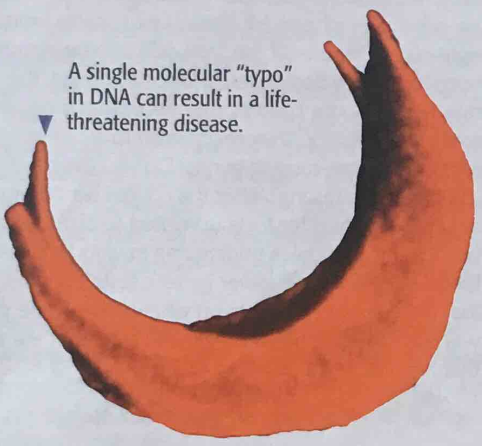
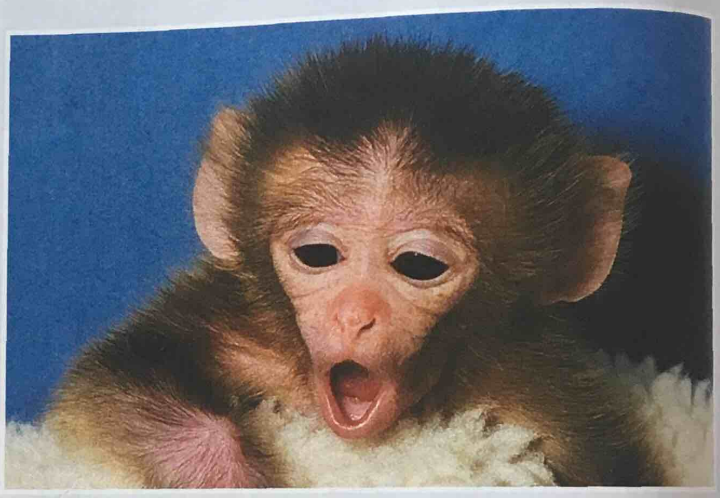
10 The Structure and Function of DNA

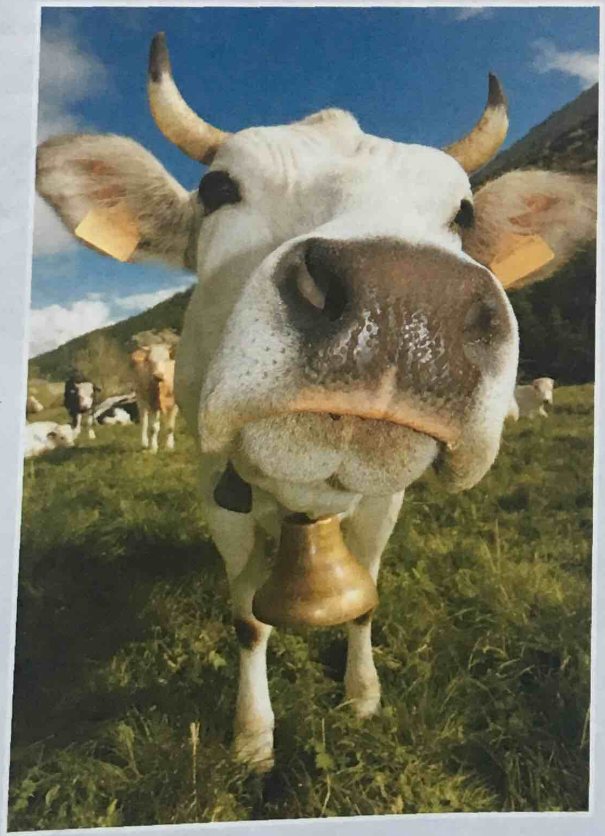
Why Molecular Biology Matters



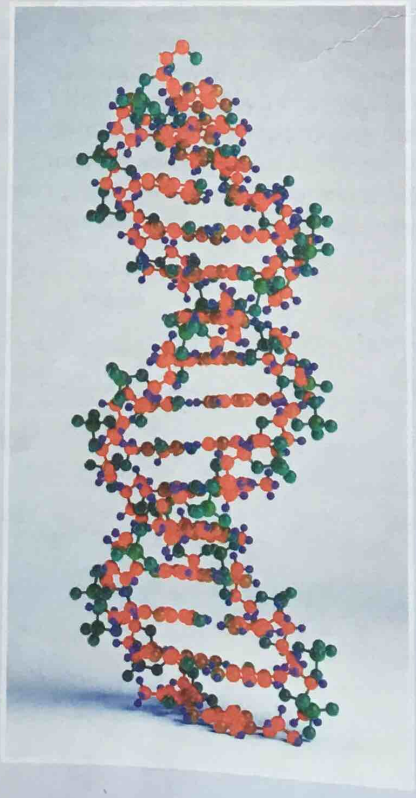
A single molecular "typo" in DNA can result in a life-threatening disease.



Because all life on Earth shares a universal genetic code, your DNA could be used to genetically modify a monkey.



Mad cow disease is caused by an abnormal molecule of protein.



Enzymes help maintain the integrity of your DNA to greater than 99.999% accuracy.

CHAPTER CONTENTS

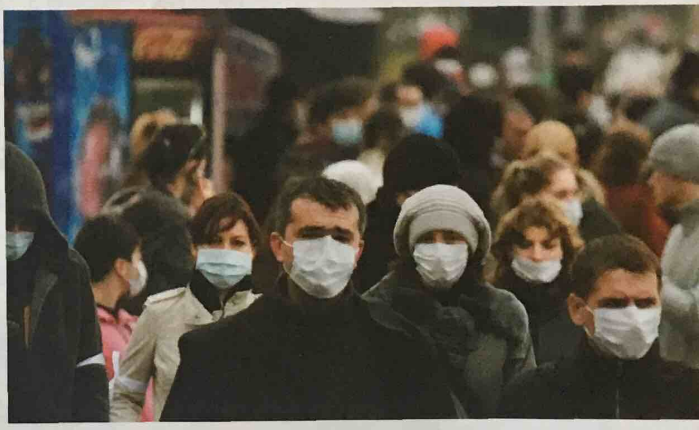
CHAPTER THREAD

The Deadliest Virus

The Deadliest Virus BIOLOGY AND SOCIETY

The First 21st-Century Pandemic

In 2009, a cluster of unusual flu cases was reported in and around Mexico City. Despite a near total shutdown of the city, the new flu virus, called 2009 H1N1, quickly spread to California and Texas. This strain was originally misnamed the "swine flu"; in fact, pigs had little impact on the spread of this virus, which, like most flu viruses, is passed from person to person through respiratory fluids. Nonetheless, the new strain received heavy media attention and caused considerable worry among the general public. In June 2009, the World Health Organization (WHO) declared H1N1 to be the first influenza pandemic (global epidemic) of the 21st century; the last flu pandemic was in 1968. In response, the WHO unveiled a comprehensive effort to contain H1N1, including increased surveillance, development of rapid testing procedures, the issuance of travel warnings, recommendations for increased sanitation (hand washing, etc.), and the production, stockpiling, and distribution of antiviral drugs. By 2010, cases of H1N1 had been confirmed in 214 countries.



The H1N1 influenza virus. Citizens of Russia try to protect themselves from H1N1 infection in late fall 2009.

Scientists soon determined that H1N1 was a hybrid flu strain, created when a previously known flu virus (which itself originated through a combination of viruses from birds, swine, and people) mixed with an Asian swine flu virus. This novel combination of genes produced some unusual features in the H1N1 strain. Most significantly, it infected young, healthy people, whereas the flu usually affects the elderly or people who are already sick. Many countries participated in the WHO-coordinated response, including widespread distribution of a new vaccine. As a result, the virus was contained, and WHO declared the pandemic over in August 2010. The WHO confirmed that this virus killed about 18,000 people, and estimates of the total number of unreported deaths topped 250,000.

You may wonder, What's the big deal about the flu? The flu is not merely a seasonal inconvenience. In fact, the influenza virus may be the deadliest pathogen known to science. In a typical year in the United States, more than 20,000 people die from influenza infection. And that is considered a good year. Once every few decades, a new flu strain explodes on the scene, causing pandemics and widespread death. H1N1 pales in comparison to the deadliest outbreak: the pandemic of 1918–1919, which in just 18 months killed 40 million people worldwide—that's more people than have died of AIDS since it was discovered more than 30 years ago. Given the nature of the threat, it's no surprise that health workers remain vigilant, always aware that a new and deadly flu virus may appear at any time.

The flu virus, like all viruses, consists of a relatively simple structure of nucleic acid (RNA in this case) and protein. Combating any virus requires a detailed understanding of life at the molecular level. In this chapter, we will explore the structure of DNA, how it replicates and mutates, and how it controls the cell by directing the synthesis of RNA and protein.

DNA: Structure and Replication

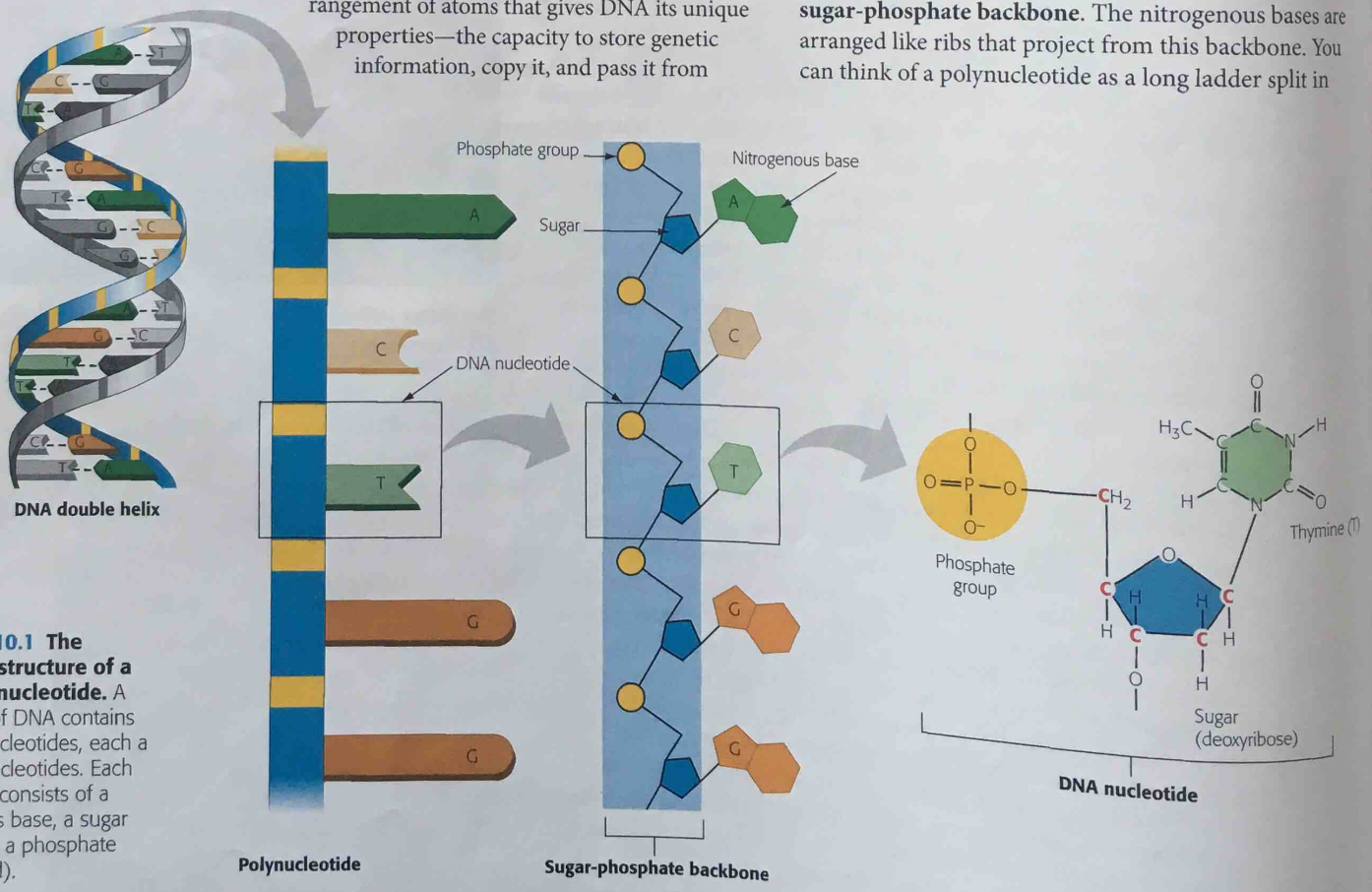
DNA was known to be a chemical component of cells by the late 1800s, but Gregor Mendel and other early geneticists did their work without any knowledge of DNA's role in heredity. By the late 1930s, experimental studies had convinced most biologists that one specific kind of molecule, rather than a complex chemical mixture, is the basis of inheritance. Attention focused on chromosomes, which were already known to carry genes. By the 1940s, scientists knew that chromosomes consist of two types of chemicals: DNA and protein. And by the early 1950s, a series of discoveries had convinced the scientific world that DNA was the molecule that acts as the hereditary material. This breakthrough ushered in the field of **molecular biology**, the study of heredity at the molecular level.

What came next was one of the most celebrated quests in the history of science: the effort to figure out the structure of DNA. A good deal was already known about DNA. Scientists had identified all its atoms and knew how they were bonded to one another. What was not understood was the specific three-dimensional arrangement of atoms that gives DNA its unique properties—the capacity to store genetic information, copy it, and pass it from generation to generation. The race was on to discover the link between the structure and function of this important molecule. We will describe that momentous discovery shortly. First, let's review the underlying chemical structure of DNA and its chemical cousin RNA.

DNA and RNA Structure

Both DNA and RNA are nucleic acids, which consist of long chains (polymers) of chemical units (monomers) called **nucleotides**. (For an in-depth refresher, see Figures 3.21–3.25.) A diagram of a nucleotide polymer, or **polynucleotide**, is shown in **Figure 10.1**. Polynucleotides can be very long and may have any sequence of the four different types of nucleotides (abbreviated A, C, T, and G), so a tremendous variety of polynucleotide chains is possible.

Nucleotides are joined together by covalent bonds between the sugar of one nucleotide and the phosphate of the next. This results in a repeating pattern of sugar-phosphate-sugar-phosphate, which is known as a **sugar-phosphate backbone**. The nitrogenous bases are arranged like ribs that project from this backbone. You can think of a polynucleotide as a long ladder split in



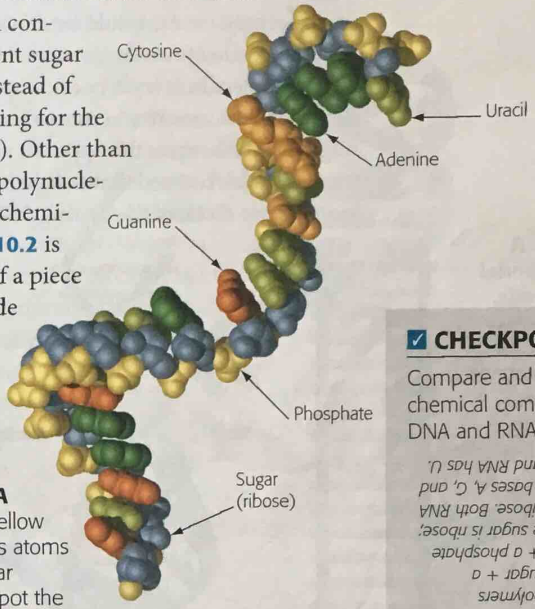
► **Figure 10.1 The chemical structure of a DNA polynucleotide.** A molecule of DNA contains two polynucleotides, each a chain of nucleotides. Each nucleotide consists of a nitrogenous base, a sugar (blue), and a phosphate group (gold).

half longwise, with rungs that come in four colors. The sugars and phosphates make up the side of the ladder, with the sugars acting as the half-rungs.

Moving from left to right across Figure 10.1, we can zoom in to see that each nucleotide consists of three components: a nitrogenous base, a sugar (blue), and a phosphate group (gold). Examining a single nucleotide even more closely, we see the chemical structure of its three components. The phosphate group, with a phosphorus atom (P) at its center, is the source of the *acid* in nucleic acid. Each phosphate has a negative charge on one of its oxygen atoms. The sugar has five carbon atoms (shown in red): 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. The full name for **DNA** is *deoxyribonucleic acid*, with *nucleic* referring to DNA's location in the nuclei of eukaryotic cells. The nitrogenous base (thymine, in our example) has a ring of nitrogen and carbon atoms with various chemical groups attached. Nitrogenous bases are basic (having a high pH, the opposite of acidic), hence their name.

