the viral DNA can remain dormant for long periods of time, until activated by a signal such as cellular stress. When activated, the viral DNA hijacks the cell's own molecules and organelles to produce new copies of the virus. Virus production eventually results in destruction of host cells—causing the sores that are characteristic of herpes diseases. The released viruses can then infect other cells.

Once a person is infected with a herpesvirus, the virus remains permanently latent (dormant) in the body, its DNA integrated into the chromosomes of nerve cells. Although many people never develop symptoms, over 75% of American adults are thought to carry herpes simplex 1 (which causes cold sores), and over 20%, herpes simplex 2 (which causes genital herpes). Herpesviruses are somewhat unusual in being able to remain latent inside our cells. Another virus with this ability is HIV, the virus that causes AIDS.

Because viruses have much less complex structures than cells, they are relatively easy to study on the molecular level, far easier than Mendel's peas or Morgan's fruit flies. For this reason, we owe our first glimpses of the functions of DNA, the molecule that controls hereditary traits, to the study of viruses.

This chapter is about **molecular biology,** the study of DNA and how it serves as the chemical basis of heredity. Here we explore the structure of DNA, how it replicates (the molecular basis of why offspring resemble their parents), and how it controls the cell by directing RNA and protein synthesis. We also look at viruses that infect bacteria, animals, and plants. We end with an examination of bacterial genetics. To start the chapter, we recount the story of how we know that DNA is the genetic material, a story in which a virus played a major role. ■ ■ ■

10.1 Experiments showed that DNA is the genetic material

Today, even schoolchildren have heard of DNA, and scientists routinely manipulate DNA in the laboratory and use it to change the heritable characteristics of cells. Early in the 20th century, however, the precise identity of the molecules of inheritance was unknown. Biologists knew that genes are located on chromosomes. Therefore, the two chemical components of chromosomes—DNA and protein—were the candidates for the genetic material. Until the 1940s, the case for proteins seemed stronger because proteins appeared to be more structurally complex and functionally specific. Biologists finally established the role of DNA in heredity through studies involving bacteria and the viruses that infect them.

We can trace the discovery of the genetic role of DNA back to 1928. British medical officer Frederick Griffith was studying two strains of a bacterium: a harmless strain and a pathogenic (disease-causing) strain that causes pneumonia. Griffith was surprised to find that when he killed the pathogenic bacteria and then mixed the bacterial remains with living harmless bacteria, some living bacterial cells were converted to the disease-causing form. Furthermore, all of the descendants of the transformed bacteria inherited the newly acquired ability to cause disease. Clearly, some chemical component of the dead bacteria could act as a "transforming factor" that brought about a heritable change.

Most biologists doubted that DNA could be Griffith's transforming factor, primarily because so little was known about DNA. However, in 1952, American biologists Alfred Hershey and Martha Chase performed a very convincing set of experiments. They showed that DNA is the genetic material of a virus called T2, which infects the bacterium *Escherichia coli* (*E. coli*). Bacterial viruses are called **bacteriophages** ("bacteria-eaters"), or **phages** for short. Figure 10.1A shows the structure of phage T2, which consists solely of DNA (blue) and protein (yellow). Resembling a lunar landing craft, T2 has a DNA-containing head and a hollow tail with six jointed fibers extending from it. The fibers attach to the surface of a susceptible bacterium. Hershey and Chase knew that T2 could reprogram its host cell to produce new phages, but they did not know which component—DNA or protein—was responsible for this ability.

Hershey and Chase found the answer by devising an experiment to determine what kinds of molecules the phage transferred to *E. coli* during infection. Their experiment used only a few relatively simple tools: chemicals containing radioactive isotopes (see Module 2.5); a radioactivity detector; a kitchen blender; and a centrifuge, a device that spins test tubes to separate particles of different weights. (These are still basic tools of molecular biology.)

Hershey and Chase used different radioactive isotopes to label the DNA and protein in T2. First, they grew T2 with *E. coli* in a solution containing radioactive sulfur (yellow in Figure 10.1B). Protein contains sulfur but DNA does not, so as new phages were made, the radioactive sulfur atoms were

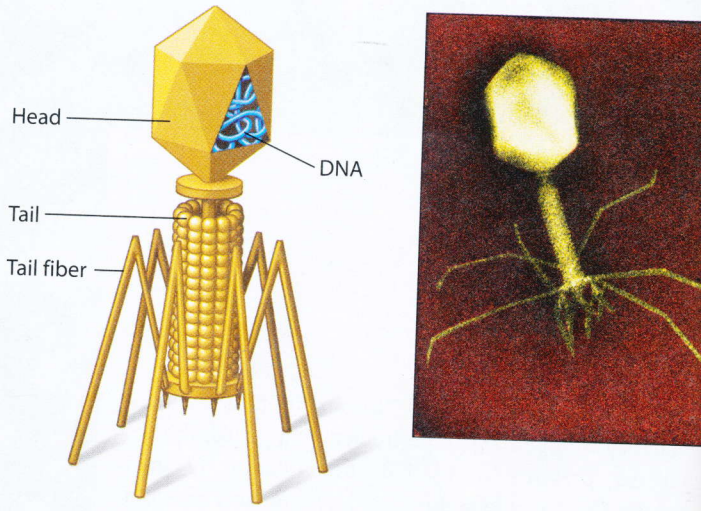

Figure 10.1A Phage T2

incorporated only into their proteins. The researchers grew a separate batch of phages in a solution containing radioactive phosphorus (green). Because nearly all the phage's phosphorus is in DNA, this labeled only the phage DNA.

Armed with the two batches of labeled T2, Hershey and Chase were ready to perform the experiment outlined in Figure 10.1B. ❶ They allowed the two batches of T2 to infect separate samples of nonradioactive bacteria. ❷ Shortly after the onset of infection, they agitated the cultures in a blender to shake loose any parts of the phages that remained outside the bacterial cells. ❸ They then spun the mixtures in a centrifuge. The cells were deposited as a pellet at the bottom of the centrifuge tubes, but phages and parts of phages, being lighter, remained suspended in the liquid. ❹ The researchers then measured the radioactivity in the pellet and the liquid.

Hershey and Chase found that when the bacteria had been infected with T2 phages containing labeled protein, the radioactivity ended up mainly in the liquid, which contained phages but not bacteria. This result suggested that the phage protein did not enter the cells. But when the bacteria had been infected with phages whose DNA was tagged, then most of the radioactivity was in the bacteria pellet. When these bacteria were returned to liquid growth medium, the bacterial cells were soon destroyed, lysing (breaking open) and releasing new phages that contained radioactive phosphorus in their DNA but no radioactive sulfur in their proteins.

Hershey and Chase concluded that T2 injects its DNA into the host cell, leaving virtually all its protein outside (as shown in Figure 10.1B). More importantly, they demonstrated that it is the injected DNA molecules that cause the cells to produce additional phage DNA and proteins—indeed, new complete phages. This indicated that the DNA contained the instructions

for making phages. Figure 10.1C outlines the reproductive cycle for phage T2 as we now understand it.

The Hershey-Chase results, added to earlier evidence, convinced most scientists that DNA is the hereditary material. What happened next was one of the most celebrated quests in the history of science: the effort to figure out the structure of DNA and how this structure enables the molecule to store genetic information and transmit it from parents to offspring.

Web Activity *The Hershey-Chase Experiment*

Web Activity *Phage T2 Reproductive Cycle*

? What convinced Hershey and Chase that DNA, rather than protein, is the genetic material of phage T2?

Radioactively labeled phage DNA, but not labeled protein, entered the cell during infection and directed the synthesis of new viruses.

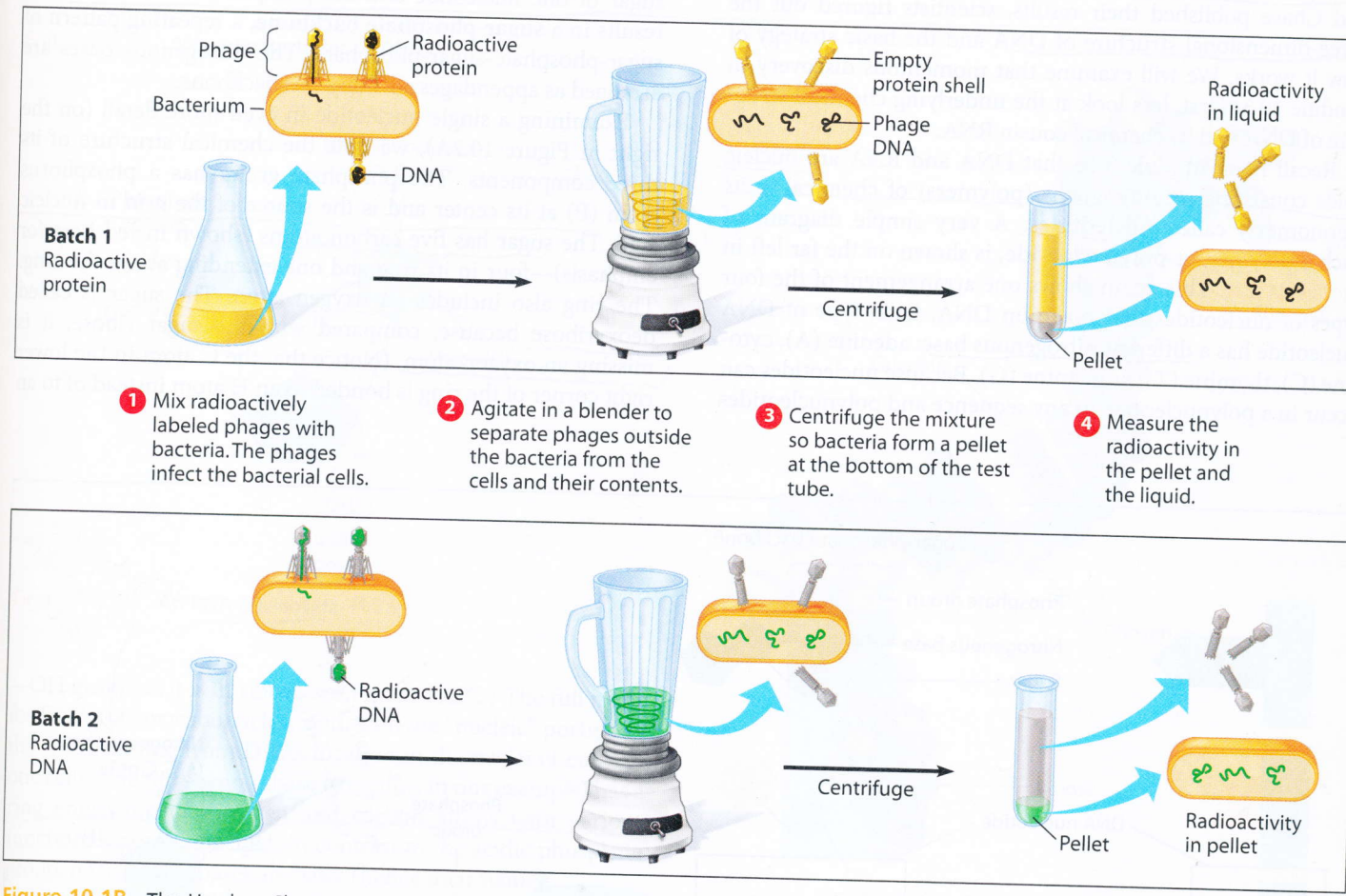

Figure 10.1B The Hershey-Chase experiment

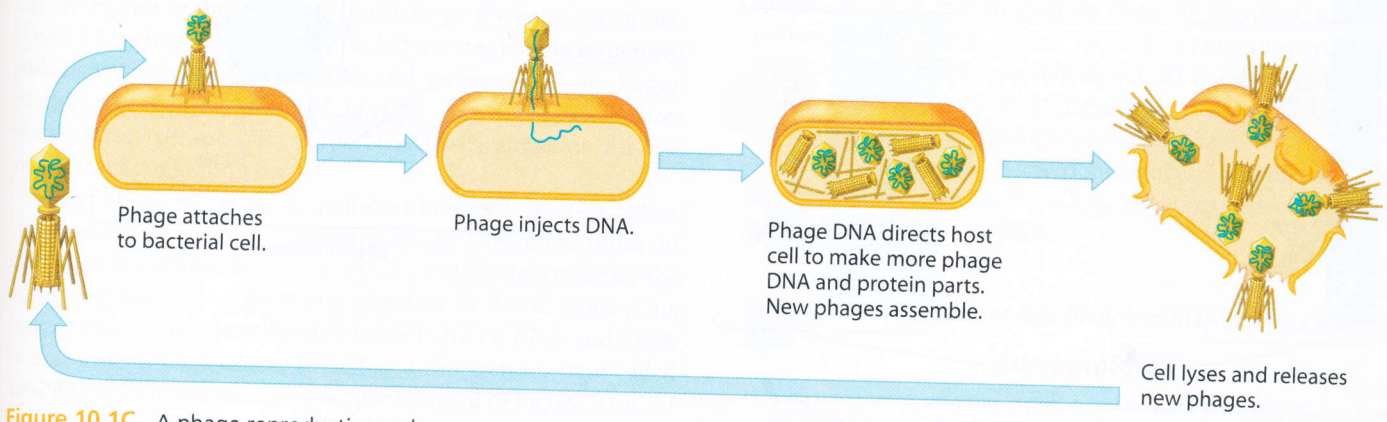

Figure 10.1C A phage reproductive cycle

10.2 DNA and RNA are polymers of nucleotides

By the time Hershey and Chase performed their experiments, much was already known about DNA. Scientists had identified all its atoms and knew how they were covalently bonded to one another. What was not understood was the specific arrangement of atoms that gave DNA its unique properties—the capacity to store genetic information, copy it, and pass it from generation to generation. However, only one year after Hershey and Chase published their results, scientists figured out the three-dimensional structure of DNA and the basic strategy of how it works. We will examine that momentous discovery in Module 10.3. First, let's look at the underlying chemical structure of DNA and its chemical cousin RNA.

