

# 7

C H A P T E R

## CHAPTER OUTLINE

- 7.1 The Central Dogma 134
  - 7.2 Nucleic Acid Structure and Function 134
    - Transcription
    - Translation
  - 7.3 Mutations 141
  - 7.4 Controlling Gene Expression 144
    - Starting Transcription
    - The Right Amount at the Right Time
    - Splicing
  - 7.5 Nucleotide Sequences—The Cell's Legacy 147
    - DNA Replication
    - DNA and Chromosomes
  - 7.6 Using DNA to Our Advantage 150
    - Genetic Engineering and Biotechnology
    - Gene Therapy
- HOW SCIENCE WORKS 7.1: Of Men (and Women!), Microbes, and Molecules 137
- OUTLOOKS 7.1: HIV Infection (AIDS) and Reverse Transcriptase 146
- OUTLOOKS 7.2: Telomeres 148

# DNA and RNA: The Molecular Basis of Heredity

## GOALS AND OBJECTIVES

**Understand the typical flow of genetic information in a cell.**

- Define *gene*, *transcription*, and *translation*.
- Describe how the processes of transcription and translation relate.

**Understand how DNA and RNA control transcription and translation.**

- State the nucleotides commonly found in DNA and RNA.
- Apply the base-pairing rules to predict the nucleotide structure of RNA.
- Explain the use of mRNA, rRNA, and tRNA in the process of translation.
- Accurately use the codon table to predict the amino acid sequence of a protein.

**Understand how mutations affect protein synthesis.**

- Provide examples of silent mutations.
- Provide examples of insertions, deletions, and frameshift mutations.

**Learn the importance of controlling gene expression.**

- State why single cellular and multicellular organisms control gene expression.
- Explain how promoters, transcription factors, and splicing affect transcription.

**Understand the relationship between DNA replication and cell division.**

- Describe DNA replication using base-pairing rules and DNA polymerase.
- Describe how DNA is organized differently in various types of cells.
- Explain how daughter cells inherit a replicated DNA molecule.

**Understand the implications that the DNA code is common to all organisms.**

- Explain how DNA from one organism is used in another organism.
- Explain how DNA can be used to uniquely identify individuals.
- Identify potential medical treatments based on DNA technology.

## 7.1 The Central Dogma

Proteins play a critical role in how cells successfully meet the challenges of living. Cells use proteins to maintain their shape and to speed up important chemical reactions such as photosynthesis and respiration. A cell will not live long if it cannot reliably create the proteins that it needs for survival.

This chapter looks at how cells reliably make proteins. To place these ideas in the proper context, remember that some proteins are enzymes that aid cells by catalyzing chemical reactions. These chemical reactions occur after the enzyme binds its substrate at the enzyme's active site. The enzyme's active site matches the substrate molecule in size, shape, and chemical properties. The size, shape, and chemical properties of an enzyme's active site are due to the combination of the enzyme's amino acids, which are the individual subunits of the enzyme. For the cell to reliably make an enzyme, the cell must be able to control the placement of amino acids in a protein during the synthesis of enzymes. • **enzymes**, p. 92

This control comes from the genetic information stored in the cell's **deoxyribonucleic acid (DNA)** molecule(s). The DNA molecule contains a type of blueprint for making the many different types of proteins that the cell needs. The portion of the DNA strand that codes for a particular protein is called a **gene**. The set of ideas that describes how the cell uses the information stored in DNA is called the **central dogma**. The first step of the central dogma is called **transcription**. Transcription uses DNA as a template to copy genetic information into the form of RNA. In turn, RNA is involved in **translation**. Translation synthesizes the protein using RNA as a template (figure 7.1). Understanding how the steps of the central dogma are carried out is important because we can then understand how mutations arise. Ultimately, mutations can affect the entire organism.

To understand how DNA is able to control the synthesis of enzymes, we need to look at how the structure of proteins and nucleic acids relates to their function—for example, the structure of an enzyme's active site and its function of binding a substrate. Another important structure/function relationship is found in the double-stranded DNA molecule. DNA has four properties that enable it to function as genetic material. It can (1) *store information* that deter-

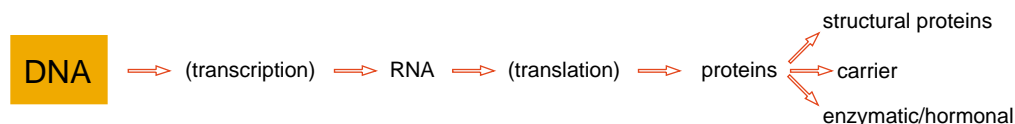
mines the characteristics of cells and organisms; (2) use this information to *direct the synthesis* of structural and regulatory proteins essential to the operation of the cell or organism; (3) *mutate*, or chemically change, and transmit these changes to future generations; and (4) *replicate* by directing the manufacture of copies of itself.

## 7.2 Nucleic Acid Structure and Function

**Nucleic acid** molecules are enormous, complex polymers made up of monomers. Each monomer is a **nucleotide**. Each nucleotide is composed of a sugar molecule containing five carbon atoms, a phosphate group, and a molecule containing nitrogen, which will be referred to as a *nitrogenous base* (figure 7.2). It is possible to classify nucleic acids into two main groups based on the kinds of sugar used in the nucleotides (i.e., DNA and RNA). • **macromolecules**, p. 46

In cells, DNA is the nucleic acid that functions as the original blueprint for the synthesis of proteins. DNA contains the sugar **deoxyribose**, phosphates, and a unique sequence of the nitrogenous bases **adenine (A)**, **guanine (G)**, **cytosine (C)**, and **thymine (T)**. **Ribonucleic acid (RNA)** is a type of nucleic acid that is directly involved in **protein synthesis**. RNA contains the sugar **ribose**, phosphates, and the nitrogenous bases adenine (A), guanine (G), cytosine (C), and **uracil (U)**. DNA and RNA share the nitrogenous bases A, G, and C. Thymine is usually only present in DNA and uracil is usually only present in RNA.

DNA and RNA differ significantly in one other respect: DNA is actually a double molecule. It consists of two flexible strands held together by attractive forces, called *hydrogen bonds*, between paired nitrogenous bases. The two strands are twisted about each other in a coil or double helix (figure 7.3). As the nitrogenous bases pair with each other in the DNA helix, they “fit” into each other like two jigsaw puzzle pieces that interlock. This arrangement is stabilized by weak chemical forces called *hydrogen bonds*. The four bases always pair in a definite way: adenine (A) with thymine (T), and guanine (G) with cytosine (C). The number of hydrogen bonds determines the pairing sets. Adenine (A) and thymine (T) both form two hydrogen bonds, while guanine (G) and cytosine (C) form three hydrogen bonds. Notice that the large molecules (A and



**Figure 7.1**

### Central Dogma Concept Map

DNA → (transcription) → RNA → (translation) → different types of proteins. This concept map outlines the steps of the central dogma. Genetic information in the cell is transferred from DNA to RNA by the process of transcription. The genetic information stored in RNA is then used by the process of translation to produce proteins. Transcription and translation are catalyzed by enzymes. The enzymes involved in this process were themselves produced by the processes of transcription and translation.

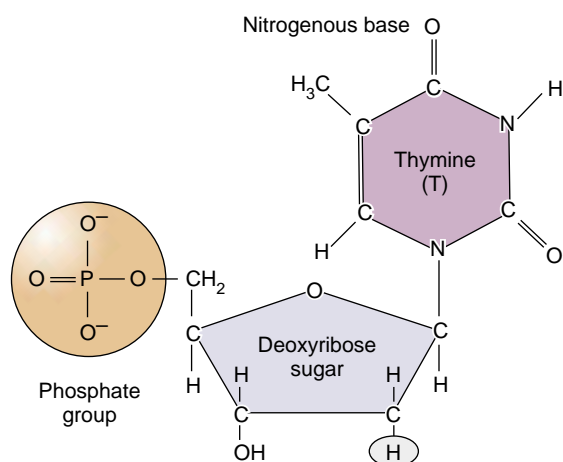
G) pair with the small ones (T and C), thus keeping the DNA double helix a constant width. The bases that pair are said to be **complementary bases** and this bonding pattern is referred to as *base-pairing rules*.

The base-pairing rules are followed throughout the steps of the central dogma. In transcription, nucleotides from DNA pair with nucleotides with RNA. In this case, guanine (G) still pairs with cytosine (C). However, because RNA does not contain thymine (T), adenine (A) in DNA pairs with uracil (U) in RNA. Later in the translation step of the central dogma, RNA nucleotides pair with RNA nucleotides. Again, guanine (G) pairs with cytosine (C) and adenine (A) pairs with uracil (U). The base-pairing rules for each of these situations are summarized in table 7.1.

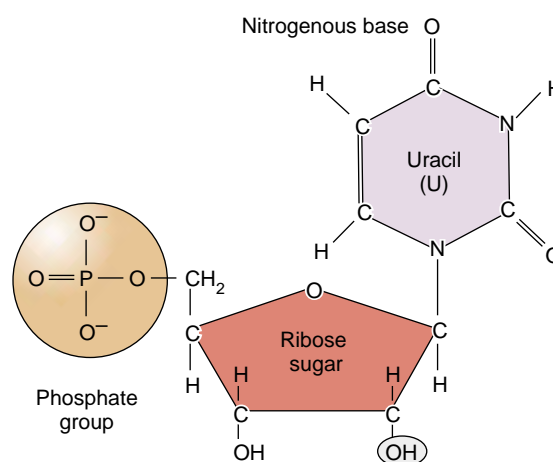
**Table 7.1****BASE-PAIRING RULES**

DNA Paired to DNA		DNA Paired to RNA		RNA Paired to RNA	
G	C	G	C	G	C
C	G	C	G	C	G
A	T	A	U	A	U
T	A	T	A	U	A

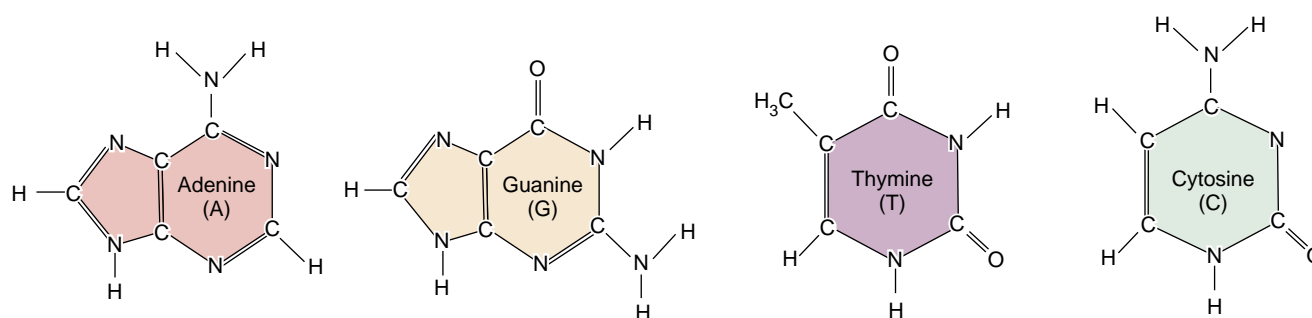
Different types of nucleic acids are capable of base-pairing with each other by following the base-pairing rules. These rules are used throughout the process of the central dogma to allow information to go from DNA to RNA and then ultimately to the amino acid sequence of proteins. In general, the rules are that G and C pair, and A and T (or U) pair. Remember that thymine is only found in DNA and uracil is found only in RNA.