The four nucleotides found in DNA differ only in their nitrogenous bases (see Figure 3.23 for a review).

The bases can be divided into two types. **Thymine (T)** and **cytosine (C)** are single-ring structures. **Adenine (A)** and **guanine (G)** are larger, double-ring structures. Instead of thymine, RNA has a similar base called **uracil (U)**. And RNA contains a slightly different sugar than DNA (ribose instead of deoxyribose, accounting for the names RNA vs. DNA). Other than that, RNA and DNA polynucleotides have the same chemical structure. **Figure 10.2** is a computer graphic of a piece of RNA polynucleotide about 20 nucleotides long. ☑



► **Figure 10.2 An RNA polynucleotide.** The yellow used for the phosphorus atoms and the blue of the sugar atoms make it easy to spot the sugar-phosphate backbone.

☑ CHECKPOINT

Compare and contrast the chemical components of DNA and RNA.

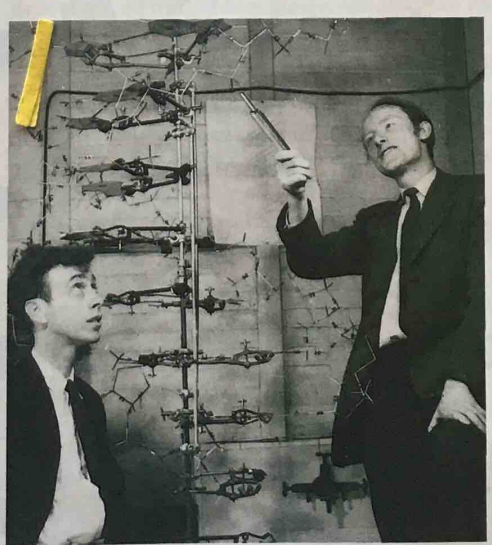
Answer: Both are polymers of nucleotides (a sugar + a nitrogenous base + a phosphate group). In RNA, the sugar is ribose; in DNA, it is deoxyribose. Both RNA and DNA have the bases A, C, and C, but DNA has T and RNA has U.

Watson and Crick's Discovery of the Double Helix

The celebrated partnership that solved the puzzle of DNA structure began soon after a 23-year-old newly minted American Ph.D. named James D. Watson journeyed to Cambridge University in England. There, a more senior scientist, Francis 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 image of DNA produced by Wilkins's colleague Rosalind Franklin. The data produced by Franklin turned out to be the key to the puzzle. A careful study of the image enabled Watson to figure out that the basic shape of DNA is a helix (spiral) with a uniform diameter. The thickness of the helix suggested that it was made up of two polynucleotide strands—in other words, a **double helix**. But how were the nucleotides arranged in the double helix?

Using wire models, Watson and Crick began trying to construct a double helix that conformed to all known data about DNA **(Figure 10.3)**. Watson placed the backbones on the outside of the model, forcing the nitrogenous bases to swivel to the interior of the molecule. As he did this, it occurred to him that the four kinds of bases must pair in a specific way. This idea of specific base pairing was a flash of inspiration that enabled Watson and Crick to solve the DNA puzzle.

▼ **Figure 10.3 Discoverers of the double helix.**



James Watson (left) and **Francis Crick.** The discoverers of the structure of DNA are shown in 1953 with their model of the double helix.

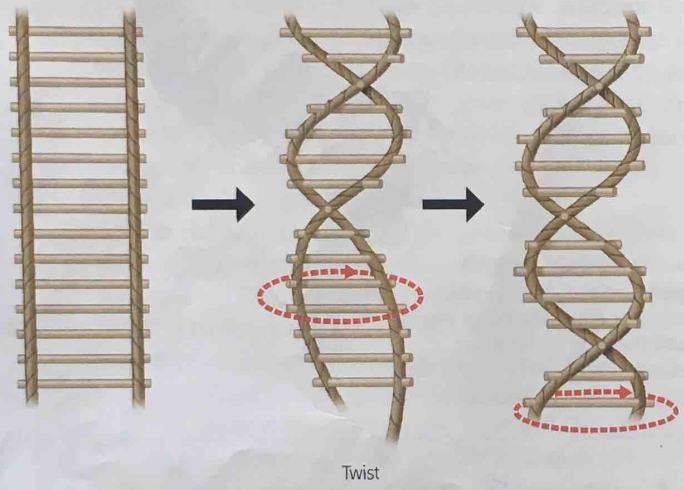
Rosalind Franklin Using X-rays, Franklin generated some of the key data that provided insight into the structure of DNA.

At first, Watson imagined that the bases paired like with like—for example, A with A, C with C. But that kind of pairing did not fit with the fact that the DNA molecule has a uniform diameter. An AA pair (made of two double-ringed bases) would be almost twice as wide as a CC pair (made of two single-ringed bases), resulting in an uneven molecule. It soon became apparent that a double-ringed base on one strand must always be paired with a single-ringed base on the opposite strand. Moreover, Watson and Crick realized that the chemical structure of each kind of base dictated the pairings even more specifically. Each base has protruding chemical groups that can best form hydrogen bonds with just one appropriate partner. This would be like having four colors of snap-together puzzle pieces and realizing that only certain colors can snap together (e.g., red can only snap together with blue). Similarly, 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.

If you imagine a polynucleotide strand as a half ladder, then you can picture the model of the DNA double helix proposed by Watson and Crick as a full ladder twisted into a spiral **(Figure 10.4)**. Figure 10.5 shows three more detailed representations of the double helix. The ribbonlike diagram in **Figure 10.5a** symbolizes the bases with shapes that emphasize their complementarity. **Figure 10.5b** is a more chemically precise version showing only four base pairs, with the helix untwisted and the individual hydrogen bonds specified by dashed lines. **Figure 10.5c** is a computer model showing part of a double helix in detail.

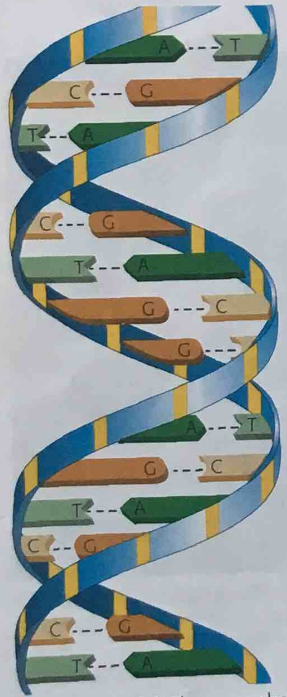
Although the base-pairing rules dictate the side-by-side combinations of bases that form the rungs of the double helix, they place no restrictions on the sequence of nucleotides along the length of a DNA strand. In fact, the sequence of bases can vary in countless ways.

In 1953, Watson and Crick rocked the scientific world with a succinct paper proposing their molecular model for

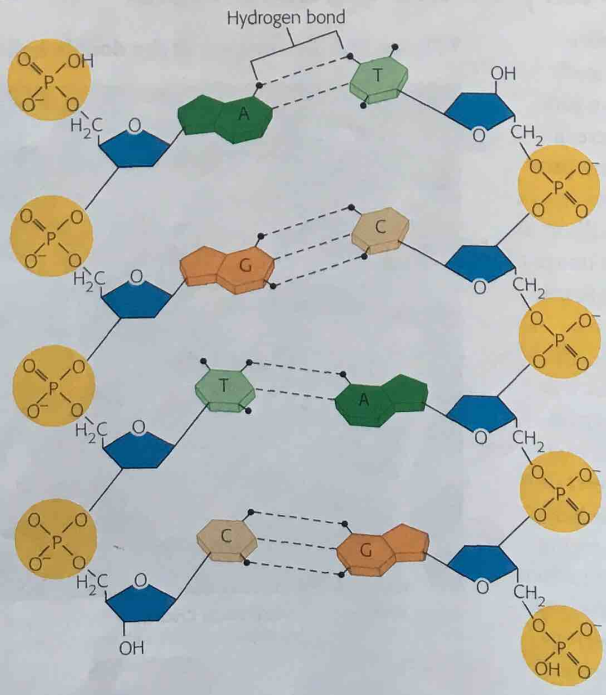


► **Figure 10.4 A rope-ladder model of a double helix.** The ropes at the sides represent the sugar-phosphate backbones. Each rung stands for a pair of bases connected by hydrogen bonds.

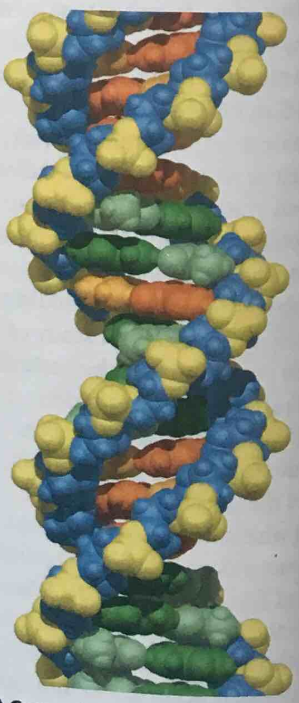
▼ **Figure 10.5 Three representations of DNA.**



(a) Ribbon model. The sugar-phosphate backbones are blue ribbons, and the bases are complementary shapes in shades of green and orange.



(b) Atomic model. In this more chemically detailed structure, you can see the individual hydrogen bonds (dashed lines). You can also see that the strands run in opposite directions: Notice that the sugars on the two strands are upside down with respect to each other.



(c) Computer model. Each atom is shown as a sphere, creating a space-filling model.

DNA. Few milestones in the history of biology have had as broad an impact as their double helix, with its A-T and C-G base pairing. In 1962, Watson, Crick, and Wilkins received the Nobel Prize for their work. (Franklin deserved a share of the prize, but she had died from cancer in 1958, and Nobel prizes are never granted posthumously.)

In their 1953 paper, Watson and Crick wrote that the structure they proposed "immediately suggests a possible copying mechanism for the genetic material." In other words, the structure of DNA points toward a molecular explanation for life's unique properties of reproduction and inheritance. Understanding how the arrangement of parts within DNA affects it actions within a cell is an excellent example of biology's important theme of the relationship of structure to function, as we see next.

Enzymes help maintain the integrity of your DNA to greater than 99.999% accuracy.

Structure/Function DNA Replication

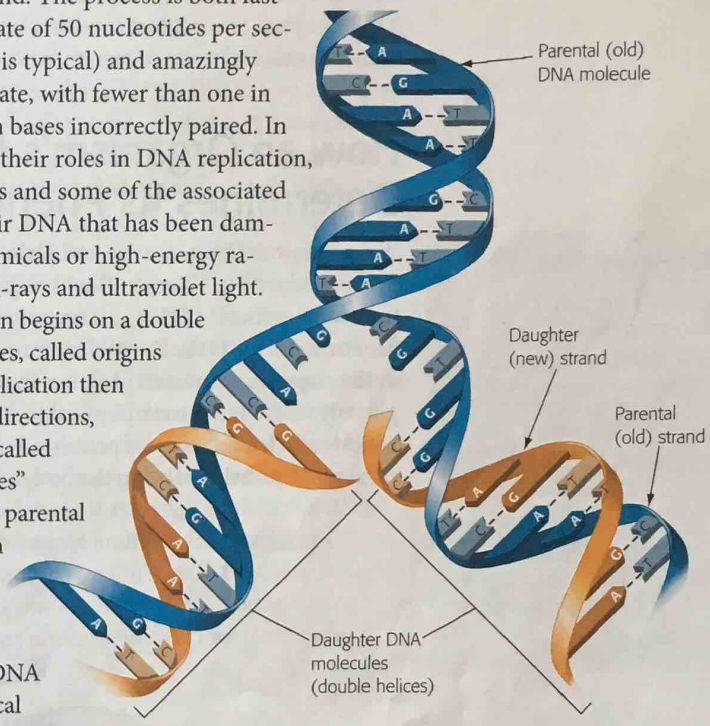
Every cell contains a DNA "cookbook" that provides complete information on how to make and maintain that cell. When a cell reproduces, it must duplicate this information, providing one copy to the new offspring cell while keeping one copy for itself. Thus, each cell must have a means of copying the DNA instructions. In a clear demonstration of how the structure of a biological system can provide insight into its function, Watson and Crick's model of DNA suggests that each DNA strand serves as a mold, or template, to guide reproduction of the other strand. If you know the sequence of bases in one strand of the double helix, you can very easily determine the sequence of bases in the other strand by applying the base-pairing rules: A pairs with T (and T with A), and G pairs with C (and C with G). For example, if one polynucleotide has the sequence AGTC, then the complementary polynucleotide in that DNA molecule must have the sequence TCAG.

Figure 10.6 shows how this model can account for the direct copying of a piece of DNA. The two strands of parental DNA separate, and each becomes a template for the assembly of a complementary strand from a supply of free nucleotides. The nucleotides are lined up one at a time along the template strand in accordance with the base-pairing rules. Enzymes link the nucleotides to form new DNA strands. The completed new molecules, identical to the parental molecule, are known as daughter DNA molecules (no gender should be inferred from this name).

The process of DNA replication requires the cooperation of more than a dozen enzymes and other proteins. The enzymes that make the covalent bonds between the nucleotides of a new DNA strand are called **DNA polymerases**. As an incoming nucleotide base-pairs with its complement on the template strand, a DNA polymerase adds it to the end of the growing daughter strand. The process is both fast (a rate of 50 nucleotides per second is typical) and amazingly accurate, with fewer than one in a billion bases incorrectly paired. In addition to their roles in DNA replication, DNA polymerases and some of the associated proteins can repair DNA that has been damaged by toxic chemicals or high-energy radiation, such as X-rays and ultraviolet light.

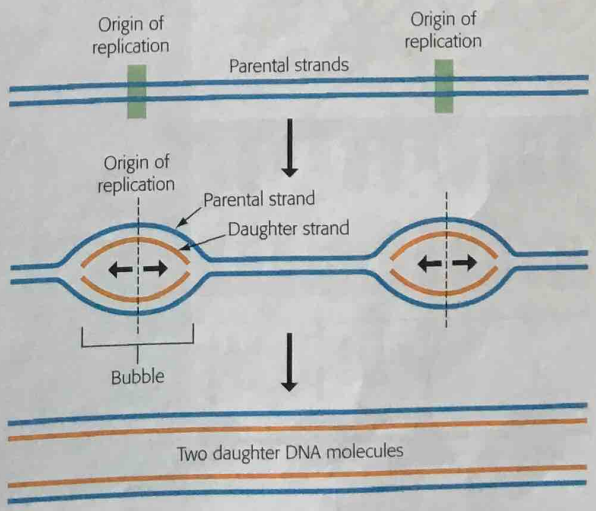
DNA replication begins on a double helix at specific sites, called origins of replication. Replication then proceeds in both directions, creating what are called replication "bubbles" **(Figure 10.7)**. The parental DNA strands open up as daughter strands elongate on both sides of each bubble. The DNA molecule of a typical eukaryotic chromosome has many origins where replication can start simultaneously, shortening the total time needed for the process. Eventually, all the bubbles merge, yielding two completed double-stranded daughter DNA molecules.

DNA replication ensures that all the body cells in a multicellular organism carry the same genetic information. It is also the means by which genetic information is passed along to offspring. ☑



▲ **Figure 10.6 DNA replication.** Replication results in two daughter DNA molecules, each consisting of one old strand and one new strand. The parental DNA untwists as its strands separate, and the daughter DNA rewinds as it forms.

▼ **Figure 10.7 Multiple "bubbles" in replicating DNA.**



☑ CHECKPOINT

1. How does complementary base pairing make DNA replication possible?
2. What enzymes connect nucleotides together during DNA replication?

Answers: **1.** When the two strands of the double helix separate, each serves as a template on which nucleotides can be arranged by specific base pairing into new complementary strands. **2.** DNA polymerases

Information Flow

From DNA to RNA to Protein

Now that we've seen how the structure of DNA allows it to be copied, let's explore how DNA provides instructions to a cell and to an organism as a whole.