Recall from Module 3.16 that DNA and RNA are nucleic acids consisting of long chains (polymers) of chemical units (monomers) called **nucleotides.** A very simple diagram of such a polymer, or **polynucleotide,** is shown on the far left in Figure 10.2A. This chain shows one arrangement of the four types of nucleotides that make up DNA. Each type of DNA nucleotide has a different nitrogenous base: adenine (A), cytosine (C), thymine (T), or guanine (G). Because nucleotides can occur in a polynucleotide in any sequence and polynucleotides vary in length from long to very long, the number of possible polynucleotides is enormous.

Looking more closely at our polynucleotide, we see in the center of Figure 10.2A that each nucleotide consists of three components: a nitrogenous base (in DNA, A, C, T, or G), a sugar (blue), and a phosphate group (yellow). The nucleotides are joined to one another by covalent bonds between the sugar of one nucleotide and the phosphate of the next. This results in a **sugar-phosphate backbone,** a repeating pattern of sugar-phosphate-sugar-phosphate. The nitrogenous bases are arranged as appendages all along this backbone.

Examining a single nucleotide in even more detail (on the right in Figure 10.2A), we note the chemical structure of its three components. The phosphate group has a phosphorus atom (P) at its center and is the source of the *acid* in nucleic acid. The sugar has five carbon atoms (shown in red here for emphasis)—four in its ring and one extending above the ring. The ring also includes an oxygen atom. The sugar is called deoxyribose because, compared with the sugar ribose, it is missing an oxygen atom. (Notice that the C atom in the lower right corner of the ring is bonded to an H atom instead of to an

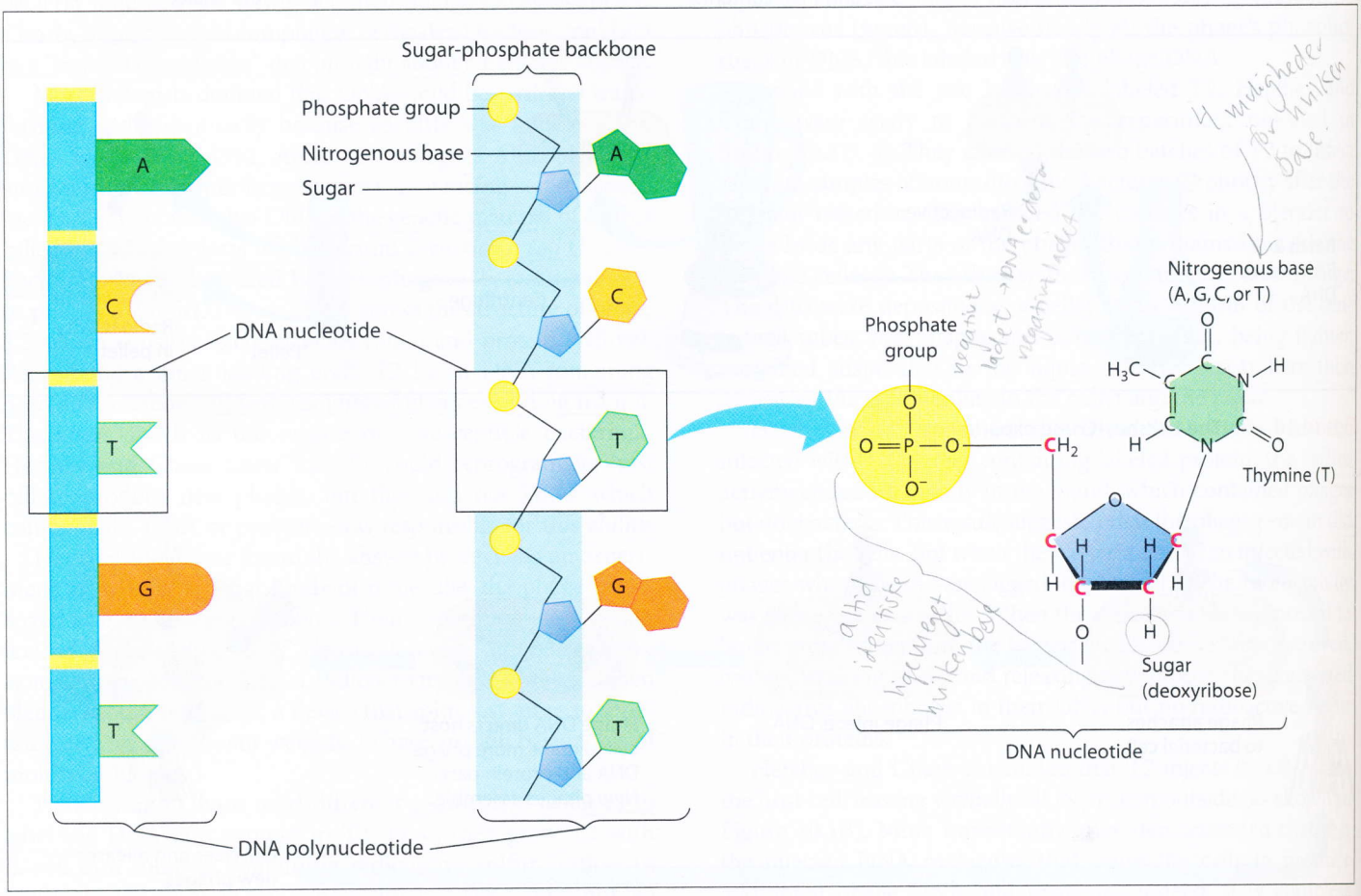

Figure 10.2A The structure of a DNA polynucleotide

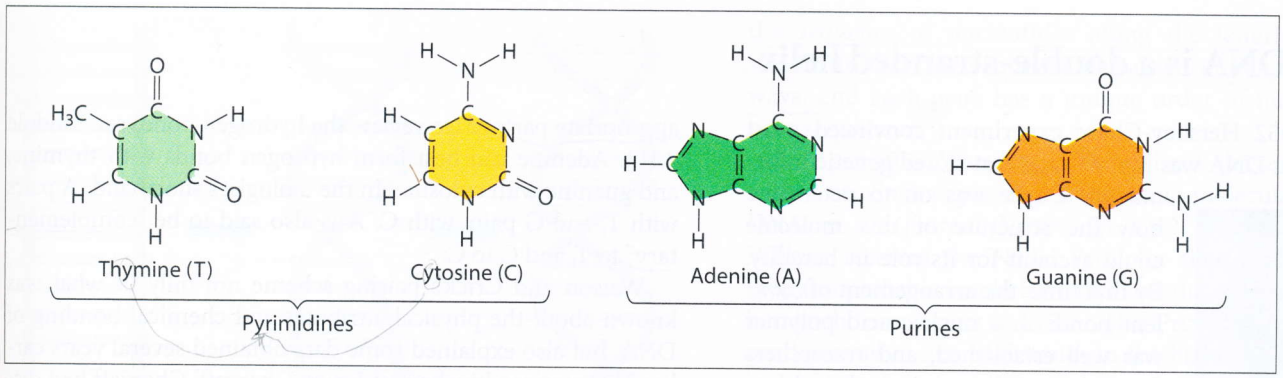

Figure 10.2B Nitrogenous bases of DNA

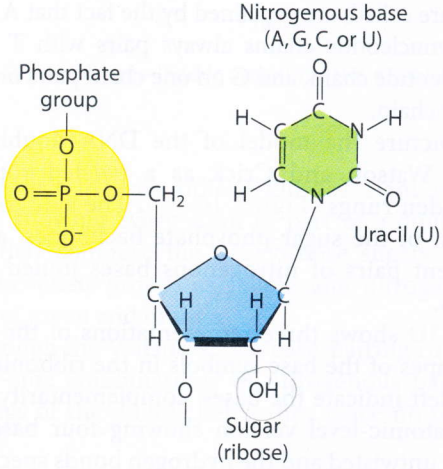

Figure 10.2C An RNA nucleotide

—OH group, as it is in ribose; see Figure 10.2C.) The full name for DNA is deoxyribonucleic acid, with the "nucleic" portion of the word coming from DNA's location in the nuclei of eukaryotic cells. The nitrogenous base (thymine, in our example) has a ring consisting of nitrogen and carbon atoms with various functional groups attached. In contrast to the acidic phosphate group, nitrogenous bases are basic (hence their name).

The four nucleotides found in DNA differ only in their nitrogenous bases. **Figure 10.2B** shows the structures of the four nitrogenous bases in DNA. At this point, the structural details are not as important as the fact that the bases are of two types. **Thymine (T)** and **cytosine (C)** are single-ring structures called *pyrimidines*. **Adenine (A)** and **guanine (G)** are larger, double-ring structures called *purines*. The one-letter abbreviations can be used for either the bases alone or for the nucleotides containing them.

What about RNA? As its name—ribonucleic acid—implies, its sugar is ribose rather than deoxyribose. Notice the ribose in the RNA nucleotide in **Figure 10.2C**; unlike deoxyribose, the sugar ring has an —OH group attached to the C atom at its lower-right corner. Another difference between RNA and DNA is that instead of thymine, RNA has a nitrogenous base called **uracil (U)**. (You can see the structure of uracil in Figure 10.2C; it is very similar to thymine.) Except for the presence of ribose and

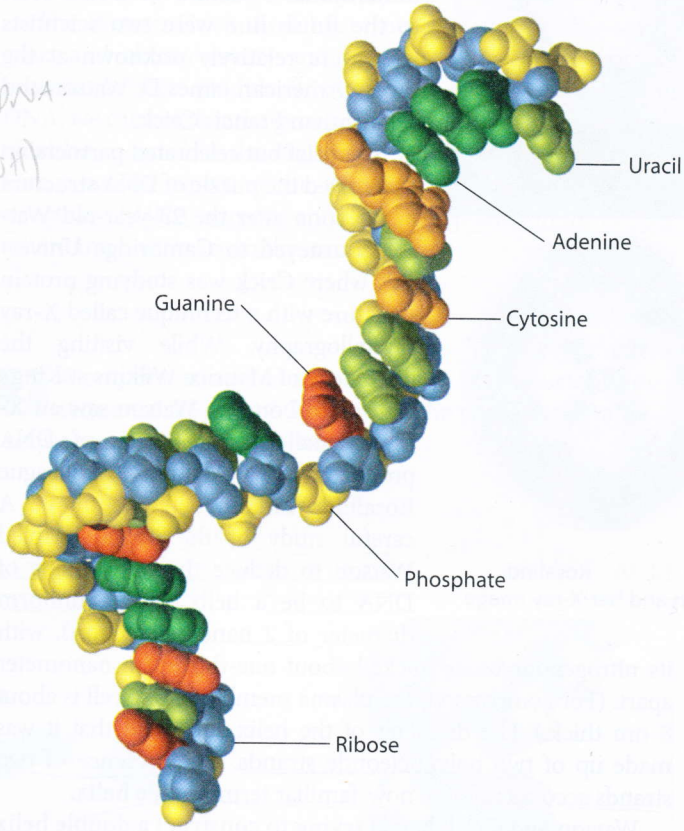

Figure 10.2D Part of an RNA polynucleotide

uracil, an RNA polynucleotide chain is identical to a DNA polynucleotide chain. **Figure 10.2D** is a computer graphic of a piece of RNA polynucleotide about 20 nucleotides long. The yellow phosphorus atoms at the center of the phosphate groups makes it easy to spot the sugar-phosphate backbone.

In this module, we reviewed the structure of a polynucleotide. In the next module, we'll see how two polynucleotides join together in a molecule of DNA.

? Compare and contrast DNA and RNA polynucleotides.

■ Both are polymers of nucleotides consisting of a sugar + a nitrogenous base + a phosphate. In RNA, the sugar is ribose; in DNA, it is deoxyribose. Both RNA and DNA have the bases A, G, and C, but DNA has a T and RNA has a U.