(a) DNA nucleotide



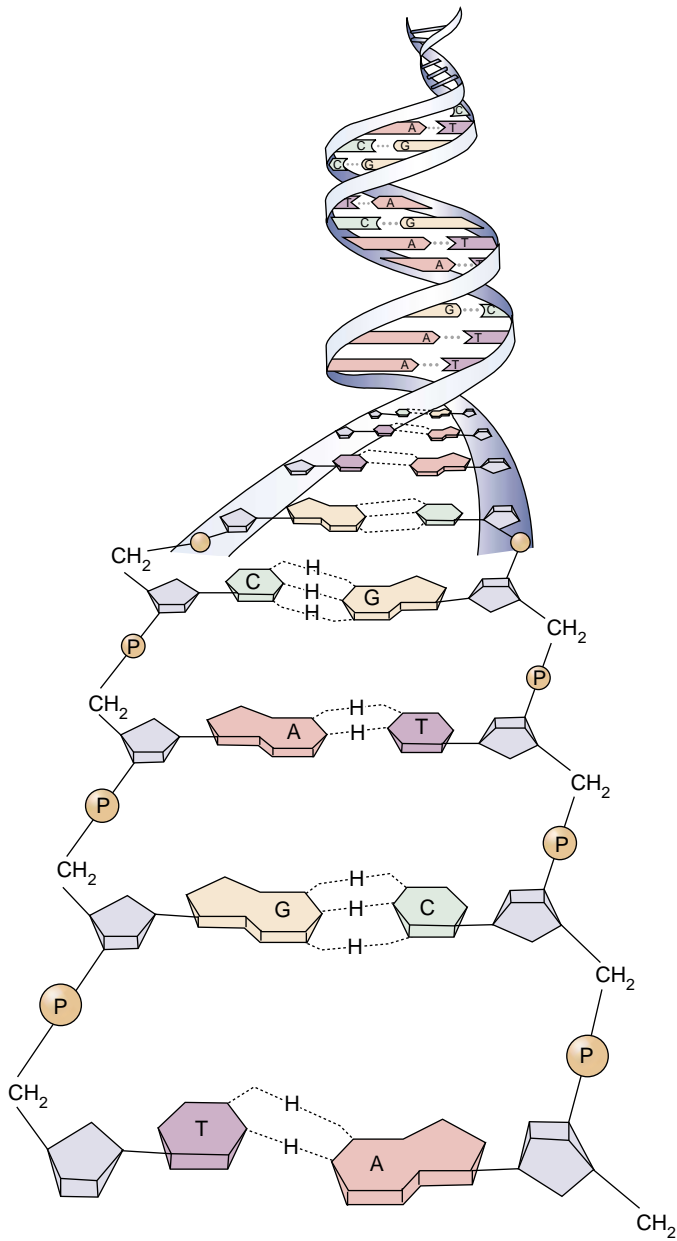
(b) RNA nucleotide



(c) The four nitrogenous bases that occur in DNA

**Figure 7.2****Nucleotide Structure**

The nucleotide is the basic structural unit of all nucleic acids and consists of a sugar, a nitrogenous base, and a phosphate group. Part (a) shows a thymine DNA nucleotide and (b) shows a uracil RNA nucleotide. Notice how both DNA and RNA nucleotides consist of the same three parts—a sugar, a nitrogenous base, and a phosphate group. One chemical difference between DNA and RNA is in the sugar. Notice that the DNA nucleotide has a circled H extending from one of its carbon atoms. In RNA, this same carbon atom has an OH group. The lack of this oxygen atom in DNA is why DNA is called deoxyribonucleic acid. (c) DNA and RNA also differ in the kinds of bases present in their nucleotides. In DNA, the nitrogenous bases can be adenine, guanine, thymine, or cytosine. In RNA, the nitrogenous bases can be adenine, guanine, uracil, or cytosine.



**Figure 7.3**

### Double-Stranded DNA

Polymerized deoxyribonucleic acid (DNA) is a helical molecule. While the units of each strand are held together by covalent bonds, the two parallel strands are interlinked by hydrogen bonds between the paired nitrogenous bases like jigsaw puzzle pieces. A and T pair with two hydrogen bonds. G and C pair with three hydrogen bonds.

The genetic information of DNA is in the form of a chemical code. The **DNA code** is the *order* of the nucleotides. When the coded information is **expressed**, it guides the assembly of particular amino acids into a specific protein. You can “write” a message in the form of a stable DNA molecule

by combining the four different DNA nucleotides (A, T, G, C) in particular sequences. In the cell, the four DNA nucleotides are used in a cellular alphabet that only consists of four letters. The letters of this alphabet are arranged in sets of three to construct words that are used to determine which amino acid is needed during translation. Each word, or **codon**, in the mRNA is always three letters (nucleotides) long and later is responsible for the placement of one amino acid in a protein. It is the sequence of nucleotides (A, T, G, C) in DNA that ultimately dictates which amino acids are used to synthesize a protein. The sequence of nucleotides and the base-pairing rules allow nucleic acids to control protein synthesis.

The fact that DNA has a sequence of nucleotides and is double-stranded allows it to fulfill all of its functions as genetic material. The linear sequence of **nitrogenous bases** in DNA allows the *storage of information* and *directs the synthesis* of proteins through base-pairing rules. The implication of this process is that changes in the linear sequence of nitrogenous bases can result in **mutations**, or changes in the amino acid sequence, used to create a protein. These changes can alter how that protein functions (How Science Works 7.1).

### Transcription

Transcription is the process of using DNA as a **template** to synthesize RNA. DNA functions in a manner that is similar to a reference library that does not allow its books to circulate. Information from the original copies of the books must be copied for use outside of the library so that the originals are not damaged or destroyed. A temporary copy is made of the necessary information in a DNA molecule so that protein synthesis can occur. This operation is called *transcription* (*scribe* = to write), which means to transfer data from one form to another. In this case, the data are copied from DNA language to RNA language. The same base-pairing rules that were discussed earlier in this section apply to the process of transcription. Using this process, the genetic information stored as a DNA chemical code is carried in the form of an RNA molecule. Later, this RNA molecule will be used as a template for creating a protein because it is RNA that is used to guide the assembly of amino acids into structural, carrier, and regulatory proteins. Without the process of transcription, genetic information would be useless in directing cell functions.

Although many types of RNA can be synthesized, the three most important are **messenger RNA (mRNA)**, **transfer RNA (tRNA)**, and **ribosomal RNA (rRNA)**. Each type of RNA is used for a unique purpose within the central dogma. Messenger RNA (mRNA) carries the blueprint for making the necessary enzyme. Transfer RNA (tRNA) and ribosomal

## HOW SCIENCE WORKS 7.1



## Of Men (and Women!), Microbes, and Molecules

As early as the 1920s, scientists didn't really understand the molecular basis of heredity. They partly understood genetics in terms of the odds that a given trait would be passed on to an individual in the next generation. This "probability" model of genetics left some questions unanswered:

- What is the nature of the genetic information?
- How does the cell use the genetic information?

As is often the case in science, serendipity played a large role in answering these questions. In 1928, a medical doctor, Frederick Griffith, was studying different bacterial strains that caused pneumonia. One of the bacterial strains killed mice very quickly and was extremely virulent. The other strain was not virulent. Griffith observed something unexpected when *dead* bacterial cells of the virulent strain were mixed with *living* cells of the less-virulent strain: The less-virulent strain took on the virulent characteristics of the dead strain. This observation was the first significant step in understanding the molecular basis of genetics because it provided scientists with a situation wherein the scientific method could be applied to ask questions and take measurements about the molecular basis of genetics. Until this point, scientists had lacked a method to provide supporting data.

This spurred the scientific community for the next 14 years to search for the identity of the "genetic molecule." A common hypothesis was that the genetic molecule would be one of the macromolecules—carbohydrates, lipids, proteins, or nucleic acids. During that period, many advances were made in how researchers studied cells. Many of the top minds in the field had formulated the hypothesis that protein was the genetic molecule. They had very good support for this hypothesis, too. Their argument boiled down to two ideas: The first idea is that proteins are found everywhere in the cell. It follows that if proteins were the genetic information, they would be found wherever that information would be used. The other reason is that proteins are structurally and chemically very complex. They are made up of 20 different monomers (amino acids) and come in a wide variety of sizes and shapes. This complexity could be used to account for all the genetic variety we observe in nature. On the other hand, very few, if any, scientists seriously considered the notion that DNA was the heritable material. After all, it was only found in the nucleus and it only consisted of four different monomers (nucleotides). How could this molecule account for the genetic complexity of life?