How an Organism's Genotype Determines Its Phenotype

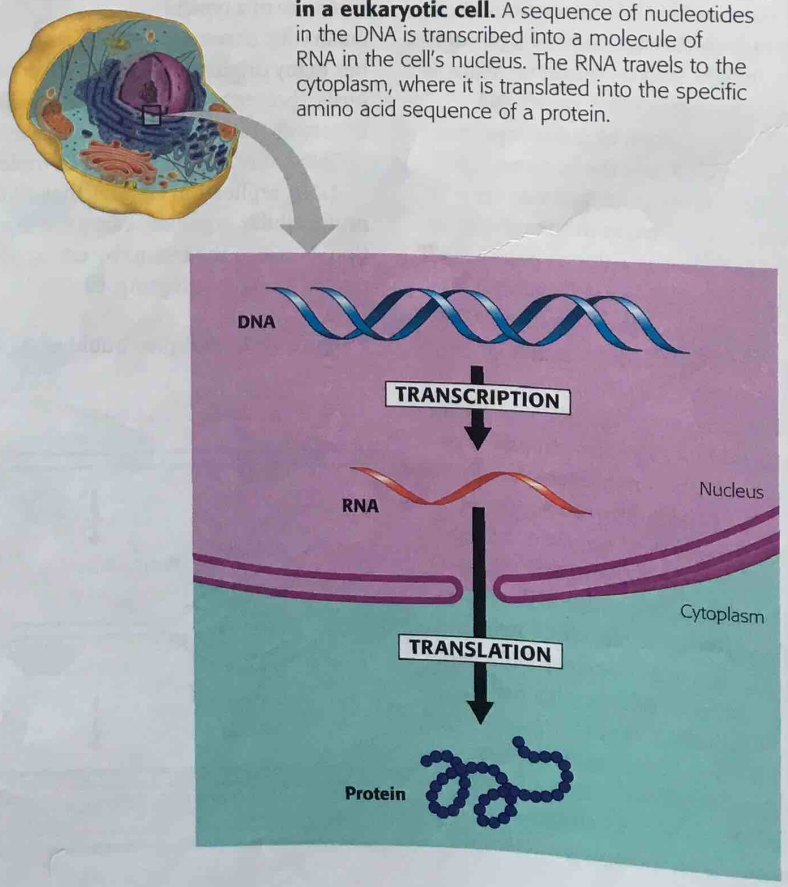
We can now define genotype and phenotype (terms first introduced in Chapter 9) with regard to the structure and function of DNA. An organism's *genotype*, its genetic makeup, is the heritable information contained in the sequence of nucleotide bases in its DNA. The *phenotype*, the organism's physical traits, arises from the actions of a wide variety of proteins. For example, structural proteins help make up the body of an organism, and enzymes catalyze the chemical reactions that are necessary for life.

DNA specifies the synthesis of proteins. However, a gene does not build a protein directly. Instead, DNA dispatches instructions in the form of RNA, which in turn programs protein synthesis. This fundamental concept in biology is summarized in **Figure 10.8**. The molecular "chain of command" is from DNA in the nucleus (purple area in the figure) to RNA to protein synthesis in the cytoplasm (blue area). The two stages are **transcription**, the transfer of genetic information from DNA into an RNA molecule, and **translation**, the transfer of the information from RNA into a polypeptide (protein strand). The relationship between genes and proteins is thus one of information flow: The function of a DNA gene is to dictate the production of a polypeptide. ☑

☑ CHECKPOINT

What are transcription and translation?

Answer: Transcription is the transfer of genetic information from DNA to RNA. Translation is the use of the information in an RNA molecule for the synthesis of a polypeptide.



▼ **Figure 10.8 The flow of genetic information in a eukaryotic cell.** A sequence of nucleotides in the DNA is transcribed into a molecule of RNA in the cell's nucleus. The RNA travels to the cytoplasm, where it is translated into the specific amino acid sequence of a protein.

From Nucleotides to Amino Acids: An Overview

Genetic information in DNA is transcribed into RNA and then translated into polypeptides, which then fold into proteins. But how do these processes occur? 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 strung together in specific sequences that convey information, much as specific sequences of letters convey information in English. In DNA, the monomers are the 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.

The language of DNA is written as a linear sequence of nucleotide bases, such as the blue sequence you see on the enlarged DNA strand in **Figure 10.9**. Every gene is made up of a specific sequence of bases, with special sequences marking the beginning and the end. A typical gene is a few thousand nucleotides in length.

When a segment of DNA is transcribed, the result is an RNA molecule. The process is called transcription because the nucleic acid language of DNA has simply been rewritten (transcribed) as a sequence of bases of RNA; the language is still that of nucleic acids. The nucleotide bases of the RNA molecule are complementary to those on the DNA strand. As you will soon see, this is because the RNA was synthesized using the DNA as a template.

Translation is the conversion of the nucleic acid language to the polypeptide language. Like nucleic acids, polypeptides are straight polymers, but the monomers that make them up—the letters of the polypeptide alphabet—are the 20 amino acids common to all organisms (represented as purple shapes in Figure 10.9). The sequence of nucleotides of the RNA molecule dictates the sequence of amino acids of the polypeptide. But remember, RNA is only a messenger; the genetic information that dictates the amino acid sequence originates in DNA.

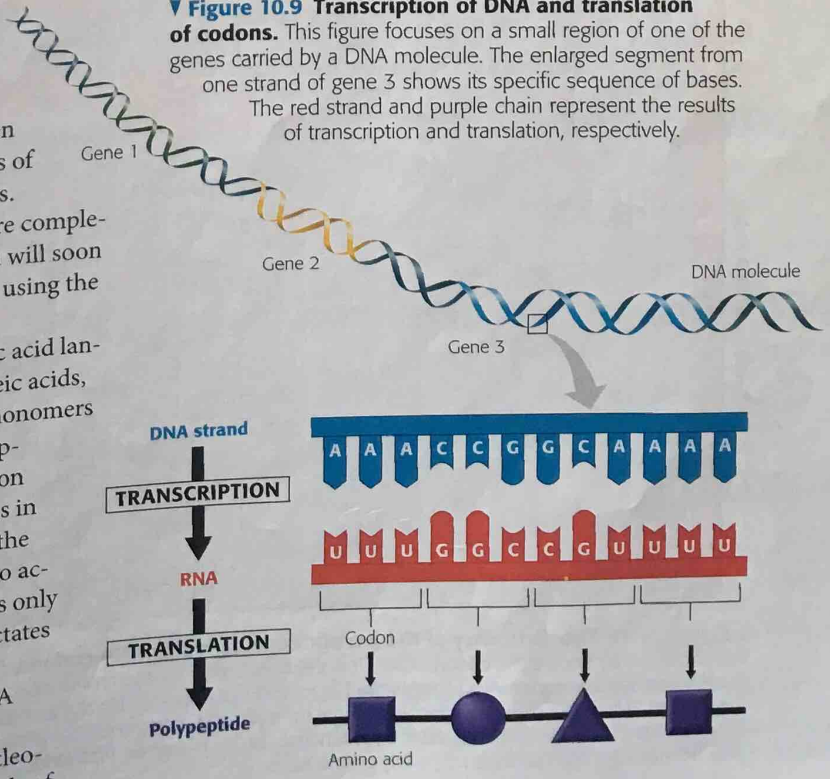
What are the rules for translating the RNA message into a polypeptide? In other words, what is the correspondence between the nucleotides of an RNA molecule and the amino acids of a polypeptide? Keep in mind that there are only four different kinds of nucleotides in DNA (A, G, C, T) and RNA (A, G, C, U). During translation, these four must somehow specify 20 amino acids. If each nucleotide base coded for one amino acid, only 4 of the 20 amino acids could be accounted for. In fact, triplets of bases are the smallest "words" of uniform length that can specify all the amino acids. There can be 64 (that is, 4^3) possible code words of this type—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 AAA and AAG both code for the same amino acid.

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**. 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. As summarized in Figure 10.9, one DNA codon (three nucleotides) → one RNA codon (three nucleotides) → one amino acid. Next we turn to the codons themselves. ☑

☑ CHECKPOINT

How many nucleotides are necessary to code for a polypeptide that is 100 amino acids long?

Answer: 300



▼ Figure 10.9 Transcription of DNA and translation of codons. This figure focuses on a small region of one of the genes carried by a DNA molecule. The enlarged segment from one strand of gene 3 shows its specific sequence of bases. The red strand and purple chain represent the results of transcription and translation, respectively.

The Genetic Code

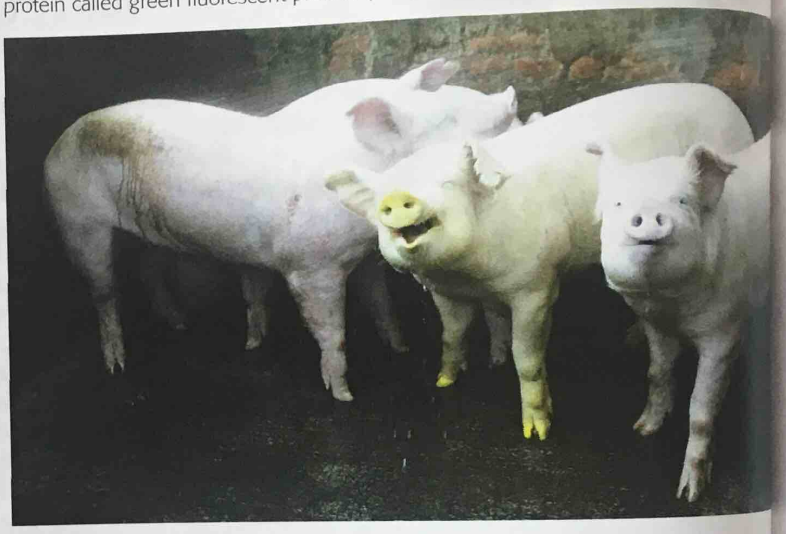
Because all life on Earth shares a universal genetic code, your DNA could be used to genetically modify a monkey.

The **genetic code** is the set of rules that convert a nucleotide sequence in RNA to an amino acid sequence. As **Figure 10.10** shows, 61 of the 64 triplets code for amino acids. The triplet AUG has a dual function: It codes for the amino acid methionine (abbreviated Met) and can also provide a signal for the start of a polypeptide chain. Three codons (UAA, UAG, and UGA) do not designate amino acids. They are the stop codons that instruct the ribosomes to end the polypeptide.

Notice in Figure 10.10 that a given RNA triplet always specifies a given amino acid. For example, although codons UUU and UUC both specify phenylalanine (Phe), neither of them ever represents any other amino acid. 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 separating the codons.

The genetic code is nearly universal, shared by organisms from the simplest bacteria to the most complex plants and animals. The universality of the genetic vocabulary suggests that it arose very early in evolution and was passed on over the eons to all the organisms living on Earth today. In fact, such universality is the key to modern DNA technologies. Because diverse organisms share a common genetic code, it is possible to program one species to produce a protein from another species by transplanting DNA (**Figure 10.11**). This allows scientists to mix and match genes from various species—a procedure with many useful genetic engineering applications in agriculture, medicine, and research (see Chapter 12 for further discussion of genetic engineering). Besides having practical purposes, a shared genetic vocabulary also reminds us of the evolutionary kinship that connects all life on Earth. ☑

▼ **Figure 10.11 A pig expressing a foreign gene.** The glowing porker in the middle was created when researchers incorporated a jelly (jellyfish) gene for a protein called green fluorescent protein (GFP) into the DNA of a standard pig.



Second base of RNA codon

First base of RNA codon	U	C	A	G	Third base of RNA codon
U	UUU, UUC Phenylalanine (Phe); UUA, UUG Leucine (Leu)	UCU, UCC, UCA, UCG Serine (Ser)	UAU, UAC Tyrosine (Tyr); UAA Stop; UAG Stop	UGU, UGC Cysteine (Cys); UGA Stop; UGG Tryptophan (Trp)	U C A G
C	CUU, CUC, CUA, CUG Leucine (Leu)	CCU, CCC, CCA, CCG Proline (Pro)	CAU, CAC Histidine (His); CAA, CAG Glutamine (Gln)	CGU, CGC, CGA, CGG Arginine (Arg)	U C A G
A	AUU, AUC, AUA Isoleucine (Ile); AUG Met or start	ACU, ACC, ACA, ACG Threonine (Thr)	AAU, AAC Asparagine (Asn); AAA, AAG Lysine (Lys)	AGU, AGC Serine (Ser); AGA, AGG Arginine (Arg)	U C A G
G	GUU, GUC, GUA, GUG Valine (Val)	GCU, GCC, GCA, GCG Alanine (Ala)	GAU, GAC Aspartic acid (Asp); GAA, GAG Glutamic acid (Glu)	GGU, GGC, GGA, GGG Glycine (Gly)	U C A G

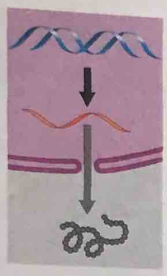
▲ **Figure 10.10 The dictionary of the genetic code, listed by RNA codons.** Practice using this dictionary by finding the codon UGG. (It is the only codon for the amino acid tryptophan, Trp.) Notice that the codon AUG (highlighted in green) not only stands for the amino acid methionine (Met), but also functions as a signal to "start" translating the RNA at that place. Three of the 64 codons (highlighted in red) function as "stop" signals that mark the end of a genetic message, but do not encode any amino acids.

☑ CHECKPOINT

An RNA molecule contains the nucleotide sequence CCAUUUACG. Using Figure 10.10, translate this sequence into the corresponding amino acid sequence.

Answer: Pro-Phe-Thr

Transcription: From DNA to RNA



Let's look more closely at transcription, the transfer of genetic information from DNA to RNA. If you think of your DNA as a cookbook, then transcription is the process of copying one recipe onto an index card (a molecule of RNA) for immediate use. **Figure 10.12a** is a close-up view of this process. As with DNA 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 RNA molecule; the other strand is unused. The nucleotides that make up the new RNA molecule take their place 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 usual base-pairing rules, except that U, rather than T, pairs with A. The RNA nucleotides are linked by the transcription enzyme **RNA polymerase**.

Figure 10.12b is an overview of the transcription of an entire gene. Special sequences of DNA nucleotides tell the RNA polymerase where to start and where to stop the transcribing process.

❶ Initiation of Transcription

The "start transcribing" signal is a nucleotide sequence called a **promoter**, which is located in the DNA at the beginning of the gene. A promoter is a specific place where RNA polymerase attaches. The first phase of transcription, called initiation, is the attachment of RNA polymerase to the promoter and the start of RNA synthesis. For any gene, the promoter dictates which of the two DNA strands is to be transcribed (the particular strand varies from gene to gene).

❷ RNA Elongation

During the second phase of transcription, elongation, the RNA grows longer. 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 Termination of Transcription

In the third phase, termination, the RNA polymerase reaches a special sequence of bases in the DNA template called a **terminator**. This sequence signals the end of the gene. At this point, the polymerase molecule detaches from the RNA molecule and the gene, and the DNA strands rejoin.

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 kinds of RNA a little later. ☑

☑ CHECKPOINT

How does RNA polymerase "know" where to start transcribing a gene?

Answer: It recognizes the gene's promoter, a specific nucleotide sequence.