10.3 DNA is a double-stranded helix

After the 1952 Hershey-Chase experiment convinced most biologists that DNA was the material that stored genetic information, a race was on to determine how the structure of this molecule could account for its role in heredity. By that time, the arrangement of covalent bonds in a nucleic acid polymer was well established, and researchers focused on discovering the three-dimensional structure of DNA. First to the finish line were two scientists who were relatively unknown at the time—American James D. Watson and Englishman Francis Crick.

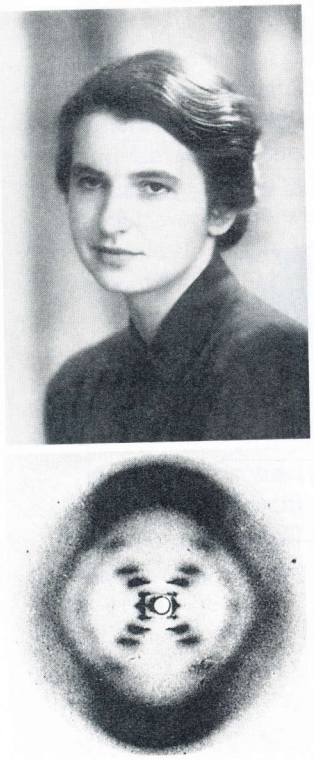

Figure 10.3A Rosalind Franklin and her X-ray image

The brief but celebrated partnership that solved the puzzle of DNA structure began soon after the 23-year-old Watson journeyed to Cambridge University, where Crick was studying protein structure with a technique called X-ray crystallography. While visiting the laboratory of Maurice Wilkins at King's College in London, Watson saw an X-ray crystallographic image of DNA produced by Wilkins's colleague Rosalind Franklin (Figure 10.3A). A careful study of the image enabled Watson to deduce the basic shape of DNA to be a helix with a uniform diameter of 2 nanometers (nm), with its nitrogenous bases stacked about one-third of a nanometer apart. (For comparison, the plasma membrane of a cell is about 8 nm thick.) The diameter of the helix suggested that it was made up of two polynucleotide strands. The presence of two strands accounts for the now-familiar term **double helix.**

Watson and Crick began trying to construct a double helix that would conform both to Franklin's data and to what was then known about the chemistry of DNA (Figure 10.3B). Franklin had concluded that the sugar-phosphate backbones must be on the outside of the double helix, forcing the nitrogenous bases to swivel to the interior of the molecule. But how were the bases arranged in the interior of the double helix?

At first, Watson and Crick imagined that the bases paired like with like—for example, A with A and C with C. But that kind of pairing did not fit the X-ray data, which suggested that the DNA molecule has a *uniform* diameter. An AA pair would be almost twice as wide as a CC pair, causing bulges in the molecule. It soon became apparent that a double-ringed base (purine) must always be paired with a single-ringed base (pyrimidine) on the opposite strand. Moreover, Watson and Crick realized that the individual structures of the bases dictated the pairings even more specifically. Each base has chemical side groups that can best form hydrogen bonds with one appropriate partner (to review the hydrogen bond, see Module 2.10). Adenine can best form hydrogen bonds with thymine, and guanine with cytosine. In the biologist's shorthand, A pairs with T, and G pairs with C. A is also said to be "complementary" to T, and G to C.

Watson and Crick's pairing scheme not only fit what was known about the physical attributes and chemical bonding of DNA, but also explained some data obtained several years earlier by American biochemist Erwin Chargaff. Chargaff had discovered that the amount of adenine in the DNA of any one species was equal to the amount of thymine and that the amount of guanine was equal to that of cytosine. Chargaff's rules, as they are called, are explained by the fact that A on one of DNA's polynucleotide chains always pairs with T on the other polynucleotide chain, and G on one chain pairs only with C on the other chain.

You can picture the model of the DNA double helix proposed by Watson and Crick as a twisted rope ladder with wooden rungs (Figure 10.3C). The side ropes are the equivalent of the sugar-phosphate backbones, and the rungs represent pairs of nitrogenous bases joined by hydrogen bonds.

Figure 10.3D shows three representations of the double helix. The shapes of the base symbols in the ribbonlike diagram on the left indicate the bases' complementarity. In the center is an atomic-level version showing four base pairs, with the helix untwisted and the hydrogen bonds specified by dotted lines. You can see that the two sugar-phosphate backbones of the double helix are oriented in opposite directions. (Notice that the sugars on the two strands are upside down

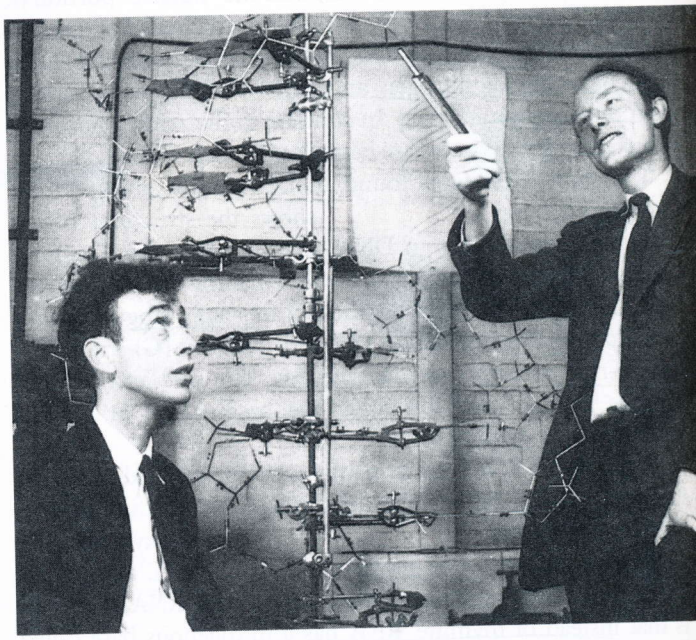

Figure 10.3B Watson and Crick in 1953 with their model of the DNA double helix

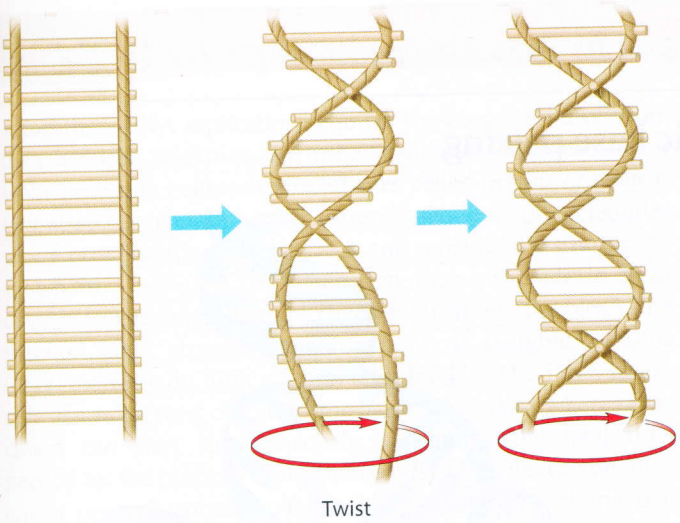

Figure 10.3C A rope-ladder model for the double helix

with respect to each other.) On the right is a computer graphic showing every atom of part of a double helix. The atoms that compose the deoxyribose sugars are shown as blue, phosphate groups as yellow, and nitrogenous bases as shades of green and orange.

Although the Watson-Crick base-pairing rules dictate the side-by-side combinations of nitrogenous bases that form the rungs of the double helix, they place no restrictions on the *sequence* of nucleotides along the length of a DNA strand. In fact, the sequence of bases can vary in countless ways, and each gene has a unique order of nucleotides, or base sequence.

In April 1953, Watson and Crick shook the scientific world with a succinct paper explaining their molecular model for DNA in the journal *Nature*. In 1962, Watson, Crick, and Wilkins received the Nobel Prize for their work. (Rosalind Franklin probably would have received the prize as well but for her death from cancer in 1958; Nobel Prizes are never awarded posthumously.) Few milestones in the history of biology have had as broad an impact as the discovery of the double helix, with its AT and CG base pairing.

The Watson-Crick model gave new meaning to the words *genes* and *chromosomes*—and to the chromosome theory of inheritance (see Module 9.16). With a complete picture of DNA, we can see that the genetic information in a chromosome must be encoded in the nucleotide sequence of the molecule. One powerful aspect of the Watson-Crick model is that the structure of DNA suggests a molecular explanation for genetic inheritance, as we see in the next module.

Web Activity *DNA and RNA Structure*

Web Activity *DNA Double Helix*

? Along one strand of a double helix is the nucleotide sequence GGCATAGGT. What is the complementary sequence for the other DNA strand?

ACCGTATCCA ■

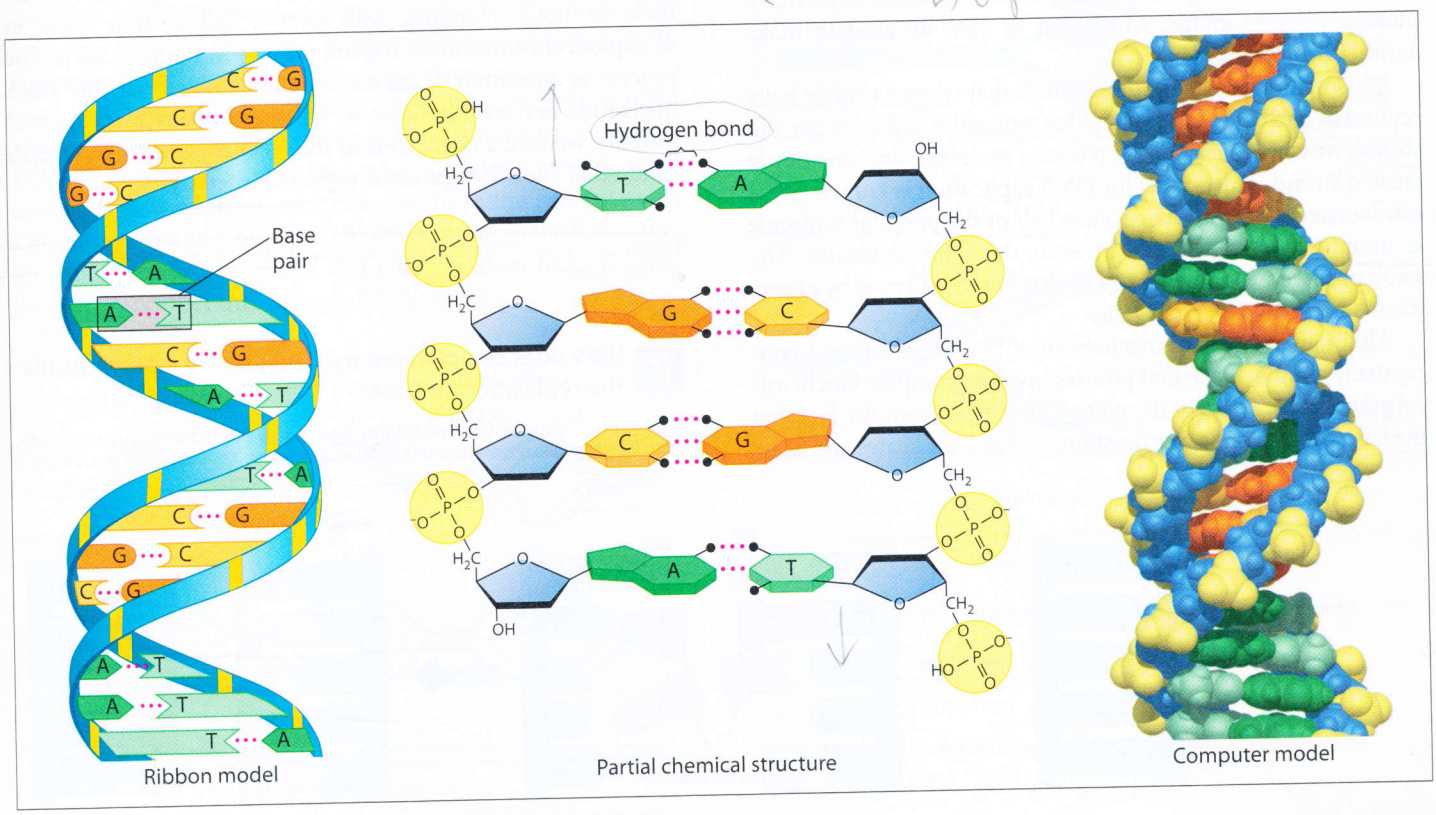

Figure 10.3D Three representations of DNA

10.4 DNA replication depends on specific base pairing

One of biology's overarching themes—the relationship between structure and function—is evident in the double helix. The idea that there is specific pairing of bases in DNA was the flash of inspiration that led Watson and Crick to the correct structure of the double helix. At the same time, they saw the functional significance of the base-pairing rules. They ended their classic 1953 paper with this statement: "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material."