In 1944, Oswald Avery and his colleagues provided the first evidence that DNA was the genetic molecule. They performed an experiment similar to Griffith's, except they used purified samples of protein, DNA, lipids, and carbohydrates. The scientific community was highly skeptical of these results for two reasons: (1) They hadn't expected this result. They expected the genetic molecule to be protein. More importantly, (2) they didn't know how to explain how DNA could function as the genetic molecule. Because of this mind set, Avery's data was largely disregarded on

the rationale that his samples were impure. This objection was a factor that he controlled for in his scientific design. He reported over 99% purity in the tested DNA samples. It took 8 additional years and a different type of experiment to concretely establish DNA as the genetic molecule.

Alfred Hershey and Martha Chase carried out this definitive experiment in 1952. Their experiment was attractive because it used a relatively simple genetic system—a bacterial phage. A phage is a type of virus that uses a bacterial cell as its host and only consists of DNA and protein. Hershey and Chase hypothesized that the phage genetic information needed to enter the bacterial cell to create new phage. By radioactively labeling the DNA and the protein of the phage, Hershey and Chase were able to track that the DNA went into the bacterial cell, while very little protein did. They reasoned that DNA must be the genetic information.

The scientific community then turned toward the issue of determining how DNA could work as the heritable material. Scientists expected that the genetic molecule would have to do a number of things such as store information, disseminate information, be able to mutate, and be able to replicate itself. Their hypothesis was that the answer was hidden in the structure of the DNA molecule itself.

Investigation of how DNA functioned as the cell's genetic information took a wide variety of different strategies. Some scientists looked at DNA from different organisms. They found that in nearly every organism, the guanine (G) and cytosine (C) nucleotides were present in equal amounts. The same held true for adenine (A) and thymine (T). Later, this provided the basis for establishing the nucleic acid base-pairing rules.

Rosalind Franklin carried out another technique called X-ray crystallography. X-ray crystallography provides clues about molecular structure by recording the reflection pattern of X rays that have been fired at a crystallized chemical sample. Franklin was able to determine the following information from her experiments:

- DNA's helical shape
- DNA's width
- DNA's composition of two parallel strands
- DNA's repeating motifs that occur along the length of the molecule

Finally, two young scientists, James Watson and Francis Crick, put it all together. They simply listened to and read the information that was being discussed in the scientific community. Their key role was the assimilation of all the data. They recognized the importance of the X-ray crystallography data in conjunction with the organic structures of the nucleotides and the data that established the base-pairing rules. Together, they created a model for the structure of DNA that could account for all the things that a genetic molecule must do. They published an article describing this model in 1952. Ten years later, they were awarded the Nobel Prize for their work.



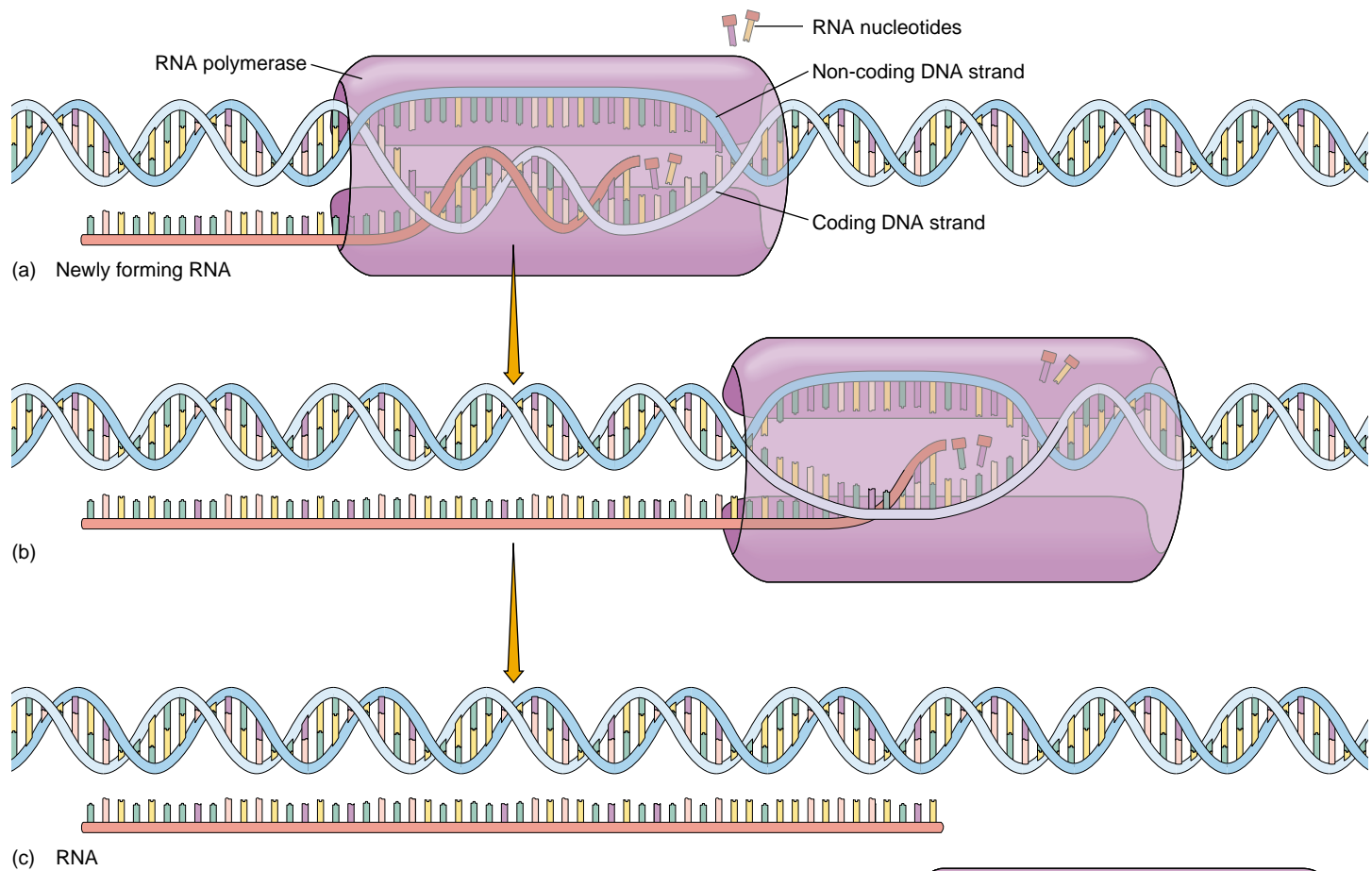
RNA (rRNA) are used later for the actual process of assembling the amino acids (table 7.2).

The actual process of transcription begins in the nucleus by separating the two strands of the double-stranded DNA with an enzyme called **RNA polymerase**. Separating the two strands of DNA exposes their nitrogenous bases so that other enzymes that carry out transcription can read one of the two DNA strands. The strand of DNA that serves as a template for the synthesis of RNA is the **coding strand**. The coding strand is only read in one direction. The strand of DNA that is not read directly by the enzymes is the *non-coding* strand of DNA. The RNA polymerase enzyme catalyzes the dehydration reactions between individual nucleotides so that a large RNA molecule is produced (figure 7.4). • enzymes, p. 92

**Table 7.2**

**TYPES OF RNA**

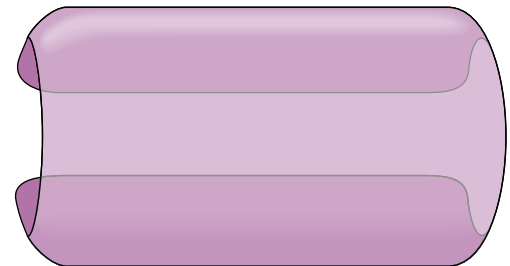
Type of RNA	Description
mRNA	A linear molecule that carries the gene's information from the cell's nucleus to the cytoplasm
tRNA	An adaptor molecule that aids in the pairing of amino acids to the correct codons of the mRNA
rRNA	Important part of ribosomal molecular structure and the ribosome's ability to carry out translation



**Figure 7.4**

**Transcription of an RNA Molecule**

This summary illustrates the basic events that occur during transcription. (a) An enzyme attaches to the DNA at a point that allows it to separate the complementary strands. (b) As RNA polymerase moves down the DNA strand, new complementary RNA nucleotides are base-paired to one of the exposed DNA strands. The base-paired RNA nucleotides are linked together by RNA polymerase to form a new RNA molecule that is complementary to the nucleotide sequence of the DNA. (c) The newly formed (transcribed) RNA is then separated from the DNA molecule and used by the cell.



## Translation

Translation is the process of using the information in RNA to direct the ordered assembly of amino acids. The word “translation” refers to the fact that nucleic acid language is being changed to protein language. To translate mRNA language into protein language, a translation dictionary is necessary. Table 7.3 shows an amino acid–mRNA nucleic acid dictionary. A 3-nucleotide combination is correlated to the single amino acid that is required in the process of translation. Each 3-nucleotide combination is called a *codon*. Consider that each codon codes for one, and only one, amino acid. The codon UUU corresponds to only phenylalanine (Phe). However, notice that more than one mRNA codon may code for the same amino acid. Phenylalanine (Phe) can be coded for by the UUU codon and the UUC codon. This is possible because there are only 20 amino acids and 64 different codons (table 7.4). The 64 codons listed in table 7.3 are all the possible ways that the four different nucleotides of RNA can be arranged in combinations of three.

The construction site of the protein molecules (i.e., the translation site) is on the **ribosome** in the cytoplasm of the cell. The ribosome is a cellular organelle that serves as the meeting place for mRNA and the tRNAs that carry amino acid building blocks. There are many ribosomes in a cell. The mRNA and the tRNAs were synthesized by the process of transcription in the cell’s nucleus and then moved to the cell’s cytoplasm. Ribosomes can be found free in the cytoplasm or attached to the endoplasmic reticulum (ER). Proteins destined to be part of the cell membrane or packaged for export from the cell are synthesized on ribosomes attached to the endoplasmic reticulum. Proteins that are to perform their function in the cytoplasm are synthesized on unattached ribosomes.

While we have initially predicted the amino acid sequence of the protein by using the codon table, the cellular process of translation is a bit more involved. The process of translation can be broken down into three basic processes, (1) initiation; (2) elongation; and (3) termination. The processes of initiation and termination happen only once for

**Table 7.3**

### AMINO ACID–NUCLEIC ACID DICTIONARY

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

**Table 7.4****THE 20 COMMON AMINO ACIDS AND THEIR ABBREVIATIONS**

These are the 20 common amino acids used in the protein synthesis operation of a cell. Each has a known chemical structure.