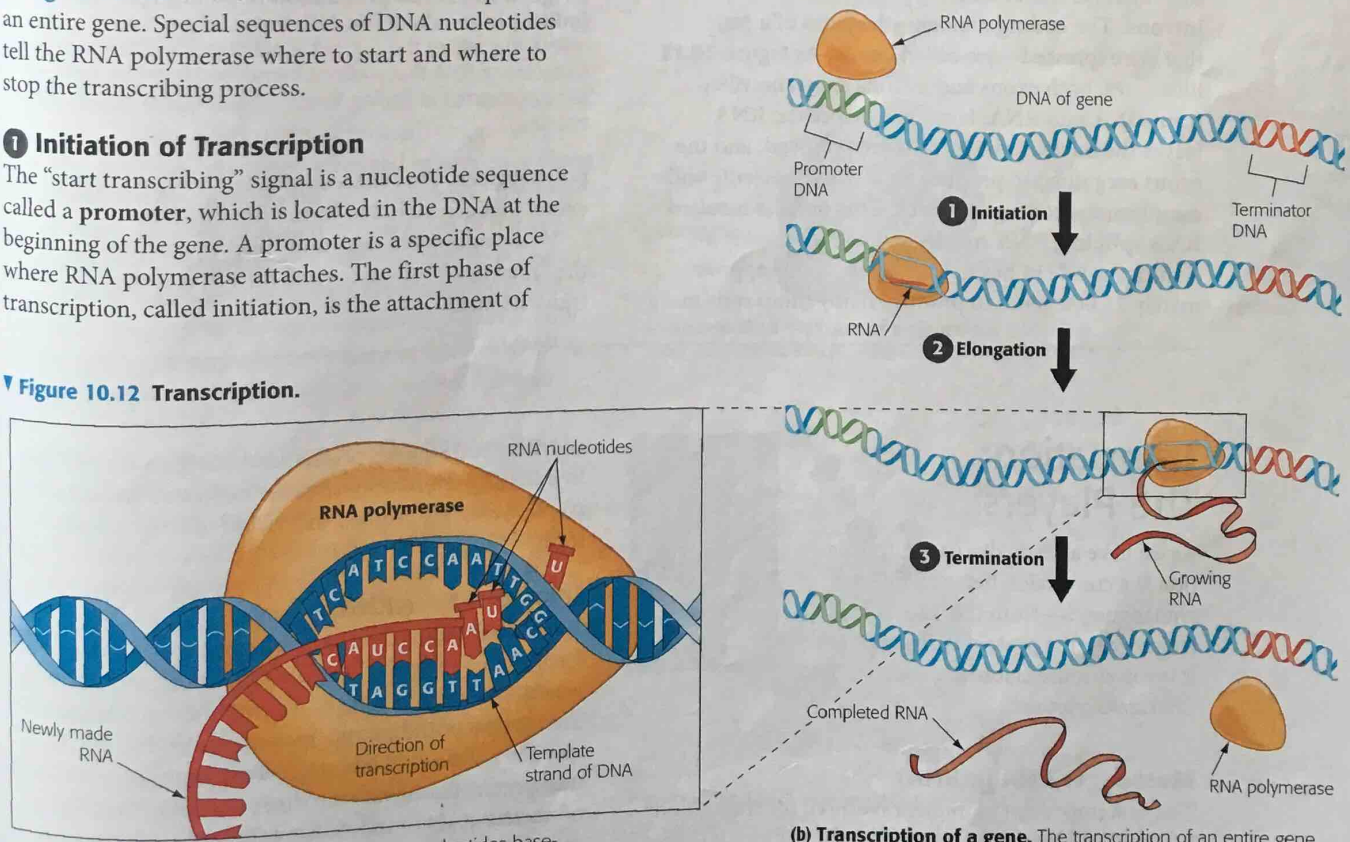


Figure 10.12 Transcription.

(a) A close-up view of transcription. As RNA nucleotides base-pair one by one with DNA bases on one DNA strand (called the template strand), the enzyme RNA polymerase (orange) links the RNA nucleotides into an RNA chain.

(b) Transcription of a gene. The transcription of an entire gene occurs in three phases: initiation, elongation, and termination of the RNA. The section of DNA where the RNA polymerase starts is called the promoter; the place where it stops is called the terminator.

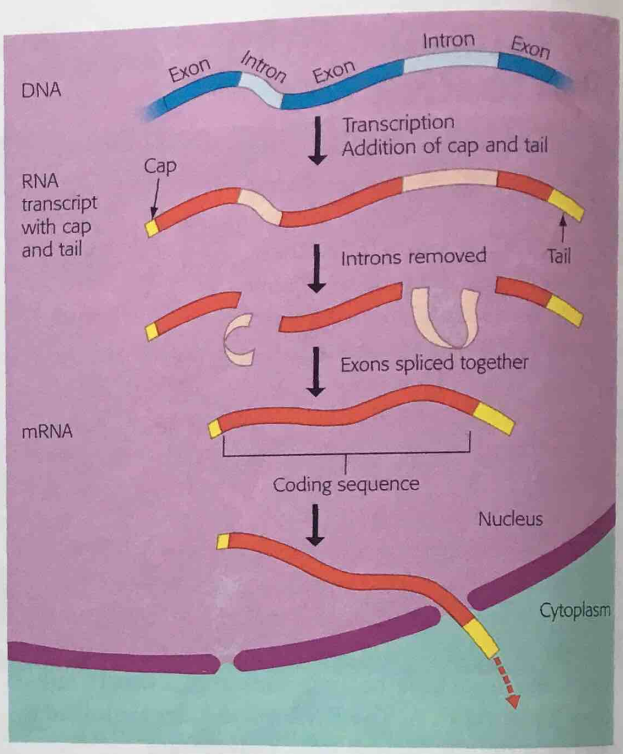
The Processing of Eukaryotic RNA

In the cells of prokaryotes, which lack nuclei, the RNA transcribed from a gene immediately functions as **messenger RNA (mRNA)**, the molecule that is translated into protein. But this is not the case in eukaryotic cells. The eukaryotic cell not only localizes transcription in the nucleus but also modifies, or processes, the RNA transcripts there before they move to the cytoplasm for translation by the ribosomes.

One kind of RNA processing is the addition of extra nucleotides to the ends of the RNA transcript. These additions, called the **cap** and **tail**, protect the RNA from attack by cellular enzymes and help ribosomes recognize the RNA as mRNA.

Another type of RNA processing is made necessary in eukaryotes by noncoding stretches of nucleotides that interrupt the nucleotides that actually code for amino acids. It is as if nonsense words were randomly interspersed within a recipe that you copied. Most genes of plants and animals, it turns out, include such internal noncoding regions, which are called **introns**. The coding regions—the parts of a gene that are expressed—are called **exons**. As **Figure 10.13** illustrates, both exons and introns 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. This process is called **RNA splicing**. RNA splicing is believed to play a significant role in humans in allowing our approximately 21,000 genes to produce many thousands more polypeptides. This is accomplished by varying the exons that are included in the final mRNA.

With capping, tailing, and splicing completed, the "final draft" of eukaryotic mRNA is ready for translation. ☑



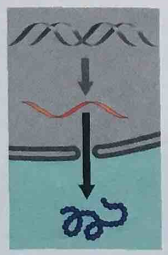
▲ **Figure 10.13 The production of messenger RNA (mRNA) in a eukaryotic cell.** Note that the molecule of mRNA that leaves the nucleus is substantially different from the molecule of RNA that was first transcribed from the gene. In the cytoplasm, the coding sequence of the final mRNA will be translated.

☑ CHECKPOINT

Why is a final mRNA often shorter than the DNA gene that coded for it?

Answer: because introns are removed from the RNA

Translation: The Players



As we have already discussed, translation is a conversion between different languages—from the nucleic acid language to the protein language—and it involves more elaborate machinery than transcription.

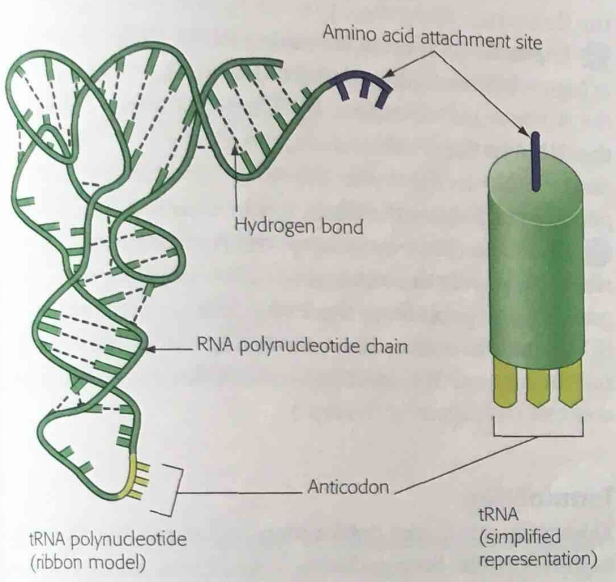
Messenger RNA (mRNA)

The first important ingredient required for translation is the mRNA produced by transcription. Once it is present, the machinery used to translate mRNA requires enzymes and sources of chemical energy, such as ATP. In addition, translation requires two other important components: ribosomes and a kind of RNA called transfer RNA.

Transfer RNA (tRNA)

Translation of any language into another requires an interpreter, someone or something that can recognize the words of one language and convert them to the other. Translation of the genetic 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 amino acid words of proteins, a cell uses a molecular interpreter, a type of RNA called **transfer RNA (tRNA)**, depicted in **Figure 10.14**.

▼ **Figure 10.14 The structure of tRNA.** At one end of the tRNA is the site where an amino acid will attach (purple), and at the other end is the three-nucleotide anticodon where the mRNA will attach (light green).



A cell that is producing proteins has in its cytoplasm a supply of amino acids. But amino acids themselves cannot recognize the codons arranged in sequence along messenger RNA. 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 distinct functions: (1) pick up the appropriate amino acids and (2) recognize the appropriate codons in the mRNA. The unique structure of tRNA molecules enables them to perform both tasks.

As shown on the left in Figure 10.14, a tRNA molecule is made of a single strand of RNA—one polynucleotide chain—consisting of about 80 nucleotides. The chain twists and folds upon itself, forming several double-stranded regions in which short stretches of RNA base-pair with other stretches. At one end of the folded molecule is a special triplet of bases called an **anticodon**. The anticodon triplet is complementary to a codon triplet on mRNA. During translation, the anticodon on the tRNA recognizes a particular codon on the mRNA by using base-pairing rules. At the other end of the tRNA molecule is a site where one specific kind of amino acid attaches. Although all tRNA molecules are similar, there are slightly different versions of tRNA for each amino acid.

Ribosomes

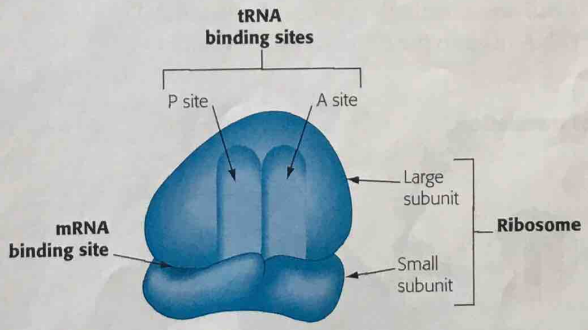
Ribosomes are the organelles in the cytoplasm that coordinate the functioning of mRNA and tRNA and actually make polypeptides. As you can see in **Figure 10.15a**, a ribosome consists of two subunits. Each subunit is made up of proteins and a considerable amount of yet another kind of RNA, **ribosomal RNA (rRNA)**. A fully assembled ribosome has a binding site for mRNA on its small subunit and binding sites for tRNA on its large subunit. **Figure 10.15b** shows how two tRNA molecules get together with an mRNA molecule on a ribosome. One of the tRNA binding sites, the P site, holds the tRNA carrying the growing polypeptide chain, while another, the A site, holds a tRNA carrying the next amino acid to be added to the chain. The anticodon on each tRNA base-pairs with a codon on the mRNA. The subunits of the ribosome act like a vise, holding the tRNA and mRNA molecules close together. The ribosome can then connect the amino acid from the tRNA in the A site to the growing polypeptide. ☑

☑ CHECKPOINT

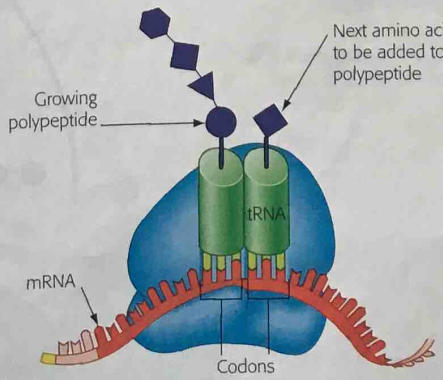
What is an anticodon?

Answer: An anticodon is the base triplet of a tRNA molecule that couples the tRNA to a complementary codon in the mRNA. The base pairing of anticodon to codon is a key step in translating mRNA to a polypeptide.

▼ **Figure 10.15 The ribosome.**



(a) A simplified diagram of a ribosome. Notice the two subunits and sites where mRNA and tRNA molecules bind.



(b) The "players" of translation. When functioning in polypeptide synthesis, a ribosome holds one molecule of mRNA and two molecules of tRNA. The growing polypeptide is attached to one of the tRNAs.

Translation: The Process

Translation is divided into the same three phases as transcription: initiation, elongation, and termination.

Initiation

This first phase brings together the mRNA, the first amino acid with its attached tRNA, and the two subunits of a ribosome. An mRNA molecule, even after splicing, is longer than the genetic message it carries (**Figure 10.16**). Nucleotide sequences at either end of the molecule (pink) are not part of the message, but along with the cap and tail in eukaryotes, they help the mRNA bind to the ribosome. The initiation process determines exactly where translation will begin so that the mRNA codons will be translated into the correct sequence of amino acids. Initiation occurs in two steps, as shown in **Figure 10.17** 1. An mRNA molecule binds to a small ribosomal subunit. A special initiator tRNA then binds to the **start codon**, where translation is to begin on the mRNA. The initiator tRNA carries the amino acid methionine (Met); its anticodon, UAC, binds to the start codon, AUG 2. A large ribosomal subunit binds to the small one, creating a functional ribosome. The initiator tRNA fits into the P site on the ribosome.

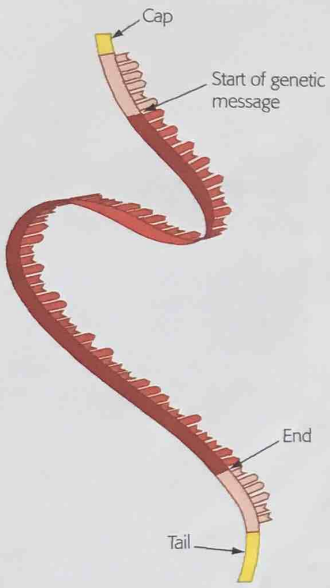
Elongation

Once initiation is complete, amino acids are added one by one to the first amino acid. Each addition occurs in the three-step elongation process shown in **Figure 10.18**. 1 The anticodon of an incoming tRNA molecule, carrying its amino acid, pairs with the mRNA codon in the A site of the ribosome. 2 The polypeptide leaves the tRNA in the P site and attaches to the amino acid on the tRNA in the A site. The ribosome creates a new peptide bond. Now the chain has one more amino acid. 3 The P site tRNA now leaves the ribosome, and the ribosome moves the remaining tRNA, carrying the growing polypeptide, to the P site. 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.