The logic behind the Watson-Crick proposal for how DNA is copied—by specific pairing of complementary bases—is quite simple. You can see this by covering one of the strands in the parental DNA molecule in Figure 10.4A. You can determine the sequence of bases in the covered strand by applying the base-pairing rules to the unmasked strand: A pairs with T, G with C.

Watson and Crick predicted that a cell applies the same rules when copying its genes. As shown in Figure 10.4A, the two strands of parental DNA (blue) separate. Each then becomes a template for the assembly of a complementary strand from a supply of free nucleotides (gray). The nucleotides line up one at a time along the template strand in accordance with the base-pairing rules. Enzymes link the nucleotides to form the new DNA strands. The completed new molecules, identical to the parental molecule, are known as daughter DNA. The copying mechanism is analogous to using a photographic negative to make a positive image, which can in turn be used to make another negative, and so on.

Watson and Crick's model predicts that when a double helix replicates, each of the two daughter molecules will have one old strand, which was part of the parental molecule, and one newly created strand. This model for DNA replication is known as the **semiconservative model** because half of the parental molecule is maintained (conserved) in each daughter molecule. The semiconservative model of replication was confirmed by experiments performed in the 1950s.

Although the general mechanism of DNA replication is conceptually simple, the actual process involves complex biochemical gymnastics. Some of the complexity arises from the fact that the helical DNA molecule must untwist as it replicates and must copy its two strands roughly simultaneously (Figure 10.4B). Another challenge is the speed of the process. *E.coli*, with about 4.6 million DNA base pairs, can copy its entire genome in less than an hour. Humans, with over 6 billion base pairs in 46 diploid chromosomes, require only a few hours. And yet the process is amazingly accurate; typically, only about one DNA nucleotide per several billion is incorrectly paired. In the next module, we take a closer look at the mechanisms of DNA replication that allow it to proceed with such speed and accuracy.

Figure 10.4B Untwisting and replication of DNA

Web Process of Science ***What Is the Correct Model for DNA Replication?***

? How does complementary base pairing make possible the replication of DNA?

■ When the two strands of the double helix separate, each serves as a "mold" for the base-pairing of the new complementary strands.

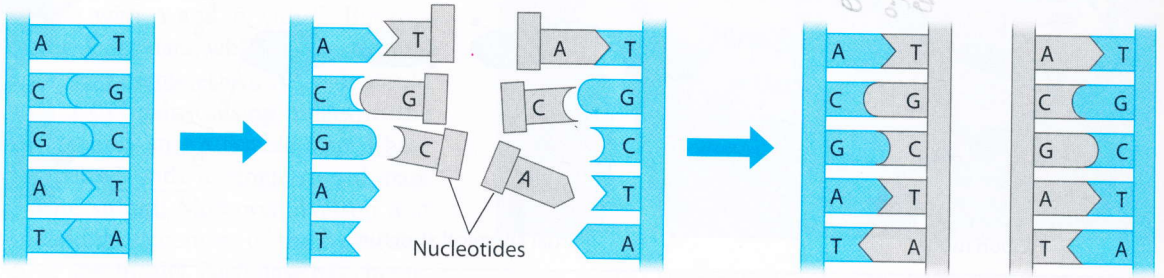

Figure 10.4A A template model for DNA replication

Parental molecule of DNA

Both parental strands serve as templates

Two identical daughter molecules of DNA

10.5 DNA replication proceeds in two directions at many sites simultaneously

Altogether, DNA replication requires the cooperation of more than a dozen enzymes and other proteins. Replication of a DNA molecule begins at special sites called *origins of replication*, stretches of DNA having a specific sequence of nucleotides where proteins attach to the DNA and separate the strands. As shown in Figure 10.5A, replication then proceeds in both directions, creating what are called replication "bubbles." The parental DNA strands (blue) open up as daughter strands (gray) elongate on both sides of each bubble. The DNA molecule of a eukaryotic chromosome has many origins where replication can start simultaneously, shortening the total time needed for the process. Thus, thousands of bubbles can be present at once. Eventually, all the bubbles merge, yielding two completed daughter DNA molecules.

Figure 10.5B shows the molecular building blocks of a tiny segment of DNA, reminding us that the DNA's sugar-phosphate backbones run in opposite directions. Notice that each strand has a 3′ ("three-prime") end and a 5′ end. The primed numbers refer to the carbon atoms of the nucleotide sugars. At one end of each DNA strand, the sugar's 3′ carbon atom is attached to an —OH group; at the other end, the sugar's 5′ carbon has a phosphate group.

The opposite orientation of the strands is important in DNA replication. The enzymes that link DNA nucleotides to a growing daughter strand, called **DNA polymerases,** add nucleotides only to the 3′ end of the strand, never to the 5′ end. Thus, a daughter DNA strand can only grow in the 5′ → 3′ direction. You see the consequences of this enzyme specificity in Figure 10.5C. The forked structure represents one side of a replication bubble. One of the daughter strands (shown in gray) can be synthesized in one continuous piece by a DNA polymerase working toward the forking point of the parental DNA. However, to make the other daughter strand, polymerase molecules must work outward from the forking point. This new strand is synthesized in short pieces as the fork opens up. Another enzyme, called **DNA ligase,** then links (ligates) the pieces together into a single DNA strand.

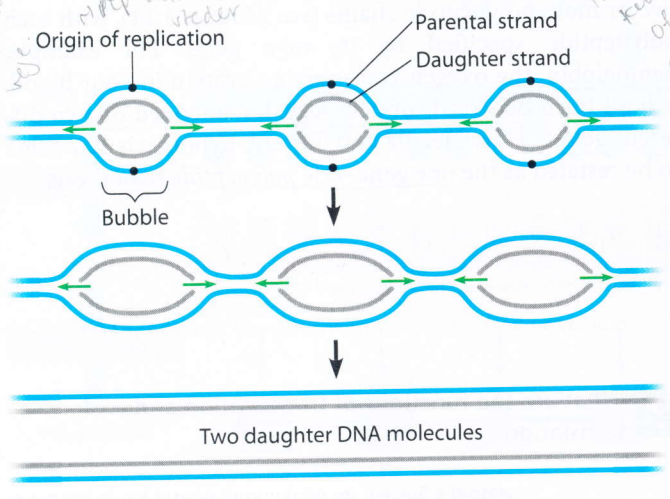

Figure 10.5A Multiple "bubbles" in replicating DNA

In addition to their roles in linking nucleotides together, DNA polymerases carry out a proofreading step that quickly removes nucleotides that have base-paired incorrectly during replication. DNA polymerases and DNA ligase are also involved in repairing DNA damaged by harmful radiation (such as ultraviolet light and X-rays) or toxic chemicals in the environment, such as those found in tobacco smoke.

DNA replication ensures that all the somatic cells in a multicellular organism carry the same genetic information. It is also the means by which genetic instructions are copied for the next generation of the organism. In the next module, we begin to pursue the connection between DNA instructions and an organism's phenotypic traits.

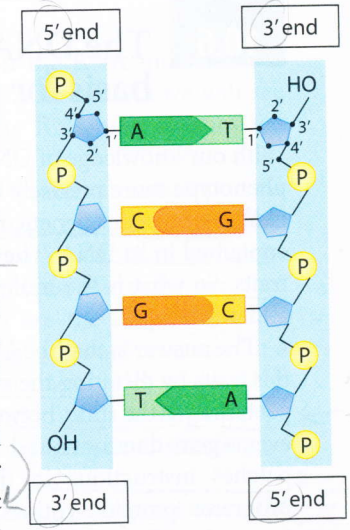

Figure 10.5B The opposite orientations of DNA strands

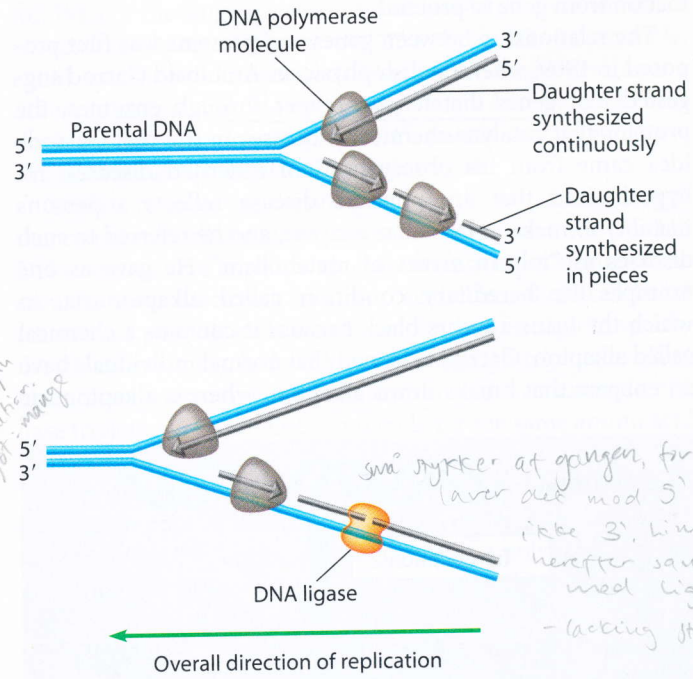

Figure 10.5C How daughter DNA strands are synthesized

Web Activity *DNA Replication*

? What is the function of DNA polymerase in DNA replication?

■ This enzyme lines up new nucleotides along an existing strand according to the base-pairing rules and then covalently connects the nucleotides into the new strand.

10.6 The DNA genotype is expressed as proteins, which provide the molecular basis for phenotypic traits

With our knowledge of DNA, we can now define genotype and phenotype more precisely than we did in Chapter 9. An organism's genotype, its genetic makeup, is the heritable information contained in its DNA. The phenotype is the organism's specific traits. So what is the molecular connection between genotype and phenotype?

The answer is that the DNA inherited by an organism specifies traits by dictating the synthesis of proteins. In other words, proteins are the links between genotype and phenotype. However, a gene does not build a protein directly. Rather, a gene dispatches instructions in the form of RNA, which in turn programs protein synthesis. This central concept in biology (termed the "central dogma" by Francis Crick in 1956) is summarized in Figure 10.6A. The chain of command is from DNA in the nucleus of the cell (purple area) to RNA to protein synthesis in the cytoplasm (tan area). The two main stages are **transcription,** the transfer of genetic information from DNA into an RNA molecule, and **translation,** the transfer of the information in the RNA into a protein. In the next nine modules, we will explore the steps in this flow of molecular information from gene to protein.

The relationship between genes and proteins was first proposed in 1909, when English physician Archibald Garrod suggested that genes dictate phenotypes through enzymes, the proteins that catalyze chemical processes in the cell. Garrod's idea came from his observations of inherited diseases. He hypothesized that an inherited disease reflects a person's inability to make a particular enzyme, and he referred to such diseases as "inborn errors of metabolism." He gave as one example the hereditary condition called alkaptonuria, in which the urine appears black because it contains a chemical called alkapton. Garrod reasoned that normal individuals have an enzyme that breaks down alkapton, whereas alkaptonuric individuals cannot make the enzyme. Garrod's hypothesis was ahead of its time, but research conducted decades later proved him right. In the intervening years, biochemists accumulated evidence that cells make and break down biologically important molecules via metabolic pathways, as in the synthesis of an amino acid or the breakdown of a sugar. As we described in Unit I, each step in a metabolic pathway is catalyzed by a specific enzyme. Therefore, individuals lacking one of the enzymes for a pathway are unable to complete it.

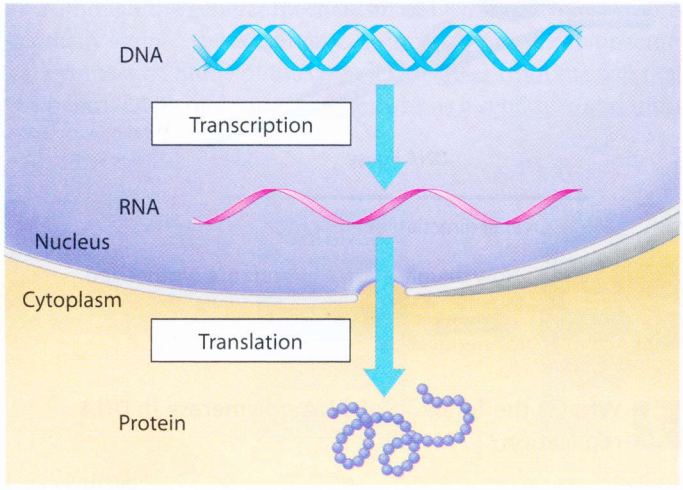

Figure 10.6A Flow of genetic information in a eukaryotic cell

The major breakthrough in demonstrating the relationship between genes and enzymes came in the 1940s from the work of American geneticists George Beadle and Edward Tatum with the bread mold *Neurospora crassa* (Figure 10.6B). Beadle and Tatum studied strains of the mold that were unable to grow on a simple growth medium. Each of these so-called nutritional mutants turned out to lack an enzyme in a metabolic pathway that produced some molecule the mold needed, such as an amino acid. Beadle and Tatum also showed that each mutant was defective in a single gene. This result suggested the one gene–one enzyme hypothesis: the function of a gene is to dictate the production of a specific enzyme.