Amino Acid	Three-Letter Abbreviation	Amino Acid	Three-Letter Abbreviation
alanine	Ala	leucine	Leu
arginine	Arg	lysine	Lys
asparagine	Asn	methionine	Met
aspartic acid	Asp	phenylalanine	Phe
cysteine	Cys	proline	Pro
glutamic acid	Glu	serine	Ser
glutamine	Gln	threonine	Thr
glycine	Gly	tryptophan	Trp
histidine	His	tyrosine	Tyr
isoleucine	Ile	valine	Val

each protein molecule that is synthesized. The second step, elongation, happens once for each amino acid that is needed in the protein.

**Initiation**

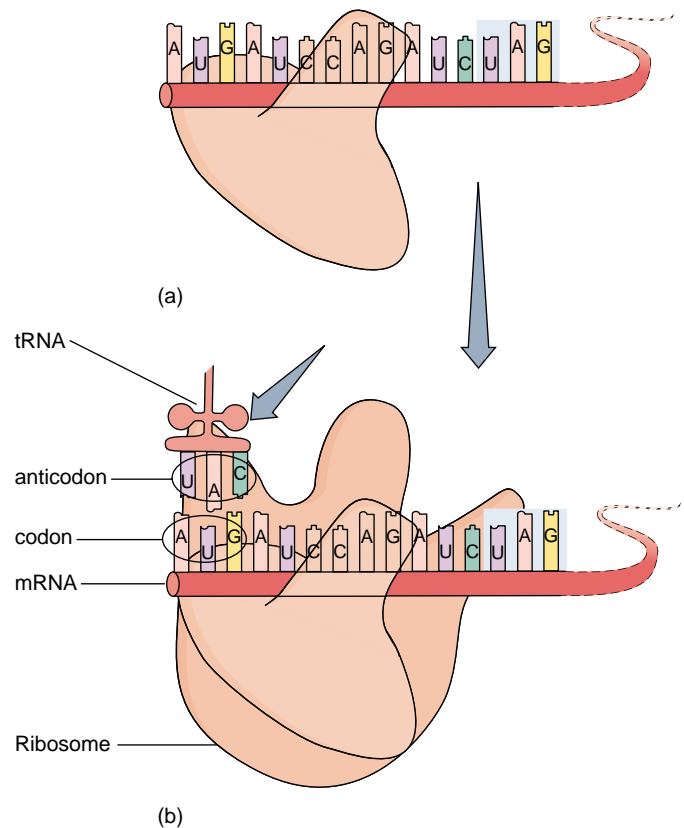
Amino acids are not nucleic acids, so their chemical properties and behaviors are not the same. Amino acids will not “play” by the nucleic acid base-pairing rules. As a result, the cell needs a way to match up amino acids with the three-letter codons in the mRNA. The cell’s machinery has compensated for this by matching up the different amino acids with an adaptor molecule that can base-pair with the nucleotides of the codons. This adaptor molecule is called *tRNA*.

Protein synthesis does not begin with the very first nucleotide of the mRNA. Rather, protein synthesis starts somewhere in the middle of the mRNA molecule. For this reason, the cell needs a signal that says, “Start protein synthesis here!” That signal is the 3-nucleotide sequence, AUG, aptly named the *start*, or **initiator codon** (figure 7.5). The initiator codon always brings in the methionine (Met) amino acid. After the start codon, the nucleotides are read sequentially in sets of 3 to bring in the correct amino acids.

The correct amino acid is correctly paired to each codon because of the *tRNA* molecule to which the amino acid is attached. The *tRNA* has bases that match the codon of the mRNA. The trio of bases in the *tRNA* that is involved in this base-pairing is called the **anticodon**.

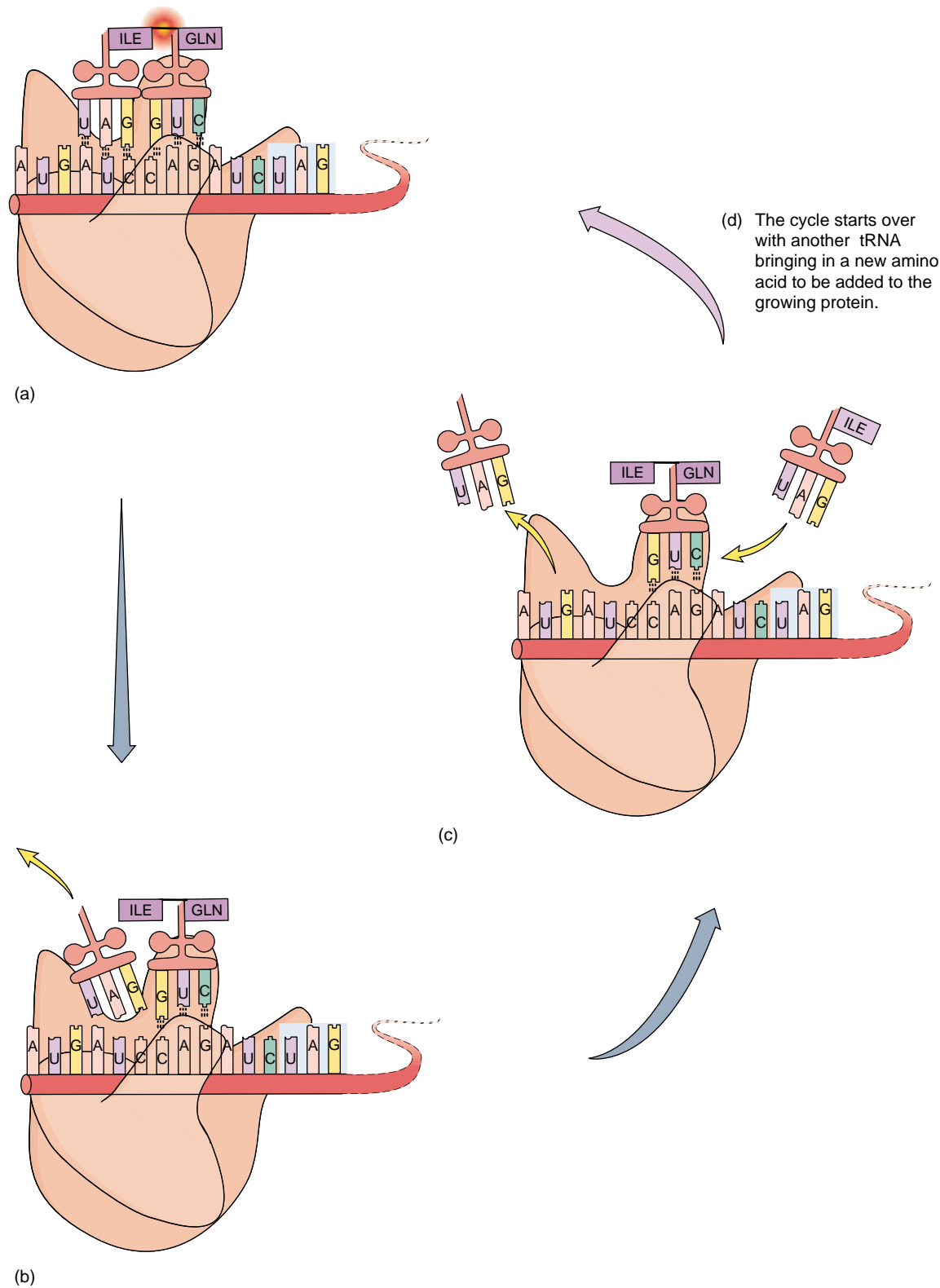
**Elongation**

Once protein synthesis is initiated, the ribosome, mRNA, and *tRNA* undergo a repetitive series of events to bring in each subsequent amino acid of the protein (figure 7.6). The

**Figure 7.5****Initiation**

(a) An mRNA molecule is positioned in the ribosome so that two codons are in position for transcription. The first of these two codons (AUG) is the initiation codon. (b) The *tRNA* lines up with mRNA over the AUG codon using base-pairing rules between the *tRNA* and the mRNA.



**Figure 7.6****Elongation**

(a) The two tRNAs align the amino acids (leu and tyr) so that they can be chemically attached to one another by a peptide bond. (b) Once the bond is formed, (b) the first tRNA detaches from its position on the mRNA. (c) The ribosome moves down one codon on the mRNA. Another tRNA now aligns so that the next amino acid (gly) can be added to the growing protein. (d) The process continues with a new tRNA, a new amino acid, and formation of a new peptide bond.

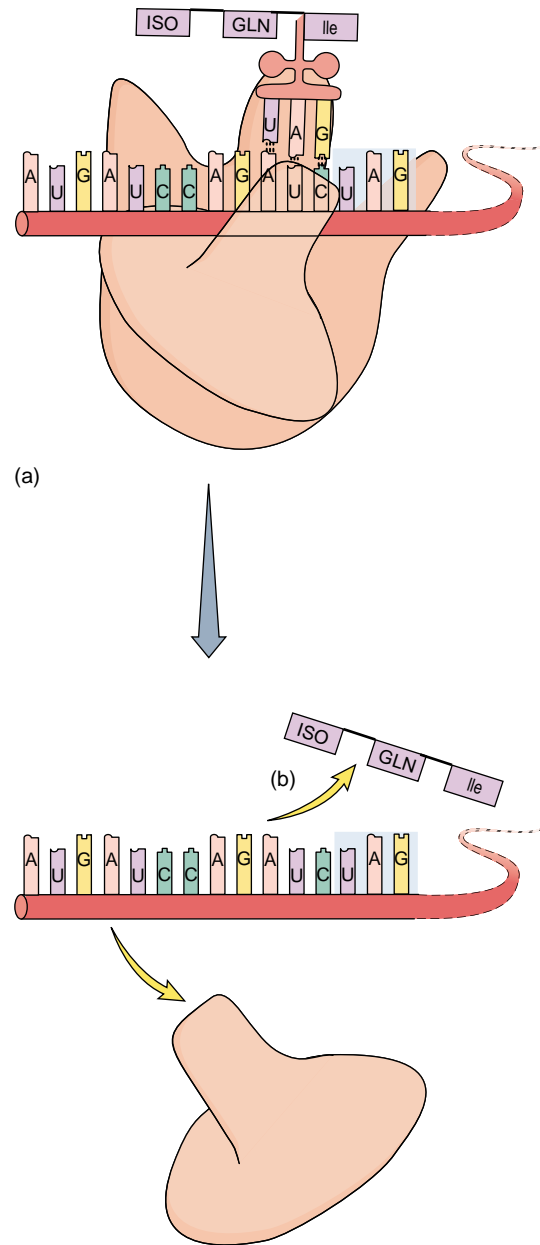
order for the entry of amino acids into the growing protein is dictated by base-pairing rules between the codon of the mRNA and the anticodon of the tRNA. The steps of protein elongation are summarized here and are then explained in a bit more detail:

1. The next tRNA, with its attached amino acid, moves into position and pairs with the codon on the mRNA on the ribosome.
2. The amino acid is chemically bonded to the growing protein.
3. The empty tRNA now moves out of position and back into the cell's cytoplasm.
4. The ribosome shifts down the mRNA to use the information in the next codon.