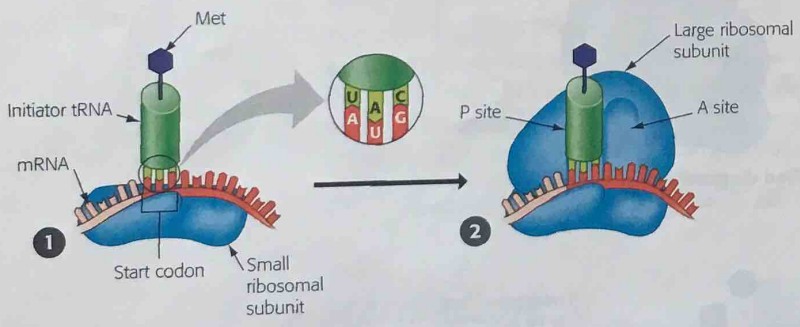
Termination

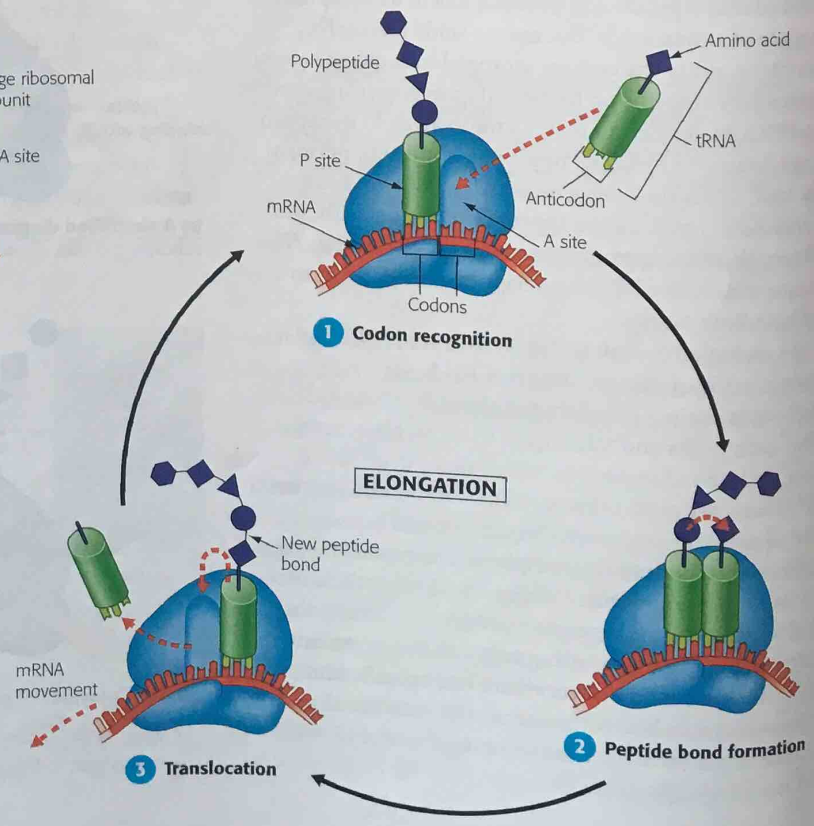
Elongation continues until a **stop codon** reaches the ribosome's A site. Stop codons—UAA, UAG, and UGA—do not code for amino acids but instead tell translation to stop. The completed polypeptide, typically several hundred amino acids long, is freed, and the ribosome splits back into its subunits. ✓

▼ Figure 10.16 A molecule of mRNA. •



▼ Figure 10.17 The initiation of translation.





► **Figure 10.18 The elongation of a polypeptide.** The dashed red arrows indicate movement.

✓ CHECKPOINT

Which of the following does not participate directly in translation: ribosomes, transfer RNA, messenger RNA, DNA?

Answer: DNA

Review: DNA → RNA → Protein

Figure 10.19 reviews the flow of genetic information in the cell, from DNA to RNA to protein. In eukaryotic cells, transcription (DNA → RNA) occurs in the nucleus, and the RNA is processed before it enters the cytoplasm. Translation (RNA → protein) is rapid; a single ribosome can make an average-sized polypeptide in less than a minute. As it is made, a polypeptide coils and folds, assuming its final three-dimensional shape.

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 flow of information originates with the specific sequence of nucleotides in a DNA gene. The gene dictates the transcription of a complementary sequence of nucleotides in mRNA. In turn, the information within the mRNA specifies the sequence of amino acids in a polypeptide. Finally, the proteins that form from the polypeptides determine the appearance and capabilities of the cell and organism.

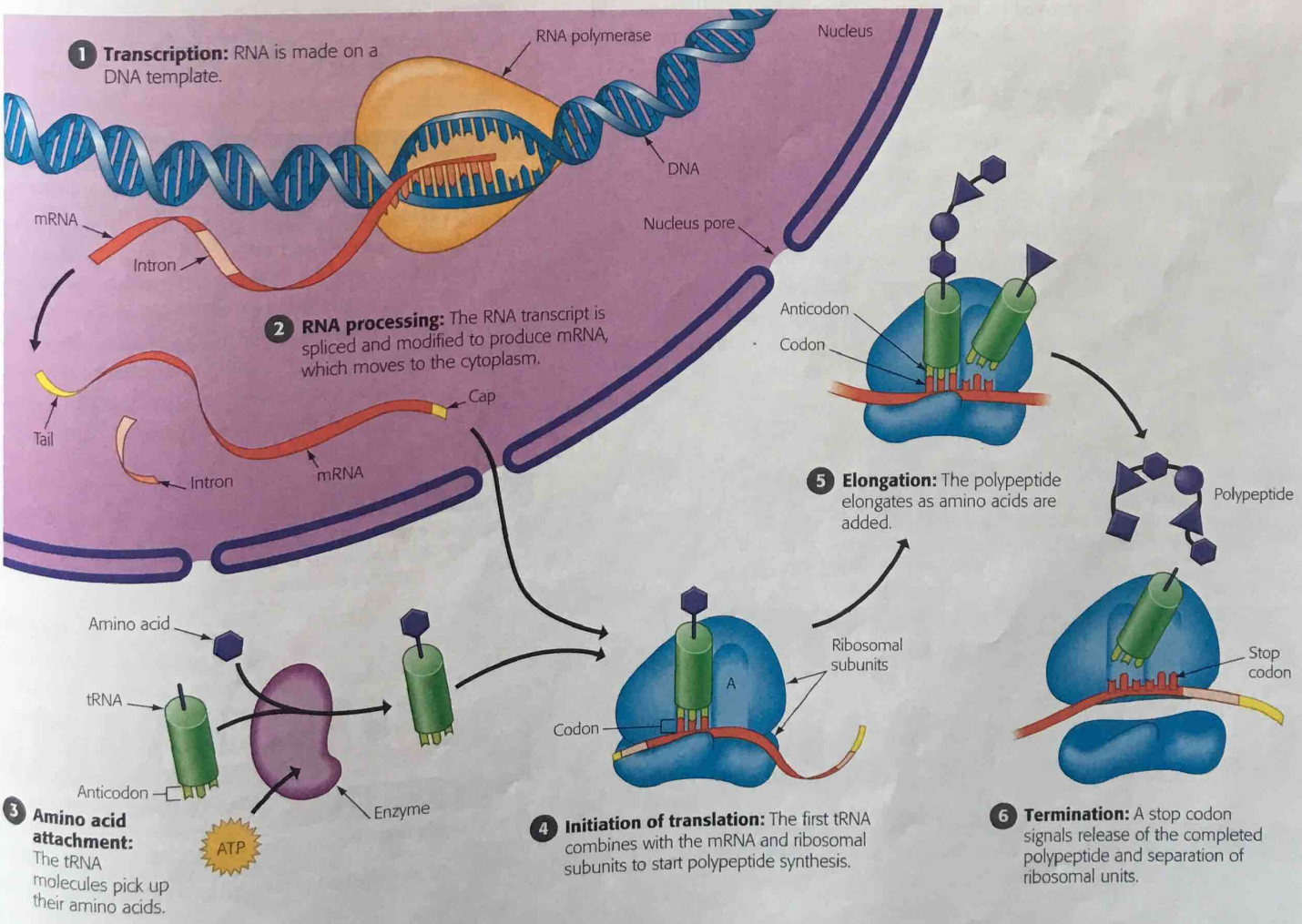
For decades, the DNA → RNA → protein pathway was believed to be the sole means by which genetic information controls traits. In recent years, however, this notion has been challenged by discoveries that point to more complex roles for RNA. (We will explore some of these special properties of RNA in Chapter 11.) ☑

☑ CHECKPOINT

1. Transcription is the synthesis of _____, using _____ as a template.
2. Translation is the synthesis of _____, with one _____ determining each amino acid in the sequence.
3. Which organelle coordinates translation?

Answers: 1. mRNA; DNA 2. protein (polypeptides); codon 3. ribosomes

Figure 10.19 A summary of transcription and translation. This figure summarizes the main stages in the flow of genetic information from DNA to protein in a eukaryotic cell.



Mutations

A single molecular "typo" in DNA can result in a life-threatening disease.

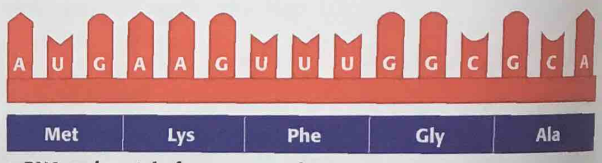
Since discovering how genes are translated into proteins, scientists have been able to describe many heritable differences in molecular terms. For instance, sickle-cell disease can be traced to a change in a single amino acid in one of the polypeptides in the hemoglobin protein (see Figure 3.19). This difference is caused by a single nucleotide difference in the DNA coding for that polypeptide **(Figure 10.20)**.

Any change in the nucleotide sequence of a cell's DNA is called a **mutation**. Mutations can involve large regions of a chromosome or just a single nucleotide pair, as in sickle-cell disease. Occasionally, a base substitution leads to an improved protein or one with new capabilities that enhance the success of the mutant organism and its descendants. Much more often, though, mutations are harmful. Think of a mutation as a typo in a recipe; occasionally, such a typo might lead to an improved recipe, but much more often it will be neutral, mildly bad, or disastrous. Let's consider how mutations involving only one or a few nucleotide pairs can affect gene translation.

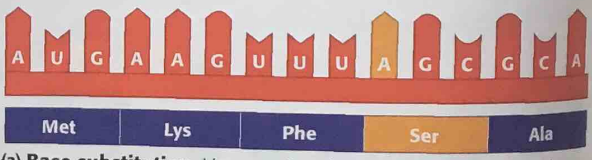
Types of Mutations

Mutations within a gene can be divided into two general categories: nucleotide substitutions and nucleotide insertions or deletions **(Figure 10.21)**. A substitution is the replacement of one nucleotide and its base-pairing partner with another nucleotide pair. For example, in the second row in Figure 10.21, A replaces G in the fourth codon of the mRNA. What effect can a substitution have? Because the genetic code is redundant, some substitutions mutations have no effect at all. For example, if a mutation causes an mRNA codon to change from GAA to GAG, no change in the protein product would result because GAA and GAG both code for the same amino acid (Glu). Such a change is called a silent mutation. In our recipe example, changing "1¼ cup sugar" to "1¼ cup sugor"

▼ Figure 10.21 Three types of mutations and their effects. Mutations are changes in DNA, but they are shown here in mRNA and the polypeptide product.

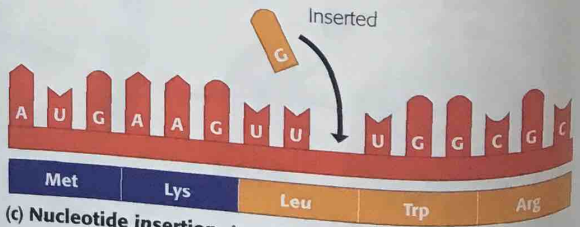


mRNA and protein from a normal gene



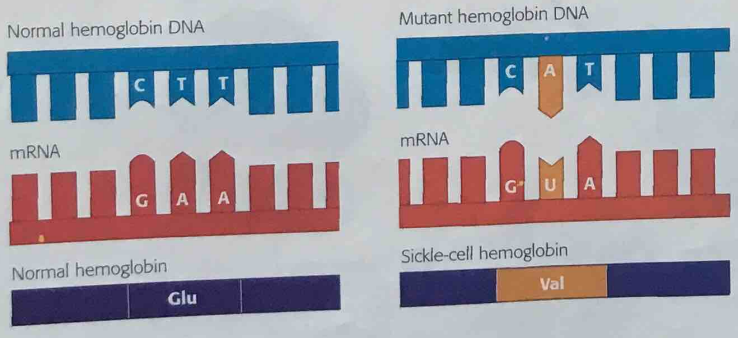
(a) Base substitution. Here, an A replaces a G in the fourth codon of the mRNA. The result in the polypeptide is a serine (Ser) instead of a glycine (Gly). This amino acid substitution may or may not affect the protein's function.

(b) Nucleotide deletion. When a nucleotide is deleted, all the codons from that point on are misread. The resulting polypeptide is likely to be completely nonfunctional.



(c) Nucleotide insertion. As with a deletion, inserting one nucleotide disrupts all codons that follow, most likely producing a nonfunctional polypeptide.

▼ Figure 10.20 The molecular basis of sickle-cell disease. The sickle-cell allele differs from its normal counterpart, a gene for hemoglobin, by only one nucleotide (orange). This difference changes the mRNA codon from one that codes for the amino acid glutamic acid (Glu) to one that codes for valine (Val).



would probably be translated the same way, just like the translation of a silent mutation does not change the meaning of the message.

Other substitutions involving a single nucleotide do change the amino acid coding. Such mutations are called missense mutations. For example, if a mutation causes an mRNA codon to change from GGC to AGC, the resulting protein will have a serine (Ser) instead of a glycine (Gly) at this position. Some missense mutations have little or no effect on the shape or function of the resulting protein; imagine changing a recipe from "1¼ cups sugar" to "1⅓ cups sugar"—this will probably have a negligible effect on your final product. However, other substitutions, as we saw in the sickle-cell case, cause changes in the protein that prevent it from performing normally. This would be like changing "1¼ cups sugar" to "6¼ cups sugar"—this one change is enough to ruin the recipe.

Some substitutions, called nonsense mutations, change an amino acid codon into a stop codon. For example, if an AGA (Arg) codon is mutated to a UGA (stop) codon, the result will be a prematurely terminated protein, which probably will not function properly. In our recipe analogy, this would be like stopping food preparation before the end of the recipe, which is almost certainly going to ruin the dish.

Mutations involving the deletion or insertion of one or more nucleotides in a gene, called frameshift mutations, often have disastrous effects (see Figure 10.21b and c). Because mRNA is read as a series of nucleotide triplets during translation, adding or subtracting nucleotides may alter the triplet grouping of the genetic message. All the nucleotides after the insertion or deletion will be regrouped into different codons. Consider this recipe example: Add one cup egg nog. Deleting the second letter produces an entirely nonsensical message—ado nec upe ggn og—which will not produce a useful product. Similarly, a frameshift mutation most often produces a nonfunctioning polypeptide.

Mutagens

Mutations can occur in a number of ways. Spontaneous mutations result from random errors during DNA replication or recombination. Other sources of mutation are physical and chemical agents called **mutagens**. The most common physical mutagen is high-energy radiation, such as X-rays and ultraviolet (UV) light. Chemical mutagens are of various types. One type, for example, consists of chemicals that are similar to normal DNA bases but that base-pair incorrectly when incorporated into DNA.

Because many mutagens can act as carcinogens, agents that cause cancer, you would do well to avoid them as much as possible. What can you do to avoid exposure to mutagens? Several lifestyle practices can help, including not smoking and wearing protective clothing and sunscreen to minimize direct exposure to the sun's UV rays. But such precautions are not foolproof, and it is not possible to avoid mutagens (such as UV radiation and secondhand smoke) entirely.

Although mutations are often harmful, they can also be beneficial, both in nature and in the laboratory. Mutations are one source of the rich diversity of genes in the living world, a diversity that makes evolution by natural selection possible **(Figure 10.22)**. Mutations are also essential tools for geneticists. Whether naturally occurring or created in the laboratory, mutations are responsible for the different alleles needed for genetic research. ☑

☑ CHECKPOINT

1. What would happen if a mutation changed a start codon to some other codon?
2. What happens when one nucleotide is lost from the middle of a gene?

Answers: **1.** mRNA transcribed from the mutated gene would be nonfunctional because ribosomes would not initiate translation. **2.** In the mRNA, the reading of the triplets downstream from the deletion is shifted, leading to a long string of incorrect amino acids in the polypeptide.

▼ **Figure 10.22 Mutations and diversity.** Mutations are one source of the diversity of life visible in this scene from the Isle of Staffa, in the North Atlantic.

Viruses and Other Noncellular Infectious Agents

Viruses share some of the characteristics of living organisms, such as having 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. (See Figure 1.4 to review the properties of life.) A **virus** is an infectious particle consisting of little more than "genes in a box": a bit of nucleic acid wrapped in a protein coat and, in some cases, an envelope of membrane **(Figure 10.23)**. A virus cannot reproduce on its own, and thus it can multiply only by infecting a living cell and directing the cell's molecular machinery to make more viruses. In this section, we'll look at viruses that infect different types of host organisms, starting with bacteria.

▲ **Figure 10.23 Adenovirus.** A virus that infects the human respiratory system, an adenovirus consists of DNA enclosed in a protein coat shaped like a 20-sided polyhedron, shown here in a computer-generated model that is magnified approximately 500,000 times the actual size. At each corner of the polyhedron is a protein spike, which helps the virus attach to a susceptible cell.