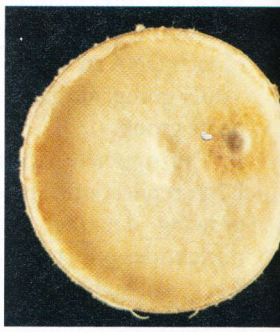

Figure 10.6B *Neurospora crassa* growing in a culture d

The one gene–one enzyme hypothesis has been amply confirmed, but with some important modifications. First it was extended beyond enzymes to include *all* types of proteins. For example, keratin—the structural protein of hair—and the hormone insulin are two examples of proteins that are not enzymes. So biologists began to think in terms of one gene–one *protein*. However, many proteins are made from two or more polypeptide chains (see Module 3.14), with each polypeptide specified by its own gene. For example, hemoglobin, the oxygen-transporting protein in your blood, is built from two kinds of polypeptides, encoded by two different genes. Thus, Beadle and Tatum's hypothesis has come to be restated as the one gene–one *polypeptide* hypothesis.

Bioflix *Protein Synthesis*

MP3 Tutor *DNA to RNA to Protein*

Web Process of Science *How Are Nutritional Mutations Identified?*

? What are the functions of transcription and translation?

■ Transcription is the transfer of information from DNA to RNA. Translation is the use of the RNA as information for making a protein.

10.7 Genetic information written in codons is translated into amino acid sequences

Genes provide the instructions for making specific proteins. But a gene does not build a protein directly. The bridge between DNA and protein synthesis is the nucleic acid RNA: DNA is transcribed into RNA, which is then translated into protein. Put another way, cells are governed by a molecular chain of command: DNA → RNA → protein.

Transcription and translation are linguistic terms, and it is useful to think of nucleic acids and proteins as having languages. To understand how genetic information passes from genotype to phenotype, we need to see how the chemical language of DNA is translated into the different chemical language of proteins.

What, exactly, is the language of nucleic acids? Both DNA and RNA are polymers made of nucleotide monomers. In DNA, there are four types of nucleotides, which differ in their nitrogenous bases (A, T, C, and G). The same is true for RNA, although it has the base U instead of T.

Figure 10.7 focuses on a small region of one of the genes (gene 3, shown in light blue) carried by a DNA molecule. DNA's language is written as a linear sequence of nucleotide bases on a polynucleotide, a sequence such as the one you see on the enlarged DNA strand in the figure. Specific sequences of bases, each with a beginning and an end, make up the genes on a DNA strand. A typical gene consists of hundreds or thousands of nucleotides in a specific sequence.

The pink strand underneath the enlarged DNA region represents the results of transcription: an RNA molecule. The process is called transcription because the nucleic acid language of DNA has been rewritten (transcribed) as a sequence of bases on RNA; the language is still that of nucleic acids.

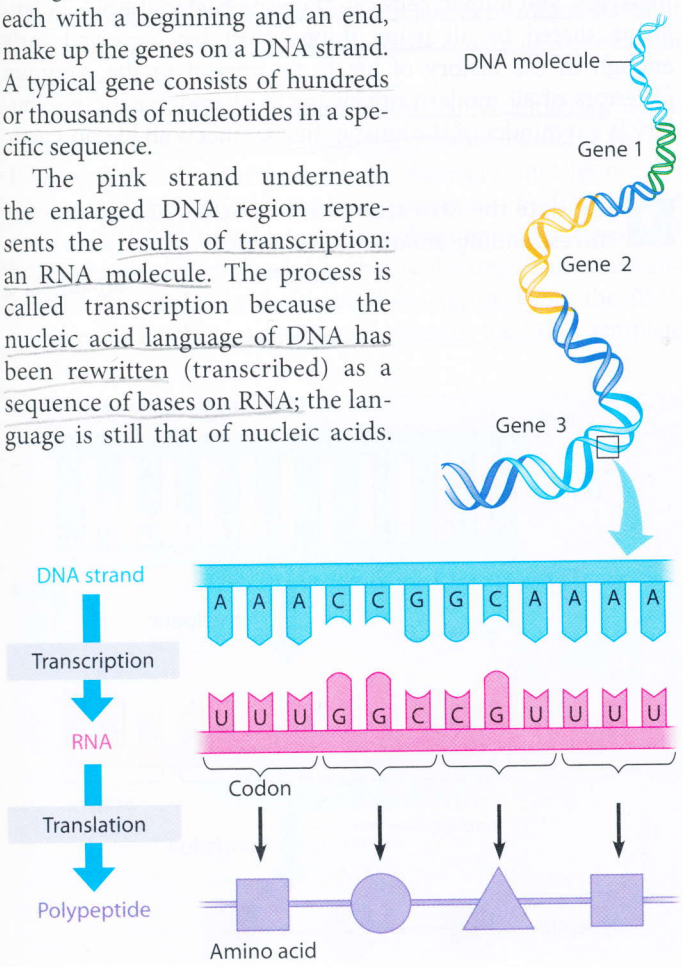

Figure 10.7 Transcription and translation of codons

Notice that the nucleotide bases on the RNA molecule are complementary to those on the DNA strand. As we will see in Module 10.9, this is because the RNA was synthesized using the DNA as a template.

The purple chain represents the results of translation, the conversion of the nucleic acid language into the polypeptide language (recall that proteins consist of one or more polypeptides). Like nucleic acids, polypeptides are polymers, but the monomers that compose them are the 20 amino acids common to all organisms. Again, the language is written in a linear sequence, and the sequence of nucleotides of the RNA molecule dictates the sequence of amino acids of the polypeptide. The RNA acts as a messenger carrying genetic information from DNA.

During translation, there is a change in language from the nucleotide sequence of the RNA into the amino acid sequence of the polypeptide. The brackets below the RNA indicate how genetic information is coded in nucleic acids. Notice that each bracket encloses *three* nucleotides on RNA. Recall that there are only four different kinds of nucleotides in DNA (A, G, C, T) and RNA (A, G, C, U). In translation, these four must somehow specify 20 amino acids. If each nucleotide base specified one amino acid, only 4 of the 20 amino acids could be accounted for. What if the language consisted of two-letter code words? If we read the bases of a gene two at a time, AG, for example, could specify one amino acid, whereas AT could designate a different amino acid. However, when the 4 bases are taken in doublets, there are only 16 (that is, 4^2) possible arrangements—still not enough to specify all 20 amino acids.

Triplets of bases are the smallest "words" of uniform length that can specify all the amino acids. Suppose each code word in DNA consists of a triplet, with each arrangement of three consecutive bases specifying an amino acid. Then there can be 64 (that is, 4^3) possible code words—more than enough to specify the 20 amino acids. Indeed, there are enough triplets to allow more than one coding for each amino acid. For example, the base triplets AAT and AAC both code for the same amino acid (leucine).

Experiments have verified that the flow of information from gene to protein is based on a **triplet code:** The genetic instructions for the amino acid sequence of a polypeptide chain are written in DNA and RNA as a series of three-base words, called **codons.** Notice in the figure that three-base codons in the DNA are transcribed into complementary three-base codons in the RNA, and then the RNA codons are translated into amino acids that form a polypeptide. We turn to the codons themselves in the next module.

? A particular protein is 100 amino acids long. How many nucleotides are necessary to code for this protein?

10.8 The genetic code is the Rosetta stone of life

In 1799, a large stone tablet was found in Rosetta, Egypt, carrying the same lengthy inscription in three ancient languages: Egyptian hieroglyphics, Egyptian script, and Greek. This stone provided the key that enabled scholars to crack the previously indecipherable hieroglyphic code.

To crack the genetic code, scientists wrote their own Rosetta stone. It was based on information gathered from a series of elegant experiments that disclosed the amino acid translations of each of the nucleotide-triplet code words. The first codon was deciphered in 1961 by American biochemist Marshall Nirenberg. He synthesized an artificial RNA molecule by linking together identical RNA nucleotides having uracil as their base. No matter where this message started or stopped, it could contain only one type of triplet codon: UUU. Nirenberg added this "poly U" to a test-tube mixture containing ribosomes and the other ingredients required for polypeptide synthesis. This mixture translated the poly U into a polypeptide containing a single kind of amino acid, phenylalanine. Thus, Nirenberg learned that the RNA codon UUU specifies the amino acid phenylalanine (Phe). By variations on this method, the amino acids specified by all the codons were soon determined.

The **genetic code** is the set of rules giving the correspondence between codons in RNA and amino acids in proteins. As Figure 10.8A shows, 61 of the 64 codons code for amino acids. The triplet AUG has a dual function: It codes for the amino acid methionine (Met) and also can provide a signal for the start of a polypeptide chain. Three of the other codons (in red boxes in the figure) do not designate amino acids. They are the stop codons that mark the end of translation.

Notice in Figure 10.8A that there is redundancy in the code but no ambiguity. For example, although codons UUU and UUC both specify phenylalanine (redundancy), neither of them ever represents any other amino acid (no ambiguity). The codons in the figure are the triplets found in RNA. They have a straightforward, complementary relationship to the codons in DNA. The nucleotides making up the codons occur in a linear order along the DNA and RNA, with no gaps or "punctuation" separating the codons.

As an exercise in translating the genetic code, consider the 12-nucleotide segment of DNA in Figure 10.8B. Let's read this as a series of triplets. Using the base-pairing rules (with U in RNA instead of T), we see that the RNA codon corresponding to the first transcribed DNA triplet, TAC, is AUG. As you can see in Figure 10.8A, AUG indicates, "place Met as the first amino acid in the polypeptide." The second DNA triplet, TTC, dictates RNA codon AAG, which designates lysine (Lys) as the second amino acid. We continue until we reach a stop codon.

The genetic code is nearly universal, shared by organisms from the simplest bacteria to the most complex plants and animals. In experiments, bacteria can translate human genetic messages, and human cells can translate bacterial RNA. A language shared by all living things must have evolved early enough in the history of life to be present in the common ancestors of all modern organisms. A shared genetic vocabulary is a reminder of the kinship that connects all life on Earth.

? Translate the RNA sequence CCAUUUACG into the corresponding amino acid sequence.

Pro-Phe-Thr ■

First base	Second base: U	Second base: C	Second base: A	Second base: G	Third base
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met or start	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

Figure 10.8A Dictionary of the genetic code (RNA codons)

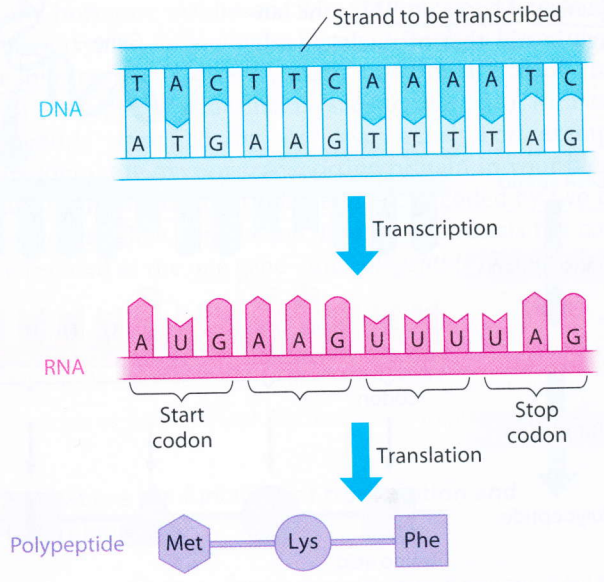

Figure 10.8B Deciphering the genetic information in DNA

10.9 Transcription produces genetic messages in the form of RNA

In eukaryotic cells, transcription, the transfer of genetic information from DNA to RNA, occurs in the nucleus. (The nucleus, after all, contains the DNA; see Figure 10.6A for a review.) An RNA molecule is transcribed from a DNA template by a process that resembles the synthesis of a DNA strand during DNA replication. Figure 10.9A is a close-up view of this process. As with replication, the two DNA strands must first separate at the place where the process will start. In transcription, however, only one of the DNA strands serves as a template for the newly forming molecule. The nucleotides that make up the new RNA molecule take their places one at a time along the DNA template strand by forming hydrogen bonds with the nucleotide bases there. Notice that the RNA nucleotides follow the same base-pairing rules that govern DNA replication, except that U, rather than T, pairs with A. The RNA nucleotides are linked by the transcription enzyme **RNA polymerase,** symbolized in the figure by the large gray shape in the background.

Figure 10.9B is an overview of the transcription of an entire prokaryotic gene. (We focus on prokaryotes here; eukaryotic transcription is a similar process but more complex.) Specific sequences of nucleotides along the DNA mark where transcription of a gene begins and ends. The "start transcribing" signal is a nucleotide sequence called a **promoter.** A promoter is a specific binding site for RNA polymerase and determines which of the two strands of the DNA double helix is used as the template in transcription.