During elongation, the ribosome actually works with 2 side-by-side codons at a time. At the beginning of the elongation cycle, 1 codon is empty of tRNA and the other codon has the forming protein attached to the last tRNA used to bring in an amino acid. With this combination of components in place, step 1 of elongation occurs. The new tRNA with the correct amino acid moves into position. Once the new tRNA is in position, the ribosome actually holds 2 tRNAs in place for a moment. The older tRNA holds the growing protein and the newer tRNA holds the next amino acid to be added to the growing protein. The ribosome moves the growing protein from the older of the 2 tRNAs to the amino acid on the newest tRNA. After this occurs, the older and empty tRNA moves out of position with regard to the ribosome. The ribosome then shifts down the mRNA one codon so that the beginning position of elongation is available again (see figure 7.6). This cycle continues until protein synthesis is terminated.

### Termination

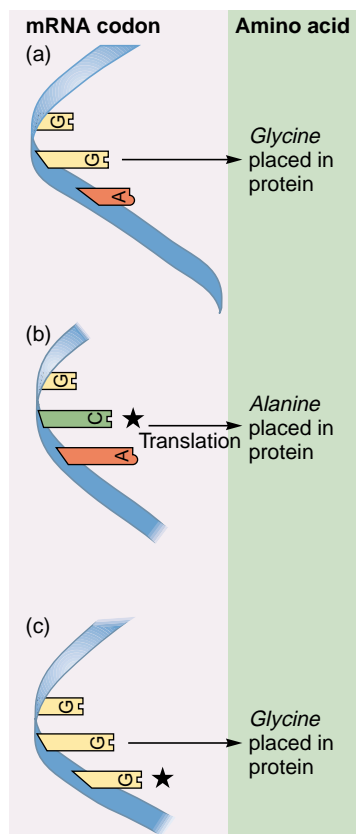
The mRNA coding for a protein is not read to its end to finish protein synthesis. Just as there was initiation codon for starting protein synthesis, there are also **termination codons** that allow protein synthesis to end in the middle of the mRNA molecule. By closely looking at table 7.3, you will see that 3 different codons can stop protein synthesis. When any of the three codons (UAA, UAG, or UGA) appear during the elongation process, a release factor rather than a new amino acid is brought into place on the ribosome (figure 7.7). The forming protein is released because no free amino acid is present. Once the protein is released, the ribosome in turn releases the mRNA. This mRNA can be used to make another protein or can be degraded by the cell's enzymes.



**Figure 7.7**

### Termination

(a) A release factor will move into position over a termination codon UAG. (b) The ribosome releases the completed amino acid chain. The ribosome disassembles and the mRNA can be used by another ribosome to synthesize another protein.



**Figure 7.8**

### DNA Mutations Can Be Silent

A nucleotide substitution changes the protein only if the changed codon results in a different amino acid being substituted into a protein chain. This feature of DNA serves to better ensure that the synthesized protein will be functional. (a) In the example provided, the original codon, GGA, calls for Gly. (b) An observable mutation changes the nucleotide in the second position of the codon, which now reads GCA. GCA calls for Ala. (c) A *silent* mutation is shown where the third position of the codon is changed. The codon GGG calls for the same amino acid as the original version (GGA). Because the proteins produced in the first and third examples will be identical in amino acid sequence, they will function the same also.

## 7.3 Mutations

Any change in the nucleotide sequence of DNA—the genetic information—is a mutation. Some mutations result in changes in the proteins produced and some do not. As stated earlier, as a genetic material, DNA must be able to change, or mutate. Different types of changes can occur and these changes have a potential impact on protein synthesis.

Normal Sequence	THE ONE BIG FLY HAD ONE RED EYE
Kind of Mutation	Sequence Change
Missense	THQ ONE BIG FLY HAD ONE RED EYE
Nonsense	THE ONE BIG
Frameshift	
Due to insertion	THE ONE QBI GFL YHA DON ERE DEY E
Due to deletion	TEO NEB IGF LYH ADO NER EDE YE
Deletion	THE ONE BIG HAD ONE RED EYE
Duplication	THE ONE BIG FLY FLY HAD ONE RED EYE
Insertion	THE ONE BIG WET FLY HAD ONE RED EYE
Expanding mutation:	
Parents	THE ONE BIG FLY HAD ONE RED EYE
Children	THE ONE BIG FLY FLY FLY HAD ONE RED EYE
Grandchildren	THE ONE BIG FLY FLY FLY FLY FLY FLY HAD ONE RED EYE

**Figure 7.9**

### Classes of Mutations

A sentence comprised of three-letter words can provide an analogy to the effect of mutations on a gene's nucleotide sequence. Examine each type of mutation to see how the information in the sentence has been confused by the change.

DNA strikes a good balance between stability and flexibility. Every change in DNA does not cause a change in a protein. For example, the codon GGX codes for the amino acid, Gly, regardless of the nucleotide found in the X position (table 7.3). This is illustrated in figure 7.8 with another example. This means that a large number of changes in DNA sequence may not result in changes in proteins; these are called *silent mutations*. However, not all such changes can be compensated for by the flexibility codon system, and an altered protein may be produced.

**Mutagenic agents** are substances or conditions that can cause the sequence of DNA to change. Agents known to cause damage to DNA are certain viruses (e.g., papillomavirus), weak or “fragile” spots in the DNA, radiation (X rays or UV light), and chemicals found in foods and other products such as nicotine in tobacco. All have been studied extensively and there is little doubt that they cause mutations.

What types of DNA mutations can be the most devastating to protein synthesis? Consider that any change in the start codon will prevent the protein from being correctly synthesized. Also consider that inserting or deleting a single nucleotide in the protein coding sequence will cause the ribosome to read the wrong set of 3 nucleotides as the

codon. This type of mutation is called a *frameshift mutation*. Figure 7.9 provides examples of the types of mutations that are commonly observed in DNA sequences.

To identify the impact that a mutation can have on a protein and, ultimately, the whole organism, consider sickle-cell anemia. In some individuals, a single nucleotide of the gene may be changed. This type of mutation is called a **point mutation**. An example of the effects of altered DNA can be seen in human red blood cells. Red blood cells contain the oxygen-carrier molecule, hemoglobin. Normal hemoglobin molecules are composed of approximately 145 amino acids in four chains—two alpha and two beta. The nucleotide sequence of the gene for the beta chain is known, as is the amino acid sequence for this chain. In normal individuals, the amino acid sequence of the hemoglobin protein begins like this:

Val-His-Leu-Thr-Pro-Glu-Glu-Lys . . .

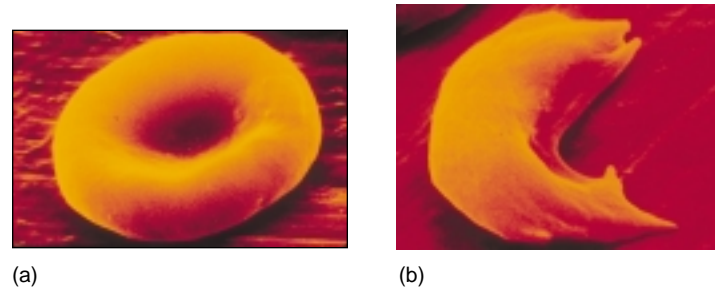
As a result of a single nucleotide in the DNA sequence, the resulting mutation is a new amino acid sequence in all the red blood cells:

Val-His-Leu-Thr-Pro-Val-Glu-Lys . . .

This single nucleotide change, which causes a single amino acid to change, may seem minor. However, it is the cause of sickle-cell anemia, a disease that affects the red blood cells by changing them from a circular to a sickle shape when oxygen levels are low (figure 7.10). When this sickling occurs, the red blood cells do not flow smoothly through capillaries. Their irregular shapes cause them to clump, clogging the blood vessels. This prevents them from delivering their oxygen load to the oxygen-demanding tissues. A number of physical disabilities may result, including weakness, brain damage, pain and stiffness of the joints, kidney damage, rheumatism, and, in severe cases, death.

Changes in the structure of DNA may have harmful effects on the next generation if they occur in the sex cells. Sex cells like sperm and eggs transmit genetic information from one generation in a species to the next. Mutations that occur to DNA molecules can only be passed on to the next generation when the mutation is present in cells like sperm and eggs that are used to start the next generation.

Some damage to DNA is so extensive that the entire strand of DNA is broken, resulting in the synthesis of abnormal proteins or a total lack of protein synthesis. A number of experiments indicate that many street drugs such as LSD (lysergic acid diethylamide) are mutagenic agents that cause DNA to break.



**Figure 7.10**

#### Normal and Sickled Red Blood Cells

(a) A normal red blood cell is shown in comparison with (b) a cell having the sickle shape. This sickling is the result of a single amino acid change in the hemoglobin molecule.