Bacteriophages

Viruses that attack bacteria are called **bacteriophages** ("bacteria-eaters"), or **phages** for short. **Figure 10.24** shows a micrograph of a bacteriophage called T4 infecting an *Escherichia coli* bacterium. The phage consists of a molecule of DNA enclosed within an elaborate structure made of proteins. The "legs" of the phage bend when they touch the cell surface. The tail is a hollow rod enclosed in a springlike sheath. As the legs bend, the spring compresses, the bottom of the rod punctures the cell membrane, and the viral DNA passes from inside the head of the virus into the cell.

Once they infect a bacterium, most phages enter a reproductive cycle called the **lytic cycle**. The lytic cycle gets its name from the fact that after many copies of the phage are produced within the bacterial cell, the bacterium lyses (breaks open). Some viruses can also reproduce by an alternative route—the **lysogenic cycle**. During a lysogenic cycle, viral DNA replication occurs without phage production or the death of the cell.

▼ **Figure 10.24 Bacteriophages (viruses) infecting a bacterial cell.**

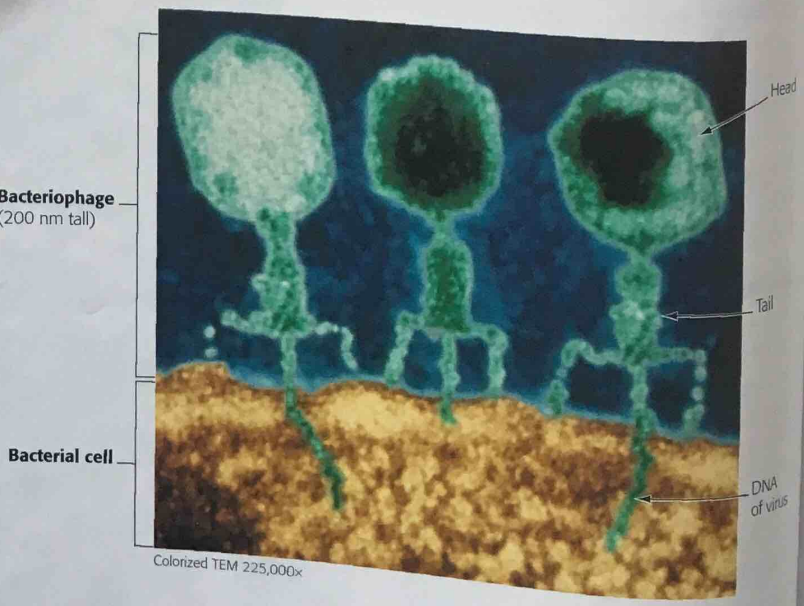


Figure 10.25 illustrates the two kinds of cycles for a phage named lambda that can infect *E. coli* bacteria. At the start of infection, (1) lambda binds to the outside of a bacterium and injects its DNA inside. (2) The injected lambda DNA forms a circle. In the lytic cycle, this DNA immediately turns the cell into a virus-producing factory. (3) The cell's own machinery for DNA replication, transcription, and translation is hijacked by the virus and used to produce copies of the virus. (4) The cell lyses, releasing the new phages.

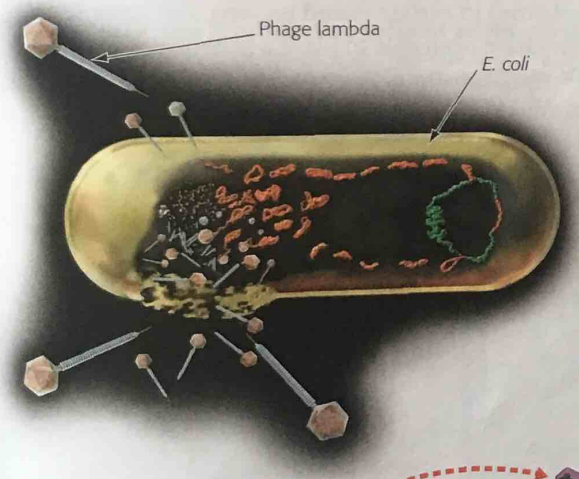
In the lysogenic cycle, (5) the viral DNA is inserted into the bacterial chromosome. Once there, the phage DNA is referred to as a **prophage**, and most of its genes are inactive. Survival of the prophage depends on the reproduction of the cell where it resides. (6) The host cell replicates the prophage DNA along with its cellular DNA and then, upon dividing, passes on both the prophage and the cellular DNA to its two daughter cells. A single infected bacterium can quickly give rise to a large population of bacteria that all carry prophages. The prophages may remain in the bacterial cells indefinitely. (7) Occasionally, however, a prophage leaves its chromosome; this event may be triggered by environmental conditions such as exposure to a mutagen. Once separate, the lambda DNA usually switches to the lytic cycle, which results in the production of many copies of the virus and lysing of the host cell.

Sometimes the few prophage genes active in a lysogenic bacterial cell can cause medical problems. For example, the bacteria that cause diphtheria, botulism, and scarlet fever would be harmless to people if it were not for the prophage genes they carry. Certain of these genes direct the bacteria to produce toxins that make people ill. ☑

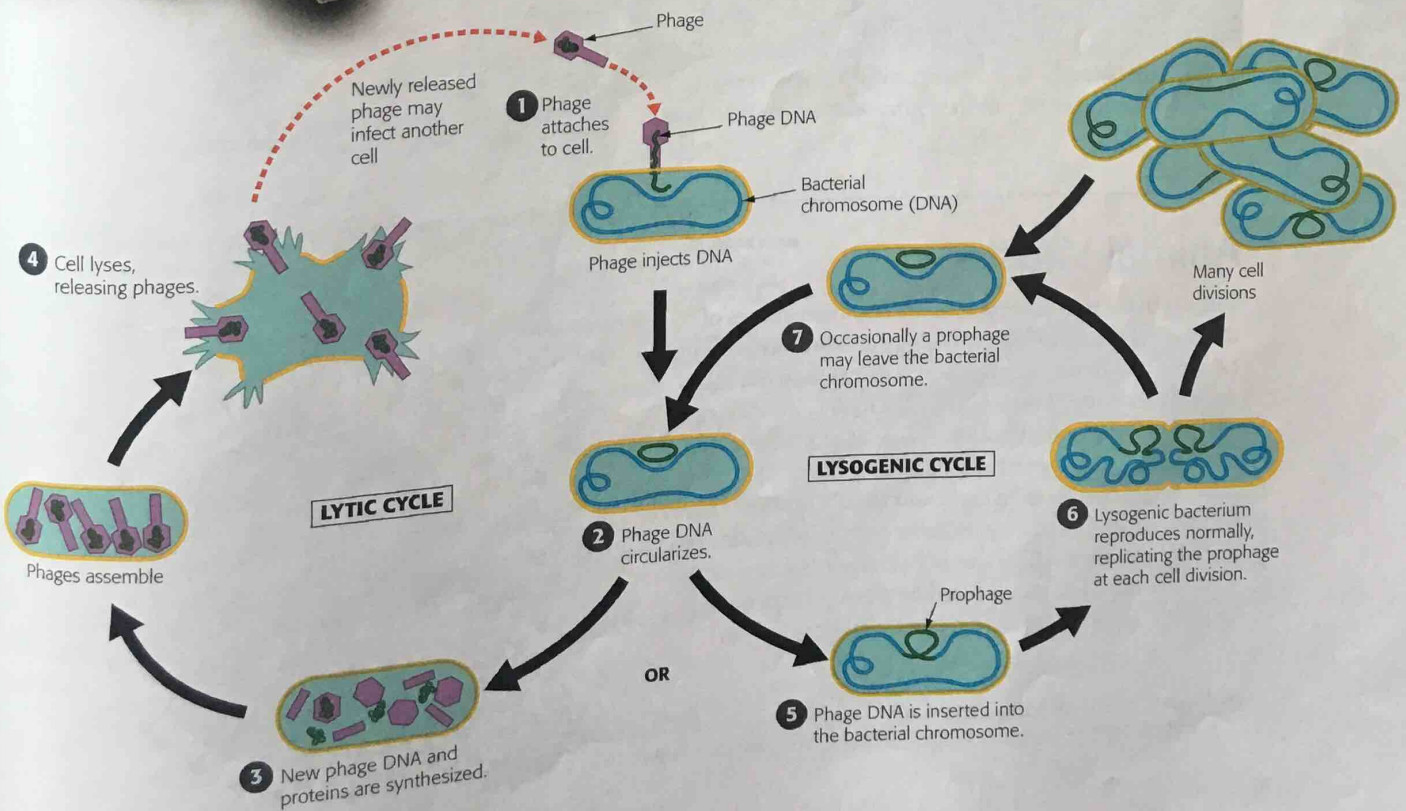
☑ CHECKPOINT

Describe one way some viruses can perpetuate their genes without immediately destroying the cells they infect.

Answer: Some viruses can insert their DNA into the DNA of the cell they infect (the lysogenic cycle). The viral DNA is replicated along with the cell's DNA every time the cell divides.



▼ Figure 10.25 Alternative phage reproductive cycles. Certain phages can undergo alternative reproductive cycles. After entering the bacterial cell, the phage DNA can either integrate into the bacterial chromosome (lysogenic cycle) or immediately start the production of progeny phages (lytic cycle), destroying the cell. Once it enters a lysogenic cycle, the phage's DNA may be carried in the host cell's chromosome for many generations.



Plant Viruses

Viruses that infect plant cells can stunt plant growth and diminish crop yields. Most known plant viruses have RNA rather than DNA as their genetic material. Many of them, like the tobacco mosaic virus (TMV) shown in **Figure 10.26**, are rod-shaped with a spiral arrangement of proteins surrounding the nucleic acid. TMV, which infects tobacco and related plants, causing discolored spots on the leaves, was the first virus ever discovered (in 1930).

To infect a plant, a virus must first get past the plant's epidermis, an outer protective layer of cells. For this reason, a plant damaged by wind, chilling, injury, or insects is more susceptible to infection than a healthy plant. Some insects carry and transmit plant viruses, and farmers and gardeners may unwittingly spread plant viruses through the use of pruning shears and other tools.

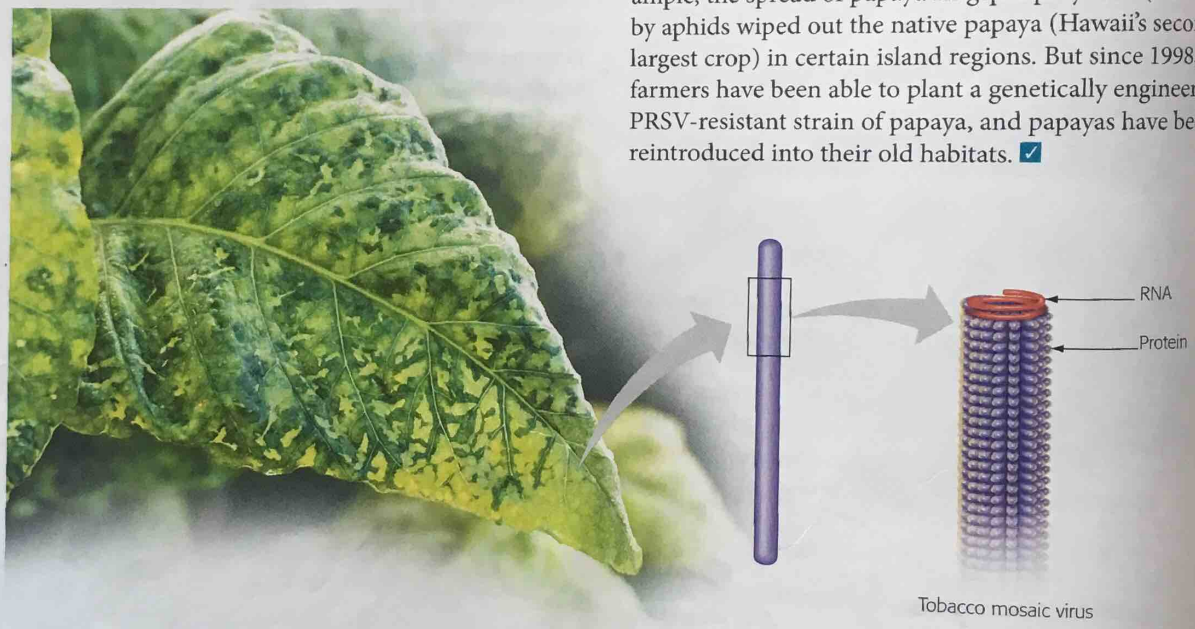
There is no cure for most viral plant diseases, and agricultural scientists focus on preventing infection and on breeding or genetically engineering varieties of crop plants that resist viral infection. In Hawaii, for example, the spread of papaya ringspot potyvirus (PRSV) by aphids wiped out the native papaya (Hawaii's second largest crop) in certain island regions. But since 1998, farmers have been able to plant a genetically engineered PRSV-resistant strain of papaya, and papayas have been reintroduced into their old habitats. ☑

☑ CHECKPOINT

What are three ways that viruses can get into a plant?

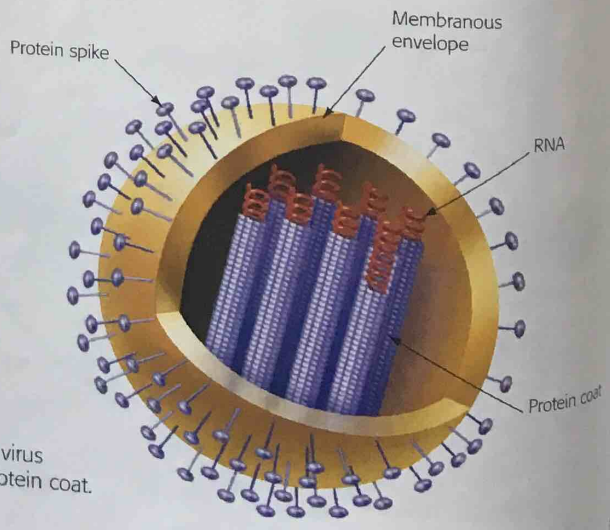
Answer: through lesions caused by injuries, transfer by insects that feed on the plant, and contaminated farming or gardening tools

▼ **Figure 10.26 Tobacco mosaic virus.** The photo shows the mottling of leaves in tobacco mosaic disease. The rod-shaped virus causing the disease has RNA as its genetic material.



Animal Viruses

Viruses that infect animal cells are common causes of disease. As discussed in the Biology and Society section, no virus is a greater human health threat than the influenza (flu) virus **(Figure 10.27)**. Like many animal viruses, this one has an outer envelope made of phospholipid membrane, with projecting spikes of protein. The envelope enables the virus to enter and leave a host cell. Many viruses, including those that cause the flu, common cold, measles, mumps, AIDS, and polio, have RNA as their genetic material. Diseases caused by DNA viruses include hepatitis, chicken pox, and herpes infections.



► **Figure 10.27 An influenza virus.** The genetic material of this virus consists of eight separate molecules of RNA, each wrapped in a protein coat.