❶ The first phase of transcription, called initiation, is the attachment of RNA polymerase to the promoter and the start of RNA synthesis. ❷ During a second phase of transcription, the RNA elongates. As RNA synthesis continues, the RNA strand peels away from its DNA template, allowing the two separated DNA strands to come back together in the region already transcribed. ❸ Finally, in the third phase, termination, the RNA polymerase reaches a sequence of bases in the DNA template called a **terminator.** This sequence signals the end of the gene; at that point, the polymerase molecule detaches from the RNA molecule and the gene.

In addition to producing RNA that encodes amino acid sequences, transcription makes two other kinds of RNA that are involved in building polypeptides. We discuss these three kinds of RNA in the next three modules.

MP3 Tutor *DNA Transcription*

Web Activity *Transcription*

? What is a promoter? What molecule binds to it?

■ A promoter is a specific nucleotide sequence at the start of a gene where RNA polymerase attaches and begins transcription.

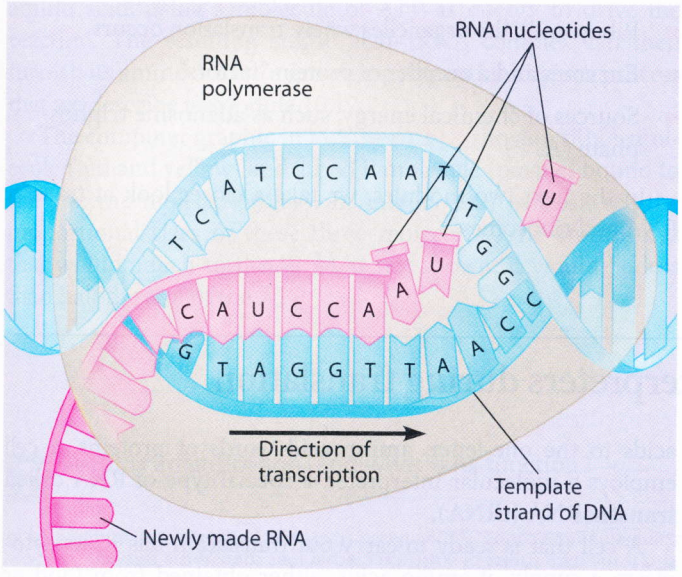

Figure 10.9A A close-up view of transcription

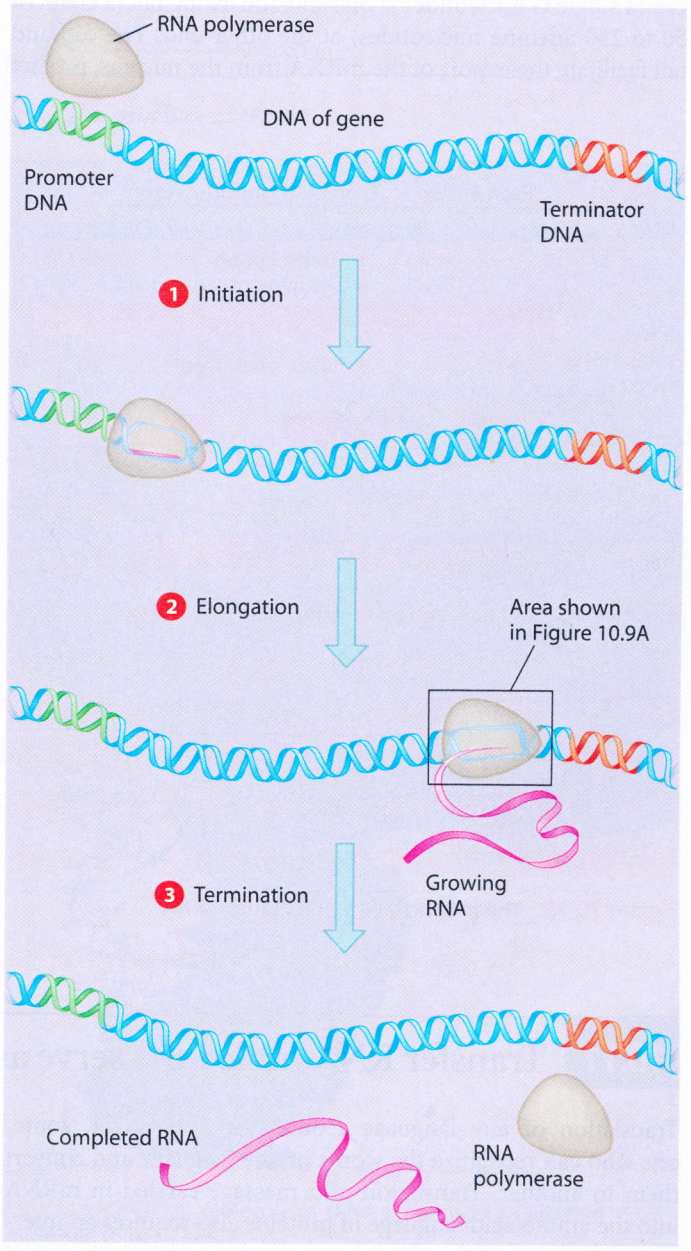

Figure 10.9B Transcription of a gene

10.10 Eukaryotic RNA is processed before leaving the nucleus

The kind of RNA that encodes amino acid sequences is called **messenger RNA (mRNA)** because it conveys genetic information from DNA to the translation machinery of the cell. Messenger RNA is transcribed from DNA, and the message in the mRNA is then translated into polypeptides. In prokaryotic cells, which lack a nucleus, transcription and translation occur in the same place (the cytoplasm). In eukaryotic cells, however, mRNA molecules and other RNA molecules required for translation must exit the nucleus via the nuclear pores and enter the cytoplasm, where the machinery for polypeptide synthesis is located.

Before leaving the nucleus as mRNA, eukaryotic transcripts are modified, or processed, in several ways. One kind of RNA processing is the addition of extra nucleotides to the ends of the RNA transcript (Figure 10.10). These additions include a small cap (a single G nucleotide) at one end and a long tail (a chain of 50 to 250 adenine nucleotides) at the other end. The cap and tail facilitate the export of the mRNA from the nucleus, protect the mRNA from attack by cellular enzymes, and help ribosomes bind to the mRNA. The cap and tail themselves are not translated into protein.

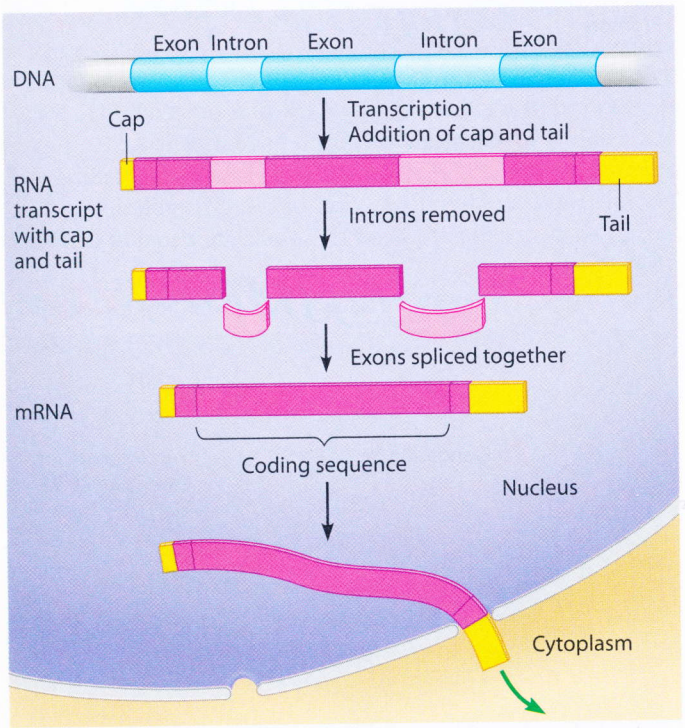

Figure 10.10 The production of eukaryotic mRNA

Eukaryotes require an additional type of RNA processing because, in most protein-coding genes, the DNA sequence that codes for the polypeptides is not continuous. Most genes of plants and animals include internal noncoding regions called **introns** (for "intervening sequences"). The coding regions—the parts of a gene that are expressed as amino acids—are called **exons.** As Figure 10.10 shows, both exons (darker color) and introns (lighter color) are transcribed from DNA into RNA. However, before the RNA leaves the nucleus, the introns are removed, and the exons are joined to produce an mRNA molecule with a continuous coding sequence. (The short noncoding regions just inside the cap and tail are considered parts of the first and last exons.) This cutting-and-pasting process is called **RNA splicing.** In most cases, RNA splicing is catalyzed by a complex of proteins and small RNA molecules, but sometimes the RNA transcript itself catalyzes the process. In other words, RNA can sometimes act as an enzyme that removes its own introns! As we will see in the next chapter (Module 11.6), RNA splicing also provides a means to produce multiple polypeptides from a single gene.

? Explain why many eukaryotic genes are longer than the mRNA that leaves the nucleus.

■ These genes have introns, noncoding sequences of nucleotides that are spliced out of the final RNA transcripts.

We are now ready to see how the translation process works. Translation of mRNA into protein involves more complicated machinery than transcription, including:

- Transfer RNA, another kind of RNA molecule
- Ribosomes, the organelles where translation occurs
- Enzymes and a number of protein "factors"
- Sources of chemical energy, such as adenosine triphosphate (ATP)

In the next two modules, we take a closer look at transfer RNA and ribosomes.

10.11 Transfer RNA molecules serve as interpreters during translation

Translation of any language requires an interpreter, someone who can recognize the words of one language and convert them to another. Translation of a message carried in mRNA into the amino acid language of proteins also requires an interpreter. To convert the three-letter words (codons) of nucleic acids to the one-letter, amino acid words of proteins, a cell employs a molecular interpreter, a special type of RNA called **transfer RNA (tRNA).**

A cell that is ready to carry out translation has in its cytoplasm a supply of amino acids, either obtained from food or

made from other chemicals. The amino acids themselves cannot recognize the codons in the mRNA. The amino acid tryptophan, for example, is no more attracted by codons for tryptophan than by any other codons. It is up to the cell's molecular interpreters, tRNA molecules, to match amino acids to the appropriate codons to form the new polypeptide. To perform this task, tRNA molecules must carry out two functions: (1) picking up the appropriate amino acids and (2) recognizing the appropriate codons in the mRNA. The unique structure of tRNA molecules enables them to perform both tasks.

As shown in Figure 10.11A, a tRNA molecule is made of a single strand of RNA—one polynucleotide chain—consisting of about 80 nucleotides. By twisting and folding upon itself, tRNA forms several double-stranded regions in which short stretches of RNA base-pair with other stretches. A single-stranded loop at one end of the folded molecule contains a special triplet of bases called an **anticodon.** The anticodon triplet is complementary to a codon triplet on mRNA. During translation, the anticodon on tRNA recognizes a particular codon on mRNA by using base-pairing rules. At the other end of the tRNA molecule is a site where an amino acid can attach.

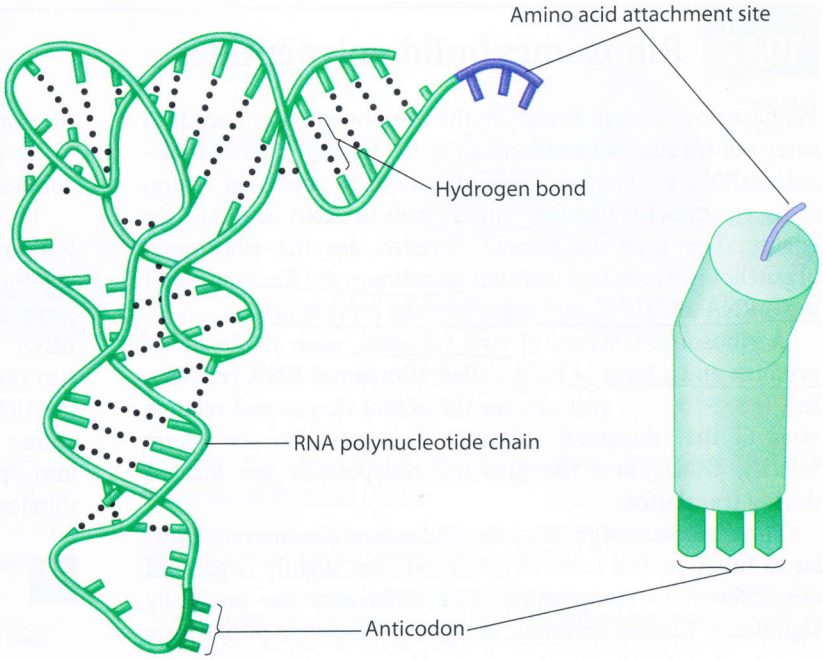

Figure 10.11A The structure of tRNA

In the modules that follow, in which we trace the process of translation, we represent tRNA with the simplified shape that is shown on the right in Figure 10.11A. This symbol emphasizes two parts of the molecule—the anticodon and the amino acid attachment site—that give tRNA its ability to match a particular nucleic acid word (codon) with its corresponding protein word (amino acid). Although all tRNA molecules are similar, there is a slightly different variety of tRNA for each amino acid.