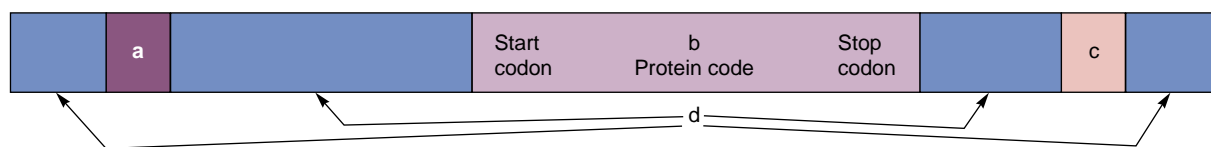
## 7.4 Controlling Gene Expression

With a solid understanding of transcription and translation, we can begin to look at the implications of that system of information exchange within the context of the cell. At a higher level of organization, you can also look at the central dogma in the context of an entire multicellular organism. One of the first things to consider is that no cell is able to survive with the ability to make just one protein. This means that even the simplest of cells have a large bank of protein blueprints to draw from. This bank of blueprints, the **genome**, is similar to a cook having many recipes to draw from in a single cookbook. To function correctly, cells need to be able to access and use their genetic information in an efficient and successful manner.

### Starting Transcription

Just like many recipes can exist on a single page of a cookbook, the blueprints in a genome are not always on separate molecules of DNA. A single molecule of DNA may have many thousands of protein-coding regions. Because there are so many genes in a cell and the codes for many different proteins can be found on a single molecule of DNA, the cell is faced with a number of problems:

1. How does the cell know where to start making the mRNA?
2. How does the cell know which strand of the double-stranded DNA molecule is used for a template of transcription?



**Figure 7.11**

### Gene Structure

There are four general regions of a gene. (a) The promoter region attracts the RNA polymerase to the site of the gene and then directs the RNA polymerase to proceed along the DNA strand in the correct direction to find the gene. (b) The protein coding region contains the information needed to correctly create a protein—this includes the start codon, the stop codon, and all the codon that are located between them. (c) The termination sequence signals the RNA polymerase to end mRNA transcription so that the mRNA can leave the nucleus and enter the cytoplasm for use in the next step of the central dogma, translation. (d) The regulatory region can be found all around the promoter region, the protein-coding region, and the termination sequence. The regulatory regions attract and bind proteins that will help or slow down the process of transcription.

### 3. How does the cell know when to stop making the mRNA?

To understand the answers to these questions, it is necessary to know that there is more information stored in a gene than the portion that codes for a protein. Genes contain four general types of information that is organized into four regions: (1) The **promoter region**, (2) the **protein-coding region**, (3) the **termination sequence**, and (4) the **regulatory regions** (figure 7.11). We have already discussed the protein-coding region, how its information is organized into codons, and that it contains very specific stop and start signals.

The promoter region contains the answer to the first two questions. The promoter region of a gene is a specific sequence of DNA that is located a number of nucleotides upstream of the region of DNA that codes for the protein. The unique sequence of the promoter is recognized by the enzyme responsible for transcription, RNA polymerase. The promoter of a gene acts in a manner similar to runway lights for an airplane. The promoter provides a signal that indicates, “Here’s a gene and it is this way!” Once the RNA polymerase has “landed” onto the promoter and is moving in the right direction toward the gene, only one of the strands of DNA can be used as a template for making the mRNA. This is due to the chemical details of how nucleotides are added to a growing strand of RNA and how two strands of nucleic acids orient to each other.

RNA polymerase finishes making an RNA molecule when it encounters a termination sequence—it releases the mRNA that is being synthesized and the DNA template. The termination sequence is a sequence of DNA that is located a number of nucleotides after the protein-coding region. The mRNA that has been created for translation actually spans the area shown in figure 7.11 that starts between the promoter and the protein-coding region and ends at the termi-

nation sequence. This mRNA is then used in translation to direct protein synthesis.

The general scheme presented here is that genetic information flows from DNA to RNA and is used to make protein. In the case of some viruses, genetic information is stored in RNA and is used to make DNA. In this instance, the flow of genetic information is RNA to DNA then to mRNA and protein (Outlooks 7.1).

### The Right Amount at the Right Time

This basic understanding of how a gene is organized with a promoter, a protein-coding region, and a termination sequence is vital to understanding how cells are able to control **gene expression**. A gene is expressed when its product, usually a protein, is produced for use by the cell. Control of gene expression involves making sure that a gene is used at the appropriate time and that its protein is being made in adequate amounts.

Regulatory regions are important in determining when and how much a gene expresses itself. The regulatory regions surround the other three regions of the gene. They work by binding specific proteins that are made by the cell. These proteins are called **transcription factors**. These proteins are unique in two ways: (1) They are only present in the cell under certain conditions, such as when a gene’s protein must be produced; and (2) they have the effect of making the promoter region more noticeable or more invisible to the RNA polymerase.

### Splicing

One of the most significant differences between prokaryotic and eukaryotic transcription is that the protein-coding region of the prokaryotic DNA is continuous, while in eukaryotic cells, it is not (figure 7.12). A eukaryotic gene begins



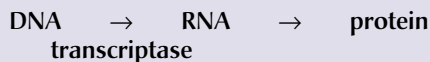


## OUTLOOKS 7.1

### HIV Infection (AIDS) and Reverse Transcriptase

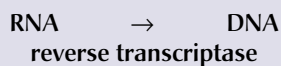
**A**IDS is an acronym for acquired immunodeficiency syndrome and is caused by a retrovirus called *human immunodeficiency virus*. HIV is a spherical virus with an outer membrane, an inside protein coat, and an RNA core. Its genetic material is RNA, not DNA. Genes are carried from one generation to the next as RNA molecules. This is not the case in humans and most other organisms where DNA is the genetic material.

Human sequence:

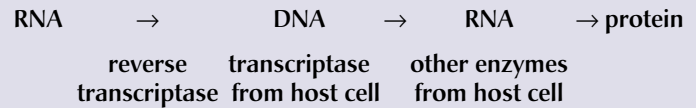


However, once having entered a suitable, susceptible host cell, HIV must convert its RNA to a DNA genome in order to integrate into the host cell's chromosome. Only then can it become an active, disease-causing parasite. This conversion of RNA to DNA is contrary, reverse, or retro to the RNA-forming process controlled by the enzyme transcriptase. Humans do not have the genetic capability to manufacture the enzyme necessary to convert RNA to DNA—reverse transcriptase.

Retrovirus sequence:



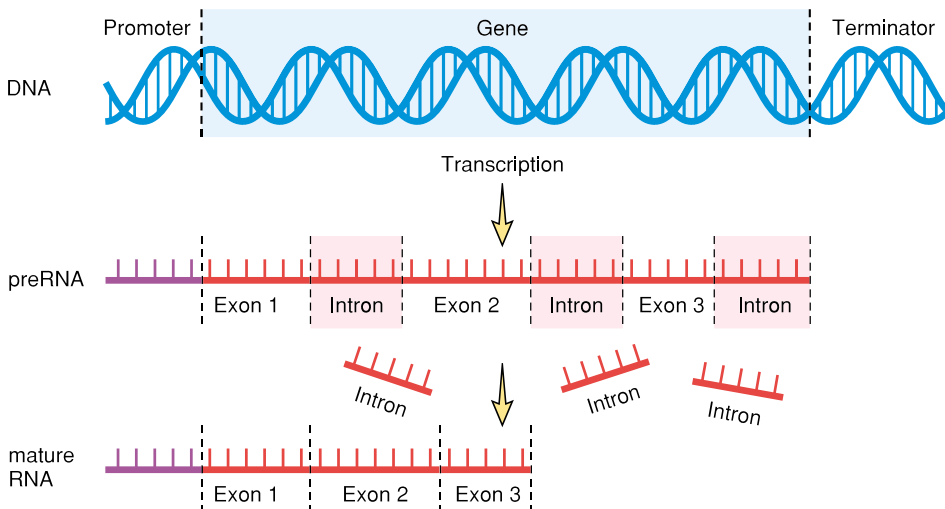
When HIV carries out gene replication and protein synthesis the process is diagrammed:



This has two important implications. First, the presence of reverse transcriptase in a human can be looked upon as an indication of retroviral infection because it is not manufactured by human cells. However, because HIV is only one of several types of retroviruses, the presence of the enzyme in an individual does not necessarily indicate an HIV infection. It does indicate “some type” of retroviral infection. Second, interference with reverse transcriptase will frustrate the virus's attempt to integrate into the host chromosome.

Treatments for HIV take advantage of a vulnerable point in the retrovirus's life cycle. The drugs fall into three basic categories: (1) blockers, (2) reverse transcriptase inhibitors, and (3) protease inhibitors. Drugs classified as blockers prevent the virus from entering the cell. Reverse transcriptase inhibitors interfere with the crucial step of converting the virus's RNA genetic information from RNA to DNA. Protease inhibitors interfere with the modification of viral proteins after translation. This prevents a complete virus from completely forming and infecting another cell. A treatment regime can include a drug from each of these categories. This combination of drugs is called a *cocktail* and is much more effective against HIV than taking any one of the drugs alone.