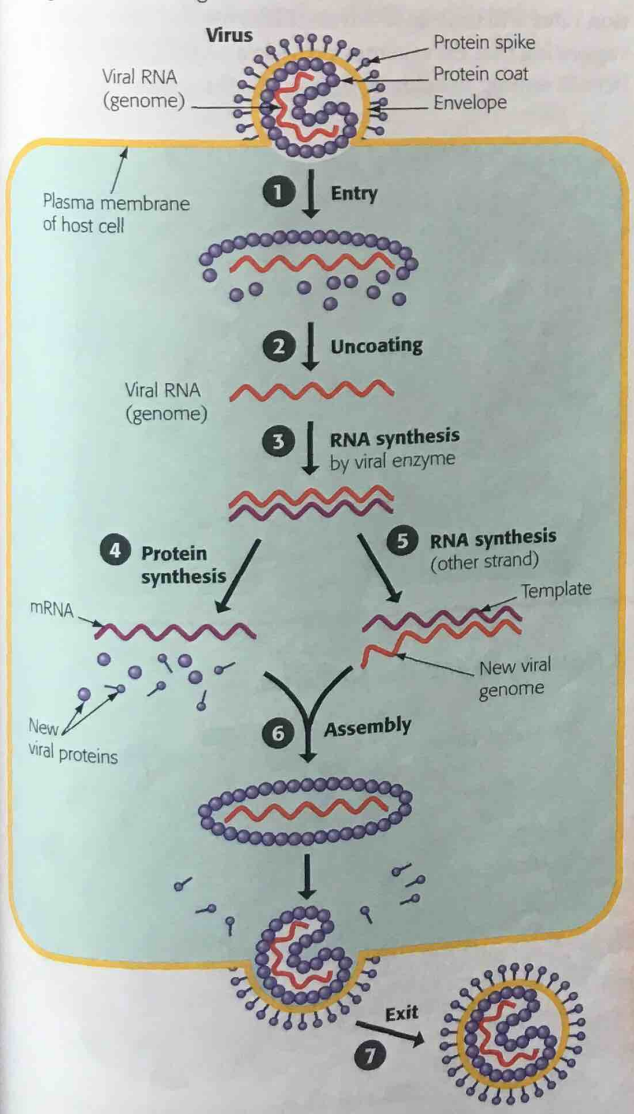
Figure 10.28 shows the reproductive cycle of the mumps virus, a typical RNA virus. Once a common childhood disease characterized by fever and swelling of the salivary glands, mumps has become quite rare in industrialized nations due to widespread vaccination. When the virus contacts a susceptible cell, protein spikes on its outer surface attach to receptor proteins on the cell's plasma membrane. ❶ The viral envelope fuses with the cell's membrane, allowing the protein-coated RNA to enter the cytoplasm. ❷ Enzymes then remove the protein coat. ❸ An enzyme that entered the cell as part of the virus uses the virus's RNA genome as a template for making complementary strands of RNA. The new strands have two functions: ❹ They serve as mRNA for the synthesis of new viral proteins, and ❺ they serve as templates for synthesizing new viral genome RNA. ❻ The new coat proteins assemble around the new viral RNA. ❼ Finally, the viruses leave the cell by cloaking themselves in plasma membrane. In other words, the virus obtains its envelope from the cell, budding off the cell without necessarily rupturing it.

Not all animal viruses reproduce in the cytoplasm. For example, herpesviruses—which cause chicken pox, shingles, cold sores, and genital herpes—are enveloped DNA viruses that reproduce in a host cell's nucleus, and they get their envelopes from the cell's nuclear membrane. Copies of the herpesvirus DNA usually remain behind in the nuclei of certain nerve cells. There they remain dormant until some sort of stress, such as a cold, sunburn, or emotional stress, triggers virus production, resulting in unpleasant symptoms. Once acquired, herpes infections may flare up repeatedly throughout a person's life. More than 75% of American adults carry herpes simplex 1 (which causes cold sores), and more than 20% carry herpes simplex 2 (which causes genital herpes).

The amount of damage a virus causes the body depends partly on how quickly the immune system responds to fight the infection and partly on the ability of the infected tissue to repair itself. We usually recover completely from colds because our respiratory tract tissue can efficiently replace damaged cells. In contrast, the poliovirus attacks nerve cells, which are not usually replaceable. The damage to such cells by polio is permanent. In such cases, the only medical option is to prevent the disease with vaccines.

How effective are vaccines? We'll examine this question next using the example of the flu vaccine. ☑

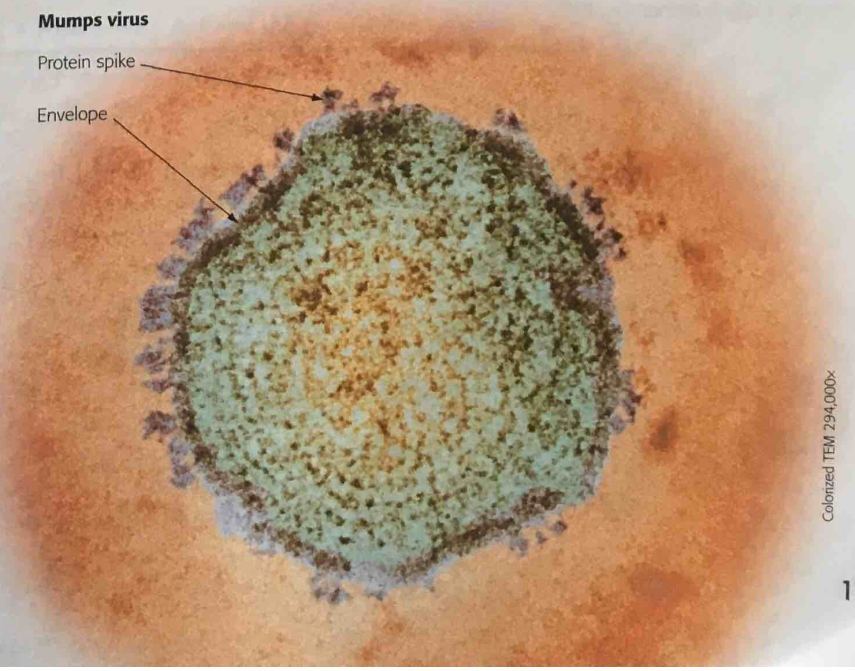
Figure 10.28 The reproductive cycle of an enveloped virus. This virus is the one that causes mumps. Like the flu virus, it has a membranous envelope with protein spikes, but its genome is a single molecule of RNA.



☑ CHECKPOINT

Why is infection by herpesvirus permanent?

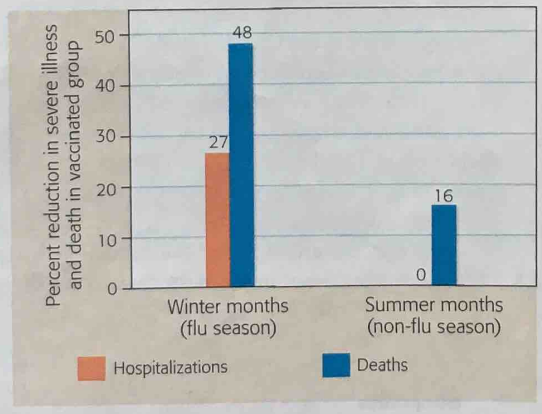
Answer: because herpesvirus leaves viral DNA in the nuclei of nerve cells



The Deadliest Virus THE PROCESS OF SCIENCE

Do Flu Vaccines Protect the Elderly?

Yearly flu vaccinations are recommended for nearly all people over the age of six months. But how can we be sure they are effective? Because elderly people often have weaker immune systems than younger people and because the elderly account for a significant slice of total health-care spending, they are an important population for vaccination efforts. Epidemiologists (scientists who study the distribution, causes, and control of diseases in populations) have made the **observation** that vaccination rates among the elderly rose from 15% in 1980 to 65% in 1996. This observation has led them to ask an important and basic **question:** Do flu vaccines decrease the mortality rate as a result of the flu among those elderly people who receive them? To find out, researchers investigated data from the general population. Their **hypothesis** was that elderly people who were immunized would have fewer hospital stays and deaths during the winter after vaccination. Their **experiment** followed tens of thousands of people over the age of 65 during the ten flu seasons of the 1990s. The **results** are summarized in **Figure 10.29**. People who were vaccinated had a 27% less chance of being hospitalized during the next flu season and a 48% less chance of dying. But could some factor other than flu shots be at play? For example, maybe people who choose to be vaccinated are healthier for other reasons. As a control, the researchers examined health data for the summer (when flu is not a factor). During these months, there was no difference in the hospitalization rates and only 16% fewer deaths for the immunized, suggesting that flu vaccines provide a significant health benefit among the elderly during the flu season.



► **Figure 10.29 The effect of flu vaccines on the elderly.** Receiving a flu vaccine greatly reduced the risk of hospitalization and death in the flu season following the shot. The reduction was much smaller or nonexistent in later summer months.

HIV, the AIDS Virus

The devastating disease **AIDS** (acquired immunodeficiency syndrome) is caused by **HIV** (human immunodeficiency virus), an RNA virus with some nasty twists. In outward appearance, HIV **(Figure 10.30)** resembles the mumps virus. Its envelope enables HIV to enter and leave a cell much the way the mumps virus does. But HIV has a different mode of reproduction. It is a **retrovirus**, an RNA virus that reproduces by means of a DNA molecule, the reverse of the usual DNA → RNA flow of genetic information. These viruses carry molecules of an enzyme called **reverse transcriptase**, which catalyzes reverse transcription: the synthesis of DNA on an RNA template.

▼ **Figure 10.30 HIV, the AIDS virus.**

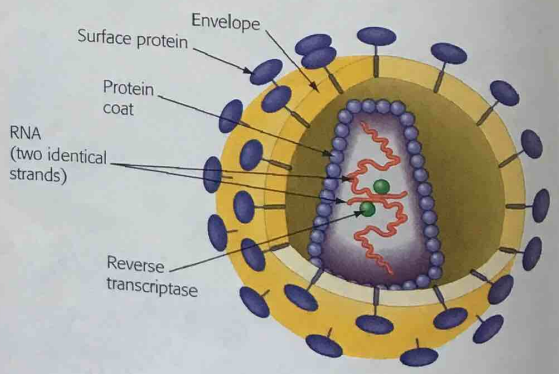


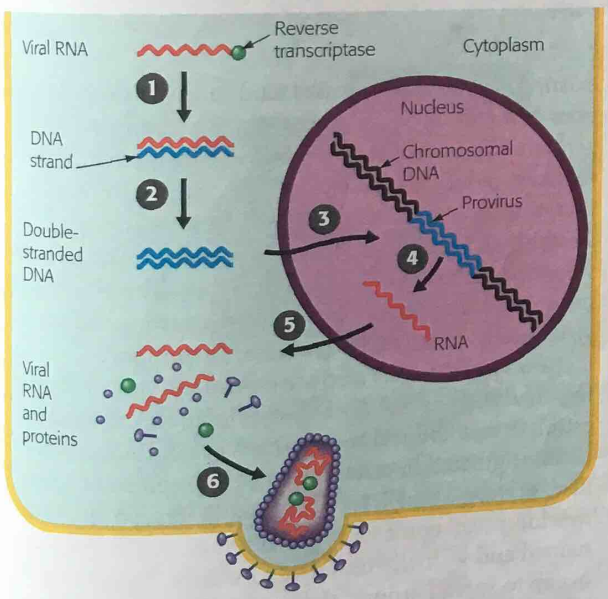
Figure 10.31 illustrates what happens after HIV RNA is uncoated in the cytoplasm of a cell. The reverse transcriptase (green) ❶ uses the RNA as a template to make a DNA strand and then ❷ adds a second, complementary DNA strand. ❸ The resulting double-stranded viral DNA then enters the cell nucleus and inserts itself into the chromosomal DNA, becoming a **provirus**. Occasionally, the provirus is ❹ transcribed into RNA ❺ and translated into viral proteins. ❻ New viruses assembled from these components eventually leave the cell and can then infect other cells. This is the standard reproductive cycle for retroviruses.

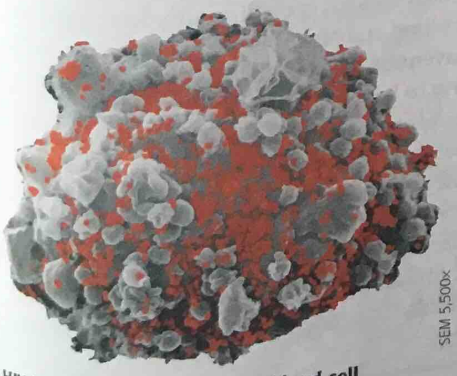
HIV infects and eventually kills several kinds of white blood cells that are important in the body's immune system. The loss of such cells causes the body to become susceptible to other infections that it would normally be able to fight off. Such secondary infections cause the syndrome (a collection of symptoms) that eventually kills AIDS patients. Since it was first recognized in 1981, HIV has infected tens of millions of people worldwide, resulting in millions of deaths.

Although there is as yet no cure for AIDS, its progression can be slowed by two categories of anti-HIV drugs. Both types of medicine interfere with the reproduction of the virus. The first type inhibits the action of enzymes called proteases, which help produce the final versions of HIV proteins. The second type, which includes the drug AZT, inhibits the action of the HIV enzyme reverse transcriptase. The key to AZT's effectiveness is its shape. The shape of a molecule of AZT is very similar to the shape of part of the T (thymine) nucleotide **(Figure 10.32)**. In fact, AZT's shape is so similar to the T nucleotide that AZT can bind to reverse transcriptase, essentially taking the place of T. But unlike thymine, AZT cannot be incorporated into a growing DNA chain. Thus, AZT "gums up the works," interfering with the synthesis of HIV DNA. Because this synthesis is an essential step in the reproductive cycle of HIV, AZT may block the spread of the virus within the body.

Many HIV-infected people in the United States and other industrialized countries take a "drug cocktail" that contains both reverse transcriptase inhibitors and protease inhibitors, and the combination seems to be much more effective than the individual drugs in keeping the virus at bay and extending patients' lives. In fact, the death rate from HIV infection can be lowered by 80% with proper treatment. However, even in combination, the drugs do not completely rid the body of the virus. Typically, HIV reproduction and the symptoms of AIDS return if a patient discontinues the medications. Because AIDS has no cure yet, prevention (namely, avoiding unprotected sex, and staying away from needle sharing) is the only healthy option. ☑

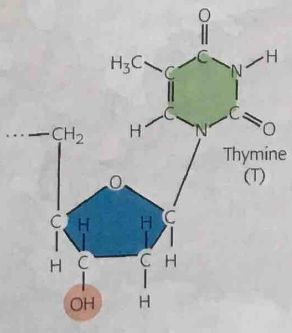
Figure 10.31 The behavior of HIV nucleic acid in an infected cell.

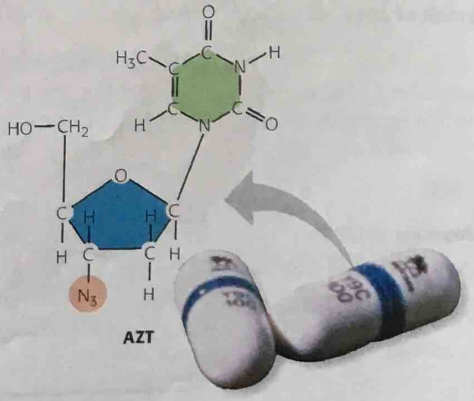




HIV (red dots) infecting a white blood cell

Figure 10.32 AZT and the T nucleotide. The anti-HIV drug AZT (right) has a chemical shape very similar to part of the T (thymine) nucleotide of DNA.





☑ CHECKPOINT

Why is HIV called a retrovirus?

Answer: Because it synthesizes DNA from its RNA genome. This is the reverse ("retro") of the usual DNA → RNA information flow.

Viroids and Prions

Viruses may be small and simple, but they are giants compared to two other classes of pathogens: viroids and prions. Viroids are small, circular RNA molecules that infect plants. Viroids do not encode proteins but can nonetheless replicate in host plant cells using cellular enzymes. These small RNA molecules may cause disease by interfering with the regulatory systems that control plant growth.

Even stranger are infectious proteins called **prions**. Prions cause a number of brain diseases in various animal species, including scrapie in sheep and goats, chronic wasting disease in deer and elk, and mad cow disease (formally called bovine spongiform encephalopathy, or BSE), which infected more than 2 million cattle in the United Kingdom in the 1980s. In humans, prions cause Creutzfeldt-Jakob disease, an extremely rare, incurable, and inevitably fatal deterioration of the brain.

Mad cow disease is caused by an abnormal molecule of protein.

How can a protein cause disease? A prion is thought to be a misfolded form of a protein normally present in brain cells. When a prion enters a cell containing the normal form of protein, the prion somehow converts the normal protein molecules to the misfolded prion version. The abnormal proteins clump together, which may lead to loss of brain tissue (although how this occurs is the subject of much debate and ongoing research). To date, there is no known cure for prion diseases, so hope rests on understanding and preventing the process of infection. ☑

☑ CHECKPOINT

What makes prions so unusual as pathogens?

Answer: Prions, unlike any other infectious agent, have no nucleic acid (DNA or RNA).