Each amino acid is joined to the correct tRNA by a specific enzyme. There is a family of 20 versions of these enzymes, one enzyme for each amino acid. Each enzyme specifically binds one type of amino acid to all tRNA molecules that code for that amino acid, using a molecule of ATP as energy to drive the reaction. The resulting amino acid–tRNA complex can then furnish its amino acid to a growing polypeptide chain, a process that we describe in Module 10.12.

The computer graphic in Figure 10.11B shows a tRNA molecule (red and yellow) and an ATP molecule (purple) bound to the enzyme molecule (blue). In this picture, you can see the proportional sizes of these three molecules. The amino acid that would attach to the tRNA is not shown; it would be less than half the size of the ATP.

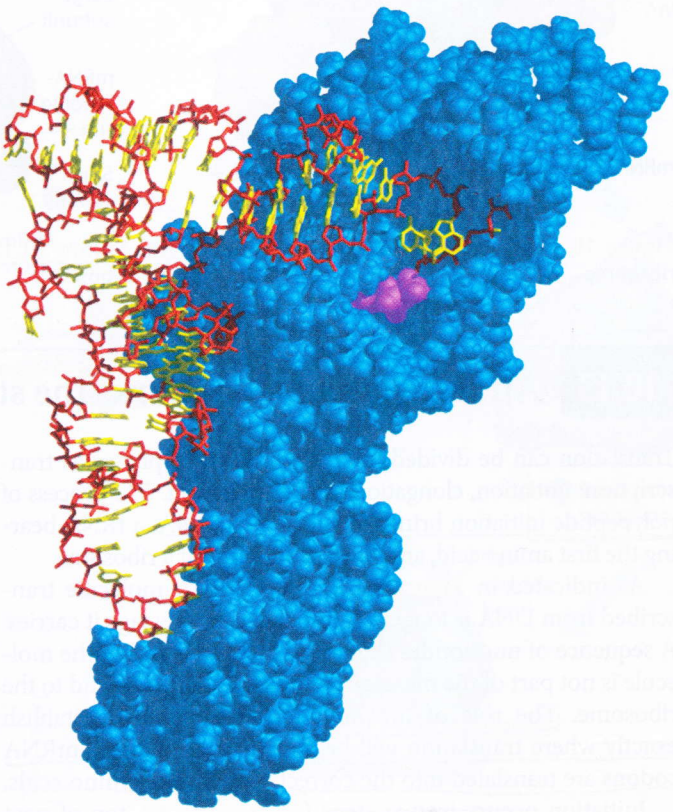

Figure 10.11B A molecule of tRNA binding to an enzyme molecule (blue)

? What is an anticodon, and what is its function?

■ It is the base triplet of a tRNA molecule that couples the tRNA to a complementary codon in the mRNA. This is a key step in translating mRNA to polypeptide.

10.12 Ribosomes build polypeptides

We have now looked at many of the components a cell needs to carry out translation: instructions in the form of mRNA molecules, tRNA to interpret the instructions, a supply of amino acids, enzymes for attaching amino acids to tRNA, and ATP for energy. The final components needed are the ribosomes, organelles in the cytoplasm that coordinate the functioning of the mRNA and tRNA and actually make polypeptides.

A ribosome consists of two subunits, each made up of proteins and a kind of RNA called **ribosomal RNA (rRNA).** In Figure 10.12A, you can see the actual shapes and relative sizes of the ribosomal subunits. You can also see where mRNA, tRNA, and the growing polypeptide are located during translation.

The ribosomes of prokaryotes and eukaryotes are very similar in function, but those of eukaryotes are slightly larger and are different in composition. The differences are medically significant. Certain antibiotic drugs can inactivate prokaryotic ribosomes while leaving eukaryotic ribosomes unaffected. These drugs, such as tetracycline and streptomycin, are used to combat bacterial infections.

The simplified drawings in Figures 10.12B and 10.12C indicate how tRNA anticodons and mRNA codons fit together on ribosomes. As Figure 10.12B shows, each ribosome has a binding site for mRNA and two binding sites for tRNA. Figure 10.12C shows tRNA molecules occupying these two sites. The subunits of the ribosome act like a vise, holding the tRNA and mRNA molecules close together, allowing the amino acids carried by the tRNA molecules to be connected into a polypeptide chain. In the next two modules, we examine the steps of translation in detail.

? How does a ribosome function in protein synthesis?

■ A ribosome holds mRNA and tRNAs together and connects amino acids from the tRNAs to the growing polypeptide chain.

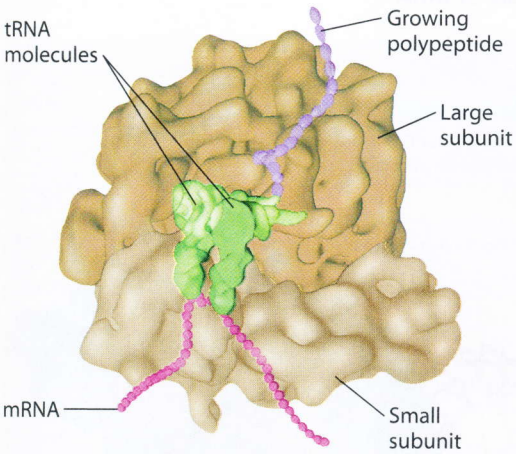

Figure 10.12A The true shape of a functioning ribosome

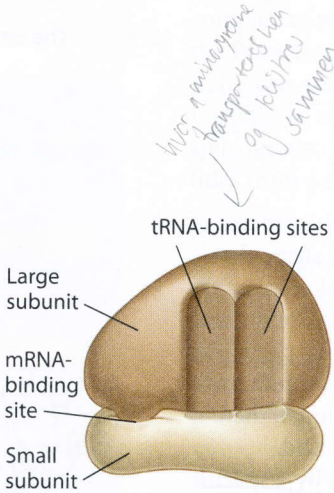

Figure 10.12B Binding sites of a ribosome

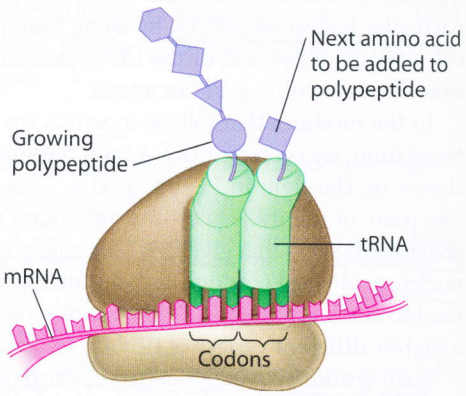

Figure 10.12C A ribosome with occupied binding sites

10.13 An initiation codon marks the start of an mRNA message

Translation can be divided into the same three phases as transcription: initiation, elongation, and termination. The process of polypeptide initiation brings together the mRNA, a tRNA bearing the first amino acid, and the two subunits of a ribosome.

As indicated in Figure 10.13A, an mRNA molecule transcribed from DNA is longer than the genetic message it carries. A sequence of nucleotides (light pink) at either end of the molecule is not part of the message but helps the mRNA bind to the ribosome. The role of the initiation process is to establish exactly where translation will begin, ensuring that the mRNA codons are translated into the correct sequence of amino acids.

Initiation occurs in two steps (Figure 10.13B, top of next page). ❶ An mRNA molecule binds to a small ribosomal subunit. A special initiator tRNA binds to the specific codon, called the **start codon,** where translation is to begin on the mRNA

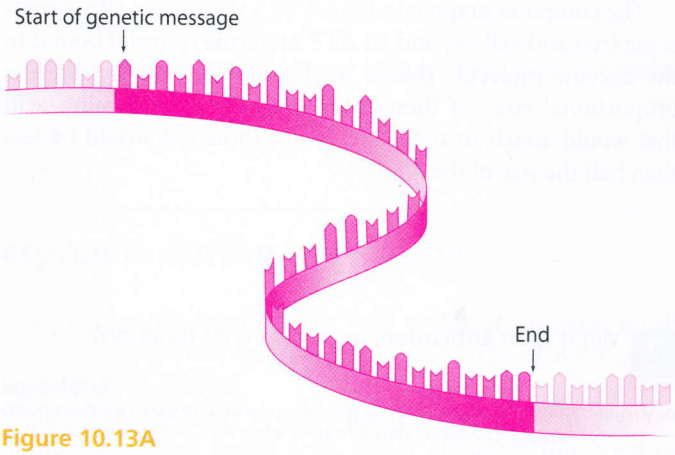

Figure 10.13A
A molecule of mRNA

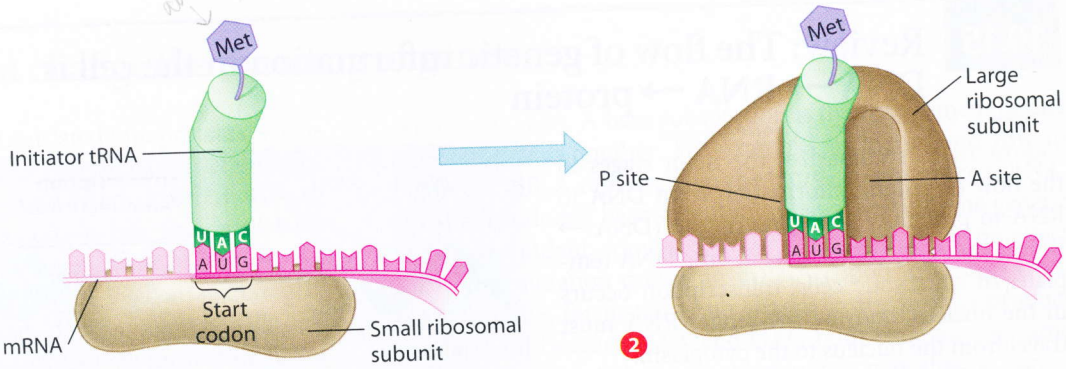

Figure 10.13B The initiation of translation

molecule. The initiator tRNA carries the amino acid methionine (Met); its anticodon, UAC, binds to the start codon, AUG. ❷ Next, a large ribosomal subunit binds to the small one, creating a functional ribosome. The initiator tRNA fits into one of the two tRNA-binding sites on the ribosome. This site, called the **P site,** will hold the growing polypeptide. The other tRNA-binding site, called the **A site,** is vacant and ready for the next amino-acid-bearing tRNA.

? What would happen if a genetic mutation changed a start codon to some other codon?

■ Any messenger RNA transcribed from the mutated gene would be nonfunctional because ribosomes could not initiate translation correctly.

10.14 Elongation adds amino acids to the polypeptide chain until a stop codon terminates translation

Once initiation is complete, amino acids are added one by one to the first amino acid. Each addition occurs in a three-step elongation process (Figure 10.14; the small green arrows indicate movement):

❶ **Codon recognition.** The anticodon of an incoming tRNA molecule, carrying its amino acid, pairs with the mRNA codon in the A site of the ribosome.

❷ **Peptide bond formation.** The polypeptide separates from the tRNA to which it was bound (the one in the P site) and attaches by a peptide bond to the amino acid carried by the tRNA in the A site. The ribosome catalyzes formation of the bond. Thus, one more amino acid is added to the chain.

❸ **Translocation.** The P site tRNA now leaves the ribosome, and the ribosome translocates (moves) the tRNA in the A site, with its attached polypeptide, to the P site. The codon and anticodon remain bonded, and the mRNA and tRNA move as a unit. This movement brings into the A site the next mRNA codon to be translated, and the process can start again with step 1.

Elongation continues until a **stop codon** reaches the ribosome's A site. Stop codons—UAA, UAG, and UGA—do not code for amino acids but instead act as signals to stop translation. This is the termination stage of translation. The completed polypeptide is released from the last tRNA and exits the ribosome, which then splits into its separate subunits.

Web Activity *Translation*

? What happens as a tRNA passes through the A and P binding sites on the ribosome?

■ In the A site, its amino acid receives the growing polypeptide from the tRNA that precedes it. In the P site, it gives up the polypeptide to the tRNA that follows it.