The Deadliest Virus EVOLUTION CONNECTION

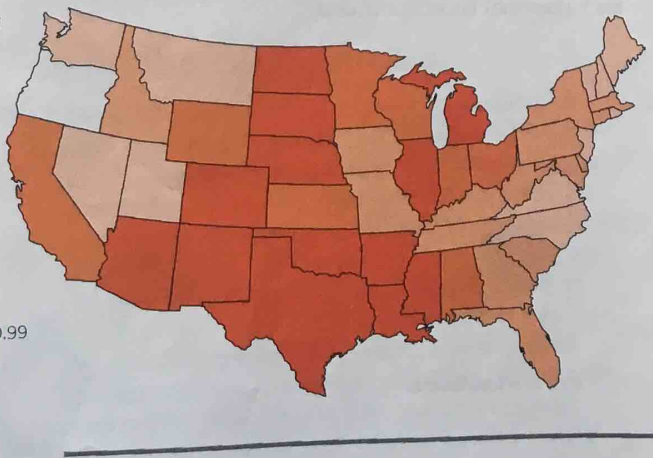
Emerging Viruses

Viruses that suddenly come to the attention of medical scientists are called **emerging viruses**. H1N1 (discussed in the Biology and Society section) is one example; another is West Nile virus, which appeared in North America in 1999 and has since spread to all 48 contiguous U.S. states. West Nile virus is spread primarily by mosquitoes, which carry the virus in blood sucked from one victim and can transfer it to another victim. West Nile virus cases surged in 2012, especially in Texas, claiming nearly 300 lives **(Figure 10.33)**.

How do such viruses burst on the human scene, giving rise to new diseases? One way is by the mutation of existing viruses. RNA viruses tend to have unusually high rates of mutation because errors in replicating their RNA genomes are not subject to proofreading mechanisms that help reduce errors during DNA replication. Some mutations enable existing viruses to evolve into new strains that can cause disease in individuals who have developed resistance to the ancestral virus. This is why we need yearly flu vaccines: Mutations create new influenza virus strains to which people have no immunity.

New viral diseases also arise from the spread of existing viruses from one host species to another. Scientists estimate that about three-quarters of new human diseases originated in other animals. The spread of a viral disease from a small, isolated population can also lead to widespread epidemics. For instance, AIDS went unnamed and virtually unnoticed for decades before it began to spread around the world. In this case, technological and social factors, including affordable international travel, blood transfusions, sexual activity, and the abuse of intravenous drugs, allowed a previously rare human disease to become a global scourge.

Acknowledging the persistent threat that viruses pose to human health, Nobel Prize winner Joshua Lederberg once warned: "We live in evolutionary competition with microbes. There is no guarantee that we will be the survivors." If we someday manage to control HIV, influenza, and other emerging viruses, this success will likely arise from our understanding of molecular biology.



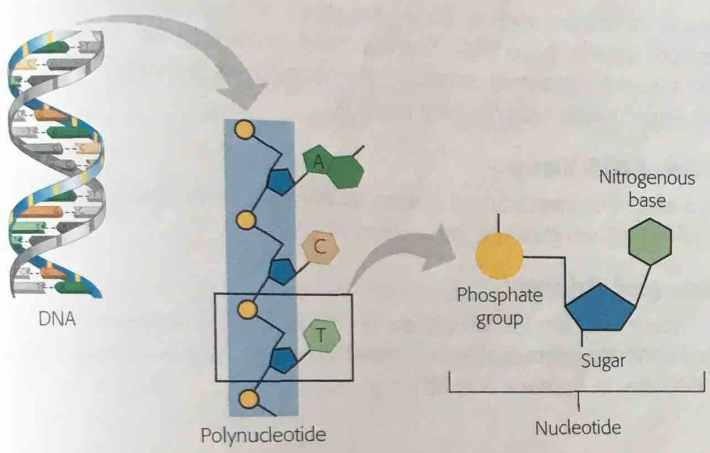
► **Figure 10.33 Mapping the West Nile virus outbreak of 2012.**

Chapter Review

SUMMARY OF KEY CONCEPTS

DNA: Structure and Replication

DNA and RNA Structure



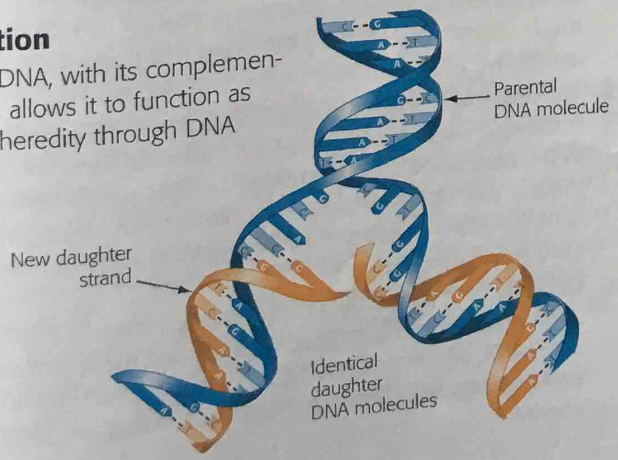
	DNA	RNA
Nitrogenous base	C G A T	C G A U
Sugar	Deoxy-ribose	Ribose
Number of strands	2	1

Structure/Function: Watson and Crick's Discovery of the Double Helix

Watson and Crick worked out the three-dimensional structure of DNA: two polynucleotide strands wrapped around each other in a double helix. Hydrogen bonds between bases hold the strands together. Each base pairs with a complementary partner: A with T, and G with C.

DNA Replication

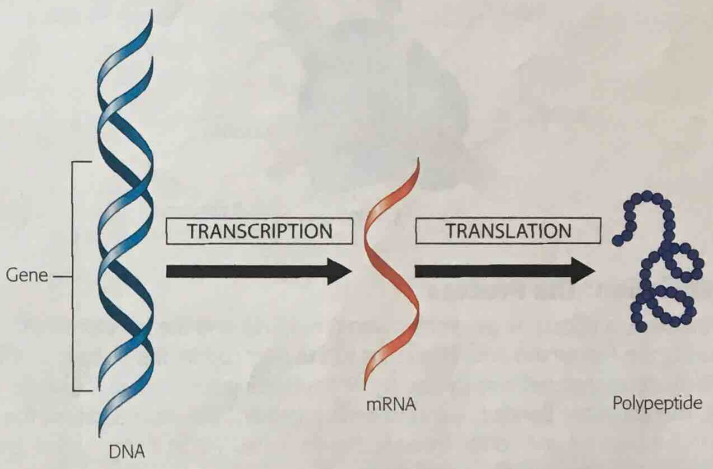
The structure of DNA, with its complementary base pairing, allows it to function as the molecule of heredity through DNA replication.



Information Flow: From DNA to RNA to Protein

How an Organism's Genotype Determines Its Phenotype

The information constituting an organism's genotype is carried in the sequence of its DNA bases. The genotype controls phenotype through the expression of proteins.



From Nucleotides to Amino Acids: An Overview

The DNA of a gene is transcribed into RNA using the usual base-pairing rules, except that an A in DNA pairs with U in RNA. In the translation of a genetic message, each triplet of nucleotide bases in the RNA, called a codon, specifies one amino acid in the polypeptide.

The Genetic Code

In addition to codons that specify amino acids, the genetic code has one codon that is a start signal and three that are stop signals for translation. The genetic code is redundant: There is more than one codon for most amino acids.

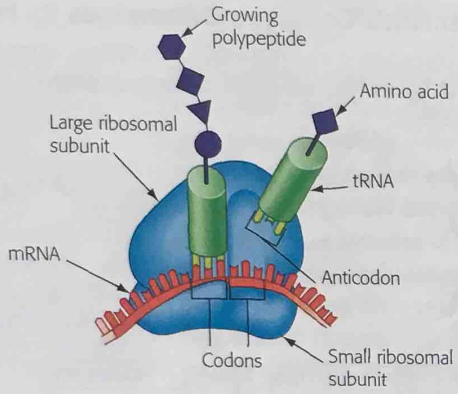
Transcription: From DNA to RNA

In transcription, RNA polymerase binds to the promoter of a gene, opens the DNA double helix there, and catalyzes the synthesis of an RNA molecule using one DNA strand as a template. As the single-stranded RNA transcript peels away from the gene, the DNA strands rejoin.

The Processing of Eukaryotic RNA

The RNA transcribed from a eukaryotic gene is processed before leaving the nucleus to serve as messenger RNA (mRNA). Introns are spliced out, and a cap and tail are added.

Translation: The Players



Translation: The Process

In initiation, a ribosome assembles with the mRNA and the initiator tRNA bearing the first amino acid. Beginning at the start codon, the codons of the mRNA are recognized one by one by tRNAs bearing succeeding amino acids. The ribosome bonds the amino acids together. With each addition, the mRNA moves by one codon through the ribosome. When a stop codon is reached, the completed polypeptide is released.

Review: DNA → RNA → Protein

The sequence of codons in DNA, via the sequence of codons in mRNA, spells out the primary structure of a polypeptide.

Mutations

Mutations are changes in the DNA base sequence, caused by errors in DNA replication or recombination or by mutagens. Substituting, deleting, or inserting nucleotides in a gene has varying effects on the polypeptide and organism.

Type of Mutation	Effect
Substitution of one DNA base for another	**Silent** mutations result in no change to amino acids.
	Missense mutations swap one amino acid for another.
	Nonsense mutations change an amino acid codon to a stop codon.
Insertions or **deletions** of DNA nucleotides	Frameshift mutations can alter the triplet grouping of codons and greatly change the amino acid sequence.

Viruses and Other Noncellular Infectious Agents

Viruses are infectious particles consisting of genes packaged in protein.

Bacteriophages

When phage DNA enters a lytic cycle inside a bacterium, it is replicated, transcribed, and translated. The new viral DNA and protein molecules then assemble into new phages, which burst from the cell. In the lysogenic cycle, phage DNA inserts into the cell's chromosome and is passed on to generations of daughter cells. Much later, it may initiate phage production.

Plant Viruses

Viruses that infect plants can be a serious agricultural problem. Most have RNA genomes. Viruses enter plants via breaks in the plant's outer layers.

Animal Viruses

Many animal viruses, such as flu viruses, have RNA genomes; others, such as hepatitis viruses, have DNA. Some animal viruses "steal" a bit of cell membrane as a protective envelope. Some, such as the herpesvirus, can remain latent inside cells for long periods.

HIV, the AIDS Virus

HIV is a retrovirus. Inside a cell it uses its RNA as a template for making DNA, which is then inserted into a chromosome.

Viroids and Prions

Even smaller than viruses, viroids are small molecules of RNA that can infect plants. Prions are infectious proteins that cause a number of degenerative brain diseases in humans and other animals.

SELF-QUIZ

1. A molecule of DNA contains two polymer strands called _____, made by bonding together many monomers called _____.
2. Name the three parts of every nucleotide.
3. Which of the following correctly ranks nucleic acid structures in order of size, from largest to smallest?
   a. gene, chromosome, nucleotide, codon
   b. chromosome, gene, codon, nucleotide
   c. nucleotide, chromosome, gene, codon
   d. chromosome, nucleotide, gene, codon
4. A scientist inserts a radioactively labeled DNA molecule into a bacterium. The bacterium replicates this DNA molecule and distributes one daughter molecule (double helix) to each of two daughter cells. How much radioactivity will the DNA in each of the two daughter cells contain? Why?
5. The nucleotide sequence of a DNA codon is GTA. What would be the nucleotide sequence of an mRNA molecule transcribed from this DNA? In the process of protein synthesis, a tRNA pairs with the mRNA codon. What is the nucleotide sequence of the tRNA anticodon that corresponds to this mRNA codon? What amino acid is attached to the tRNA (see Figure 10.10)?
6. Describe the process by which the information in a gene is transcribed and translated into a protein. Correctly use these terms in your description: tRNA, amino acid, start codon, transcription, mRNA, gene, codon, RNA polymerase, ribosome, translation, anticodon, peptide bond, stop codon.

7. Match the following molecules with the cellular process or processes in which they are primarily involved.

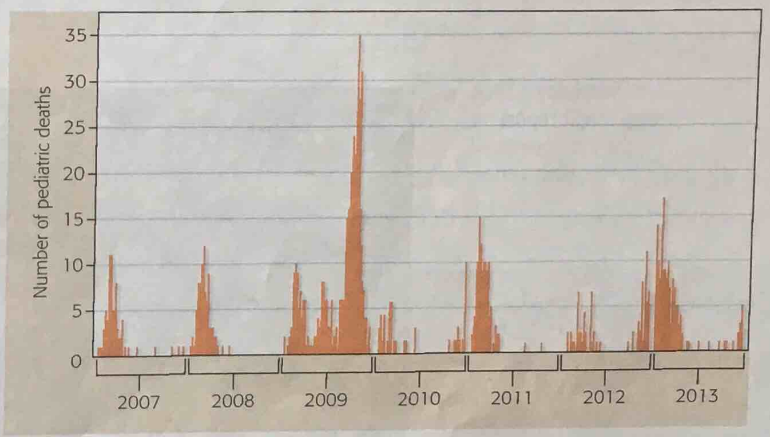
a. ribosomes	1. DNA replication
b. tRNA	2. transcription
c. DNA polymerases	3. translation
d. RNA polymerase	
e. mRNA	

8. A geneticist finds that a particular mutation has no effect on the polypeptide encoded by the gene. This mutation probably involves
   a. deletion of one nucleotide.
   b. alteration of the start codon.
   c. insertion of one nucleotide.
   d. substitution of one nucleotide.
9. Scientists have discovered how to put together a bacteriophage with the protein coat of phage A and the DNA of phage B. If this composite phage were allowed to infect a bacterium, the phages produced in the cell would have
   a. the protein of A and the DNA of B.
   b. the protein of B and the DNA of A.
   c. the protein and DNA of A.
   d. the protein and DNA of B.
10. How do some viruses reproduce without ever having DNA?
11. HIV requires an enzyme called _____ to convert its RNA genome to a DNA version. Why is this enzyme a particularly good target for anti-AIDS drugs? (*Hint*: Would you expect such a drug to harm the human host?)

Answers to these questions can be found in Appendix: Self-Quiz Answers.

THE PROCESS OF SCIENCE

12. A cell containing a single chromosome is placed in a medium containing radioactive phosphate, making any new DNA strands formed by DNA replication radioactive. The cell replicates its DNA and divides. Then the daughter cells (still in the radioactive medium) replicate their DNA and divide, resulting in a total of four cells. Sketch the DNA molecules in all four cells, showing a normal (nonradioactive) DNA strand as a solid line and a radioactive DNA strand as a dashed line.
13. In a classic 1952 experiment, biologists Alfred Hershey and Martha Chase labeled two batches of bacteriophages, one with radioactive sulfur (which only tags protein) and the other with radioactive phosphorus (which only tags DNA). In separate test tubes, they allowed each batch of phages to bind to nonradioactive bacteria and inject its DNA. After a few minutes, they separated the bacterial cells from the viral parts that remained outside the bacterial cells and measured the radioactivity of both portions. What results do you think they obtained? How would these results help them to determine which viral component—DNA or protein—was the infectious portion?
14. **Interpreting Data** The graph below summarizes the number of children who died of all strains of flu from 2007 until 2013. Each bar represents the number of child deaths occurring in one week. Why does the graph have the shape it does, with a series of peaks and valleys? Looking over the Biology and Society section at the start of the chapter, why does the graph reach its highest points near the middle? Based on these data, when does flu season begin and end in a typical year?



BIOLOGY AND SOCIETY

15. Scientists at the National Institutes of Health (NIH) have worked out thousands of sequences of genes and the proteins they encode, and similar analyses are being carried out at universities and private companies. Knowledge of the nucleotide sequences of genes might be used to treat genetic defects or produce lifesaving medicines. NIH and some U.S. biotechnology companies have applied for patents on their discoveries. In Britain, the courts have ruled that a naturally occurring gene cannot be patented. Do you think individuals and companies should be able to patent genes and gene products? Before answering, consider the following: What are the purposes of a patent? How might the discoverer of a gene benefit from a patent? How might the public benefit? What negative effects might result from patenting genes?
16. Your college roommate seeks to improve her appearance by visiting a tanning salon. How would you explain the dangers of this to her?
17. Flu vaccines have been shown to be safe, are very reliable at reducing the risk of hospitalization or death from influenza, and are inexpensive. Should children be required to obtain a flu vaccine before going to school? What about hospital workers before reporting to work? Defend your answers to these questions.