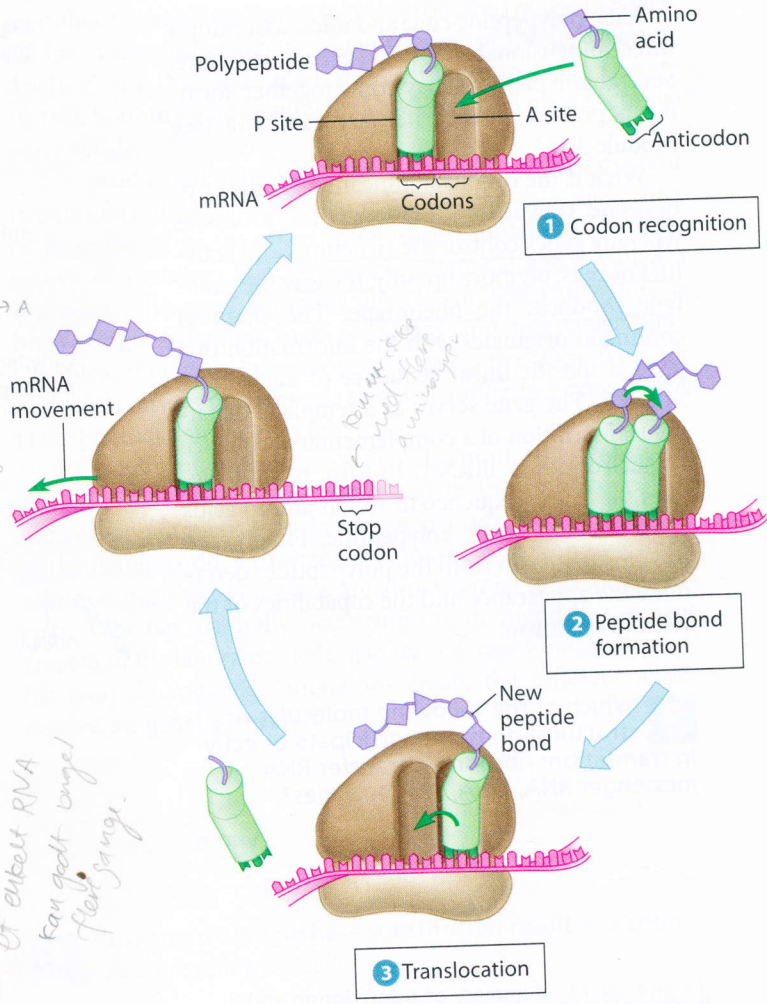

Figure 10.14 Polypeptide elongation

10.15 Review: The flow of genetic information in the cell is DNA → RNA → protein

Figure 10.15 summarizes the main stages in the flow of genetic information from DNA to RNA to protein. ❶ In transcription (DNA → RNA), the RNA is synthesized on a DNA template. In eukaryotic cells, transcription occurs in the nucleus, and the messenger RNA must travel from the nucleus to the cytoplasm.

❷–❺ Translation (RNA → protein) can be divided into four steps, all of which occur in the cytoplasm. When the polypeptide is complete at the end of step 5, the two ribosomal subunits come apart, and the tRNA and mRNA are released (not shown in this figure). Translation is rapid; a single ribosome can make an average-sized polypeptide in less than a minute. Typically, an mRNA molecule is translated simultaneously by a number of ribosomes. Once the start codon emerges from the first ribosome, a second ribosome can attach to it; thus, several ribosomes may trail along on the same mRNA molecule.

Each polypeptide coils and folds, assuming a three-dimensional shape, its tertiary structure. Several polypeptides may come together, forming a protein with quaternary structure (see Module 3.14).

What is the overall significance of transcription and translation? These are the processes whereby genes control the structures and activities of cells, or, more broadly, the way the genotype produces the phenotype. The chain of command originates with the information in a gene, a specific linear sequence of nucleotides in DNA. The gene serves as a template, dictating transcription of a complementary sequence of nucleotides in mRNA. In turn, mRNA dictates the linear sequence in which amino acids appear in a specific polypeptide. Finally, the proteins that form from the polypeptides determine the appearance and the capabilities of the cell and organism.

? Which of the following molecules or structures does not participate directly in translation: ribosomes, transfer RNA, messenger RNA, DNA, ATP, enzymes?

DNA ■

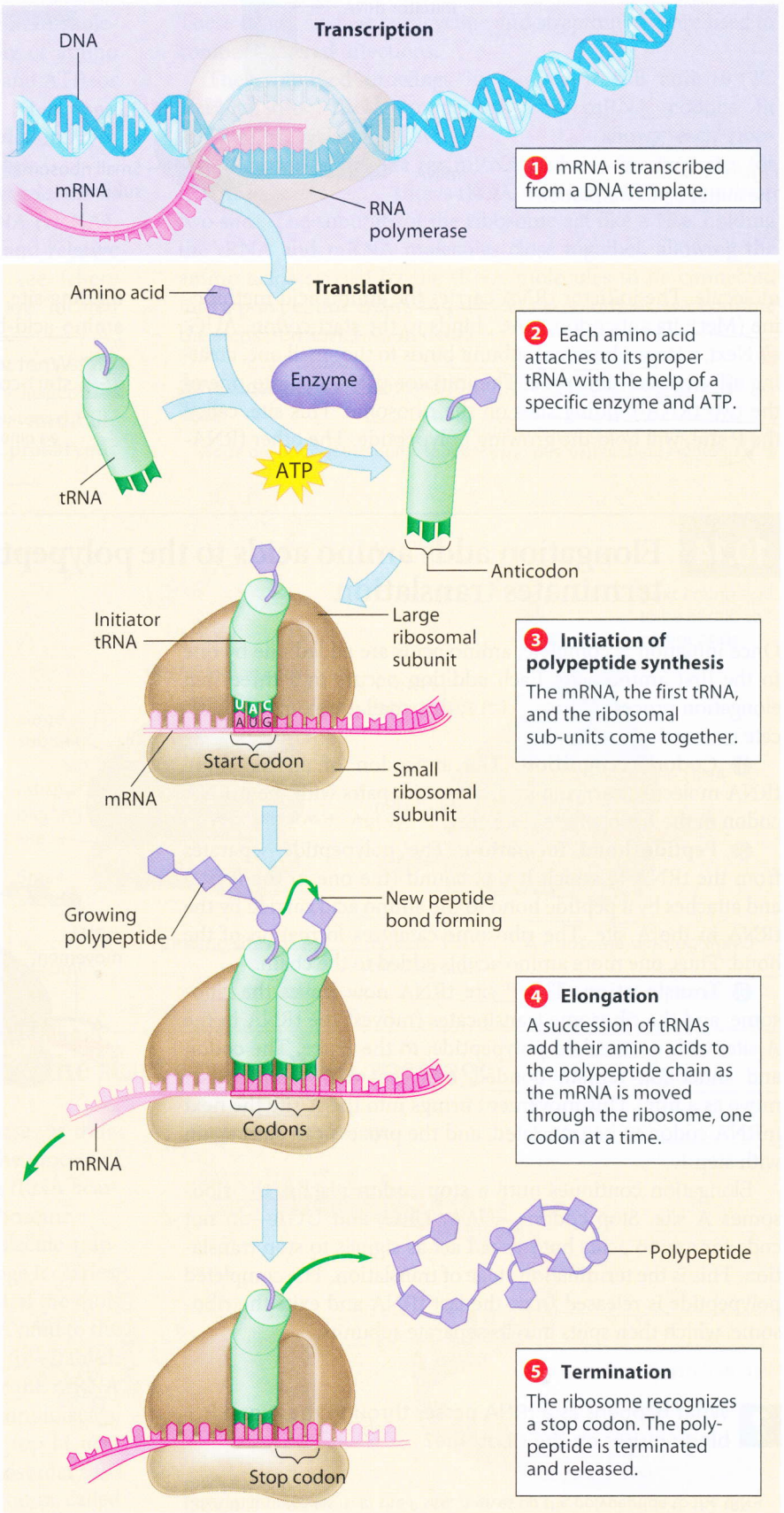

Figure 10.15 Summary of transcription and translation

10.16 Mutations can change the meaning of genes

Since discovering how genes are translated into proteins, scientists have been able to describe many inherited traits in molecular terms. For instance, when a child is born with sickle-cell disease (see Module 9.9), the condition can be traced back through a difference in a protein to one tiny change in a gene. In one of the two kinds of polypeptides in the hemoglobin protein, the sickle-cell child has a single different amino acid. This difference is caused by the change of a single nucleotide in the coding strand of DNA (Figure 10.16A). In the double helix, one base *pair* is changed.

We now know that the alternative alleles of many genes result from changes in single base pairs in DNA. Any change in the nucleotide sequence of DNA is called a **mutation.** Mutations can involve large regions of a chromosome or just a single nucleotide pair, as in sickle-cell disease. Here we consider how mutations involving only one or a few nucleotide pairs can affect gene translation.

Mutations within a gene can be divided into two general categories: base substitutions and base insertions or deletions (Figure 10.16B). A base substitution is the replacement of one nucleotide with another. For example, in the second row in Figure 10.16B, A replaces G in the fourth codon of the mRNA. What effect does a base substitution have? Because the genetic code is redundant, some substitution mutations have no effect at all. If a mutation causes an mRNA codon to change from GAA to GAG, for instance, no change in the protein product would result because GAA and GAG both code for the same amino acid (Glu). Other base substitutions may alter an amino acid but have little effect (perhaps because that amino acid is uninvolved in the protein's function). But as in the example in Figure 10.16A, base substitutions may cause changes in a protein that prevent it from functioning normally. Occasionally, a base substitution leads to an improved protein that enhances the success of the mutant organism and its descendants. Much more often, though, mutations are harmful. And if a base substitution changes an amino-acid codon to a stop codon, a shortened, probably nonfunctional, polypeptide will result.

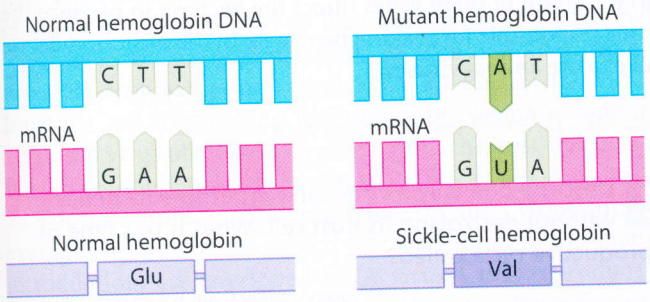

Figure 10.16A The molecular basis of sickle-cell disease

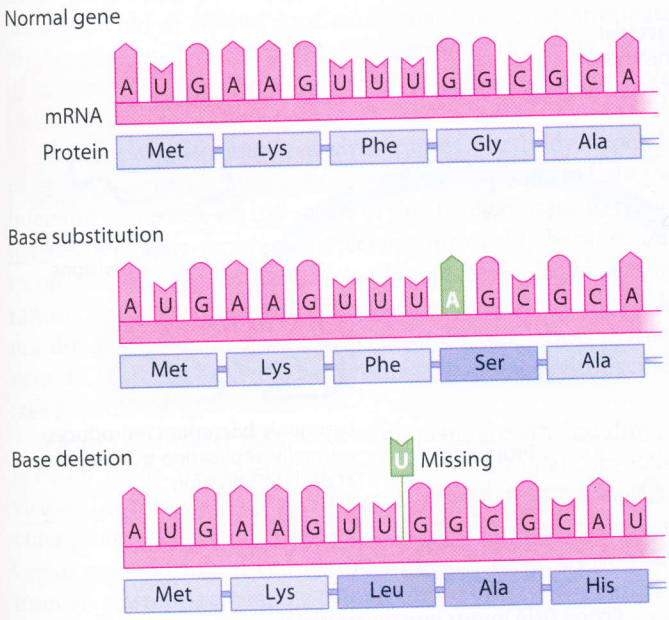

Figure 10.16B Types of mutations and their effects

Mutations involving the insertion or deletion of one or more nucleotides in a gene often have disastrous effects. Because mRNA is read as a series of nucleotide triplets (codons) during translation, adding or subtracting nucleotides may alter the **reading frame** (triplet grouping) of the message. All the nucleotides that are "downstream" of the insertion or deletion will be regrouped into different codons (Figure 10.16B, bottom). The result will most likely be a nonfunctional polypeptide.

The production of mutations, called **mutagenesis,** can occur in a number of ways. Errors that occur during DNA replication or recombination are called spontaneous mutations. Another source of mutation is a physical or chemical agent, called a **mutagen.** The most common physical mutagen in nature is high-energy radiation, such as X-rays and ultraviolet light. Chemical mutagens fall into several categories. One type, for example, consists of chemicals that are similar to normal DNA bases but that pair incorrectly.

Although mutations are often harmful, they are also extremely useful, both in nature and in the laboratory. It is because of mutations that there is such a rich diversity of genes in the living world, a diversity that makes evolution by natural selection possible. Mutations are also essential tools for geneticists. Whether naturally occurring (as in Mendel's peas) or created in the laboratory (Morgan used X-rays to make most of his fruit fly mutants), mutations create the different alleles needed for genetic research.

Web Process of Science *Connection: How Do You Diagnose a Genetic Disorder?*

? How could a single base substitution result in a shortened protein product?

■ A substitution that changed an amino acid codon into a stop codon would produce a prematurely terminated polypeptide.