However, HIV mutates very quickly because its genetic information is stored as a single-stranded molecule of RNA. This high mutation rate quickly produces strains of the HIV virus that are resistant to these drugs. • evolution, p. 222



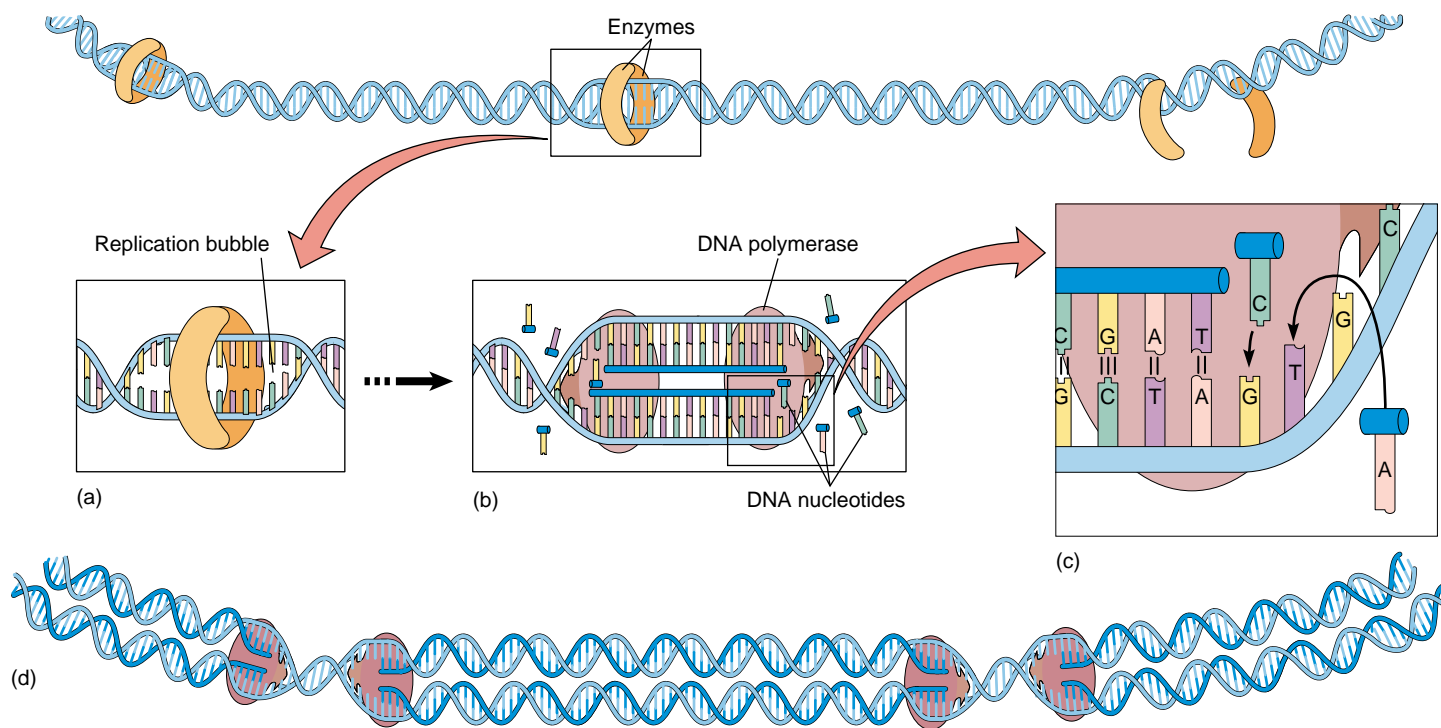
**Figure 7.12**

#### Transcription of mRNA in Eukaryotic Cells

This is a summary of the events that occur in the nucleus during the manufacture of mRNA in a eukaryotic cell. Notice that the original nucleotide sequence is first transcribed into an RNA molecule that is later “clipped” and then rebonded to form a shorter version of the original. It is during this time that the introns are removed.

with a promoter region and an initiation code and ends with a termination code and region. However, the intervening gene sequence contains patches of nucleotides that apparently do not code for protein but do serve important roles in maintaining the cell. If they were used in protein synthesis, the resulting proteins would be worthless. To remedy this

problem, eukaryotic cells prune these segments from the mRNA after transcription. When such split genes are transcribed, RNA polymerase synthesizes a strand of pre-mRNA that initially includes copies of both exons (meaningful mRNA coding sequences) and introns (intervening DNA sequences). Soon after its manufacture, this pre-mRNA mole-



**Figure 7.13**

### DNA Replication

These illustrations summarize the basic events that occur during DNA replication. (a) Enzymes break apart the two strands of DNA. (b and c) As the DNA strands are separated, new DNA nucleotides are added to the new strands by DNA polymerase. The new DNA strands are synthesized according to base-pairing rules for nucleic acids. (d) By working in two directions at once along the DNA strand, the cell is able to more quickly replicate the DNA. Each new daughter cell will receive one of these copies.

cule has the meaningless introns clipped out and the exons spliced together into the final version, or mature mRNA, which is used by the cell. In humans, it has been found that the exons of a single gene may be spliced together in three different ways, resulting in the production of three different, mature messenger RNAs. This means that a single gene can be responsible for the production of three different proteins. Learning this information has led geneticists to revise their estimate of the total number of genes found in the human genome from 100,000 to an estimated 30,000.

## 7.5 Nucleotide Sequences— The Cell's Legacy

We have looked at how cells use genetic information in DNA to make proteins, but cells use this information in a second, *very different* manner: They pass it on to their offspring. The process of passing DNA on to the next generation of cells helps to ensure that the offspring cells will have the genetic information necessary to carry out the needed chemical reactions of life.

### DNA Replication

**DNA replication** occurs in cells in preparation for cell division. Without replication, daughter cells would not receive the library of information required to sustain life. Because all cells must maintain a complete set of genetic material, there must be a doubling of DNA to produce enough copies to pass on to the offspring. DNA replication is the process of duplicating the genetic material prior to its distribution to daughter cells. When a cell divides into two *daughter cells*, each new cell must receive a complete and exact copy of the parent cell's genetic information or it will not be able to manufacture all the proteins vital to its existence.

DNA replication involves a few general steps:

1. The DNA replication process begins as an enzyme breaks the attachments between the paired bases of the two strands of DNA. In eukaryotic cells, this occurs in hundreds of different positions forming replication bubbles along the length of the DNA. In prokaryotic cells, this occurs in only one spot, the origin of replication (figure 7.13a).
2. As the two DNA strands are separated in both directions along the DNA molecule, new DNA

nucleotides bond to the nucleotides on the existing DNA strands using the base-pairing rules and an enzyme called DNA polymerase (figure 7.13*b*).

3. DNA polymerase sequentially attaches each incoming nucleotide to the previous one by chemical bonds between the phosphate of one nucleotide and the deoxyribose of the next (figure 7.13*c*).
4. Eventually all the replication bubbles combine and a new strand of DNA is formed on each of the original strands (figure 7.13*d*). The result is two identical double-stranded DNA molecules.

Note that the overall pattern is that the new complementary strand of DNA forms on each of the old DNA strands, resulting in the formation of two double-stranded, duplex DNA molecules. In this way, the exposed nitrogenous bases of the original DNA serve as a template, or pattern, for the formation of the new DNA. As the new DNA is completed, it twists into its double-helix shape.

The completion of the DNA replication process yields two double helices that are identical in their nucleotide sequences. Half of each is new; half is the original parent DNA molecule. The DNA replication process is highly accurate. It has been estimated that there is only one error made for every  $2 \times 10^9$  nucleotides. A human cell contains 46 chromosomes consisting of about 3 billion base pairs. This averages to about five errors per cell! Don't forget that this figure is an estimate. Whereas some cells may have five errors per replication, others may have more, and some may have no errors at all. It is also important to note that some errors can be major and deadly, whereas others are insignificant. Because this error rate is so small, DNA replication is considered to be essentially error-free. Following DNA replication, the cell now contains two complete copies of genetic information and is ready to begin the process of distributing one set of genetic information to each of its two daughter cells.

The distribution of DNA involves splitting the cell and distributing a set of genetic information to the two new daughter cells. In this way, each new cell has the necessary information to control its activities. The mother cell ceases to exist when it divides. Its contents are divided between the two smaller daughter cells.

A cell does not really die when it reproduces itself; it merely starts over again. This is called the *life cycle* of a cell. A cell can divide and redistribute its genetic information to the next generation in a number of ways. • **mitosis**, p. 159, **meiosis**, p. 170.

## DNA and Chromosomes

After DNA is replicated, cells wind the DNA up like thread on a bobbin or spool so that the huge bulk of DNA is more manageable. Eukaryotic and prokaryotic cells do this in different ways.

The genetic material of humans and other eukaryotic organisms consists of strands of coiled, double-stranded DNA,

## OUTLOOKS 7.2



### Telomeres

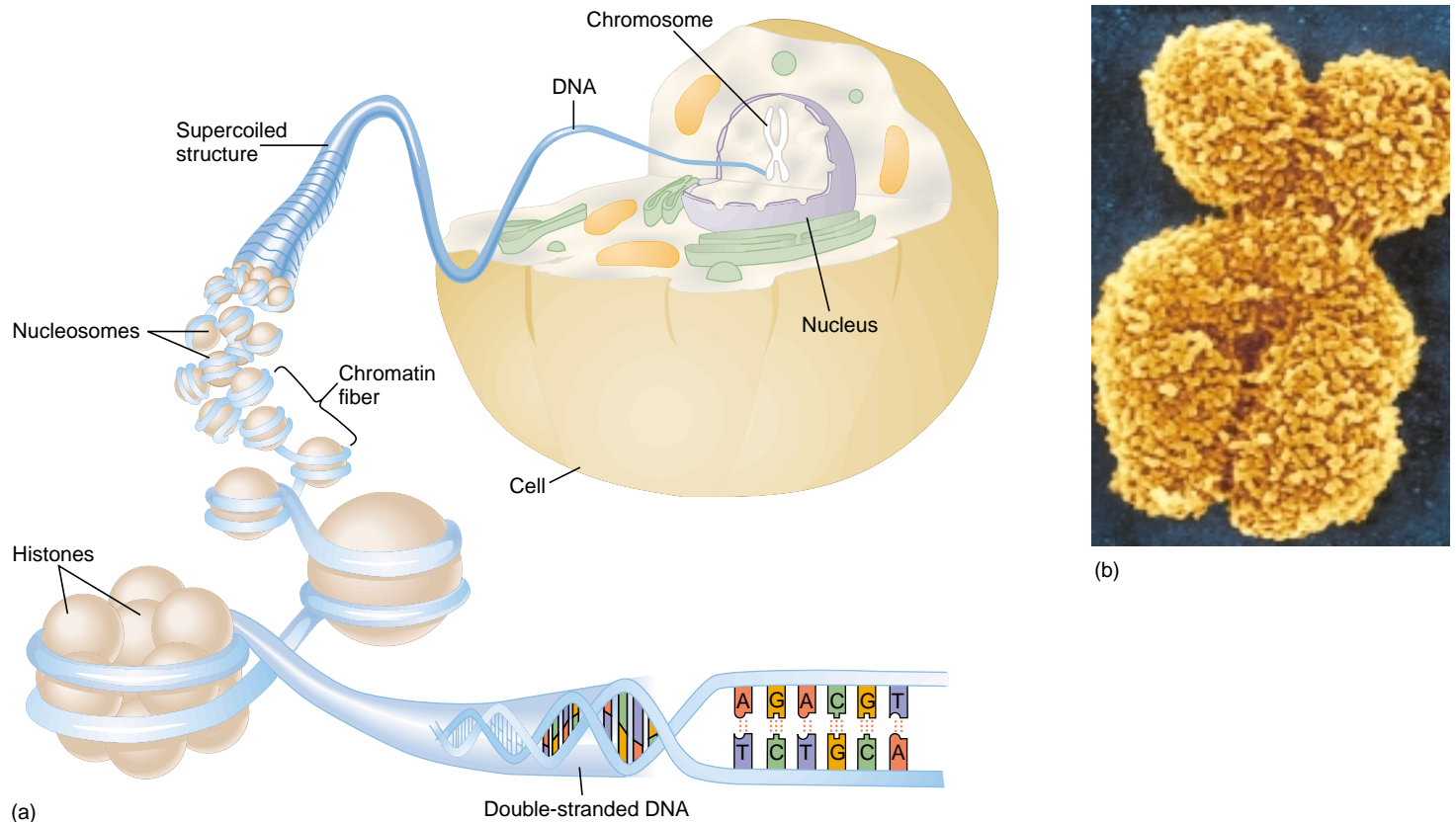
The ends of a chromosome contain a special sequence of nucleotides called a **telomere**. In humans, these chromosome “caps” contain the nucleotide base-pair sequence:

TTAGGG  
AATCCC

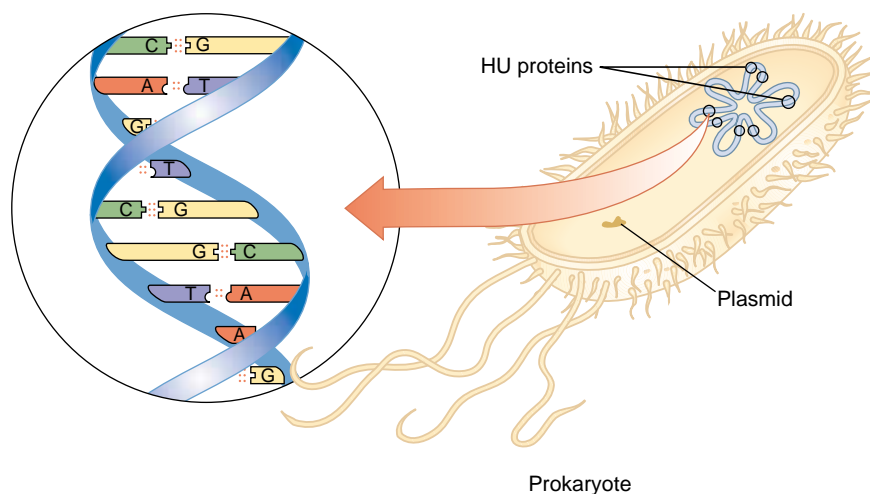
This sequence is repeated many times over. Telomeres are very important segments of the chromosome. They are required for chromosome replication; they protect the chromosome from being destroyed by dangerous DNAase enzymes (enzymes that destroy DNA), and they keep chromosomes from bonding to one another end to end. Evidence shows that the loss of telomeres is associated with cell “aging,” whereas not removing them has been linked to cancer. Every time a cell reproduces itself, it loses some of its telomeres. However, in cells that have the enzyme telomerase, new telomeres are added to the end of the chromosome each time the cells divide. Therefore, cells that have telomerase do not age as other cells do, and cancer cells are immortal as a result of this enzyme. This enables them to maintain, if not increase, the number of telomeres from one cell generation to the next.

which has histone proteins attached along its length. These coiled DNA strands with attached proteins become visible during cell division and are called **nucleoproteins** or **chromatin fibers**. The histone protein and DNA are not arranged randomly, but come together in a highly organized pattern. The double-stranded DNA spirals around repeating clusters of eight histone spheres. Histone clusters with their encircling DNA are called **nucleosomes** (figure 7.14*a*). When eukaryotic chromatin fibers coil into condensed, highly knotted bodies, they are seen easily through a microscope after staining them with dye. Condensed like this, a chromatin fiber is referred to as a **chromosome** (figure 7.14*b*; Outlooks 7.2).

The genetic material in bacteria is also double-stranded DNA, but the ends of the molecule are connected to form a loop and they do not form condensed chromosomes (figure 7.15). However, prokaryotic cells have an attached protein called *HU protein*. In certain bacteria, there is an additional loop of DNA called a *plasmid*. Plasmids are considered extra DNA because they appear not to contain genes that are required for the normal function of the cell. However, they can play two important roles in bacteria: (1) Some plasmids have genes that enable the cell to resist certain antibiotics such as the penicillins. The gene may be for the production of the enzyme beta lactamase (formerly known as penicillinase), which is capable of destroying certain forms of penicillin. It is important to note that plasmids are capable of carrying many kinds of genes. We tend to

**Figure 7.14****Eukaryotic Genome Packaging**

(a) Eukaryotic cells contain DNA in their nuclei that takes the form of a double-stranded helix. The two strands fit together and are bonded by weak hydrogen bonds formed between the complementary, protruding nitrogenous bases according to the base-pairing rule. To form a chromosome, the DNA molecule is wrapped around a group of several histone proteins. Together the histones and the DNA form a structure called the nucleosome. The nucleosomes are packaged to form a chromosome. (b) During certain stages in the reproduction of a eukaryotic cell, the nucleoprotein coils and “supercoils,” forming tightly bound masses. When stained, these are easily seen through the microscope. In their supercoiled form, they are called *chromosomes*, meaning “colored bodies.”

**Figure 7.15****Prokaryotic Genome Packaging**

The nucleic acid of prokaryotic cells does not have the histone protein; rather, it has proteins called *HU* proteins that serve a similar function to histones. In addition, the ends of the giant nucleoprotein molecule overlap and bind with one another to form a loop. The additional small loop of DNA is a plasmid, which contains genes that are not essential for the daily life of the cell.



be most concerned about plasmids that carry antibiotic resistance because of the medical implications.

A second important role of plasmids enables the cell to become involved in *genetic recombination*—the transfer of genes from one cell (the donor) to another (the recipient). By transferring genes from one cell to another, cells that receive the genes can become genetically diverse and are more likely to survive threatening environmental hazards.

Each chromatin strand is different because each strand has a different chemical code. Coded DNA serves as a central cell library. Tens of thousands of messages are in this storehouse of information. This information tells the cell such things as (1) how to produce enzymes required for the digestion of nutrients; (2) how to manufacture enzymes that will metabolize the nutrients and eliminate harmful wastes; (3) how to repair and assemble cell parts; (4) how to reproduce healthy offspring; (5) when and how to react to favorable and unfavorable changes in the environment; and (6) how to coordinate and regulate all of life's essential functions. If any of these functions are not performed properly, the cell will die. The importance of maintaining essential DNA in a cell becomes clear when we consider cells that have lost it. For example, human red blood cells lose their nuclei as they become specialized to carry oxygen and carbon dioxide throughout the body. Without DNA, they are unable to manufacture the essential cell components needed to sustain themselves. They continue to exist for about 120 days, functioning only on enzymes manufactured earlier in their lives. When these enzymes are gone, the cells die. Because these specialized cells begin to die the moment they lose their DNA, they are more accurately called *red blood corpuscles* (RBCs): “little dying red bodies.”

### Chromosomal Aberrations

A **chromosomal aberration** is a major change in DNA that can be observed at the level of the chromosome. There are four types of aberrations: inversions, translocations, duplications, and deletions. An *inversion* occurs when a chromosome is broken and a piece becomes reattached to its original chromosome but in reverse order—it has been cut out and flipped around. A *translocation* occurs when one broken segment of DNA becomes integrated into a different chromosome. *Duplications* occur when a portion of a chromosome is replicated and attached to the original section in sequence. *Deletion* aberrations result when a broken piece becomes lost or is destroyed before it can be reattached.

---

## 7.6 Using DNA to Our Advantage

A powerful new science of gene manipulation—**biotechnology**—suggests genetic diseases can be controlled or cured in the future. Since 1953, when the structure of the DNA molecule was first described, there has been a rapid succession of advances in the field of genetics. It is now possible to transfer DNA from one organism to another. This has

made possible the manufacture of human genes and gene products by bacteria.

### Genetic Engineering and Biotechnology

Biotechnology includes the use of a method of splicing genes from one organism into another, resulting in a new form of DNA called **recombinant DNA**. Organisms with these genetic changes are referred to as **genetically modified (GM)**, or **transgenic organisms**. These organisms or their offspring have been engineered so that they contain genes from at least one unrelated organism, which could be a virus, a bacterium, a fungus, a plant, or an animal.

As this highly sophisticated procedure has been refined, it has become possible to quickly and accurately splice genes from a variety of species into host bacteria, making possible the synthesis of large quantities of medically important products. For example, recombinant DNA procedures are responsible for the production of human insulin, used in the control of diabetes; interferon, used as an antiviral agent; human growth hormone, used to stimulate growth in children lacking this hormone; and somatostatin, a brain hormone also implicated in growth. Over 200 such products have been manufactured using these methods.

Revolutionary possibilities open up with the manipulation of DNA. These methods enable cells to produce molecules that they would not normally make. Some research laboratories have even spliced genes into laboratory-cultured human cells. Should such a venture prove to be practical, genetic diseases such as sickle-cell anemia could be controlled. The process of recombinant DNA gene splicing also enables cells to be more efficient at producing molecules that they normally synthesize.

Some of the likely rewards are (1) production of additional, medically useful proteins; (2) mapping of the locations of genes on human chromosomes; (3) a more complete understanding of how genes are regulated; (4) production of crop plants with increased yields; and (5) development of new species of garden plants (figure 7.16).

The discovery of the structure of DNA nearly 50 years ago seemed very far removed from the practical world. The importance of this “pure” or “basic” research is just now being realized. Many companies are involved in recombinant DNA research with the aim of alleviating or curing disease.

The field of **bioengineering** is advancing as quickly as is the electronics industry. The first bioengineering efforts focused on developing genetically modified (GM) crops that had improvements over past varieties, such as increased resistance to infectious plant disease. This was primarily accomplished through selective breeding and irradiation of cells to produce desirable mutations. The second wave of research involved directly manipulating DNA using the more sophisticated techniques such as the polymerase chain reaction (PCR), genetic fingerprinting, and cloning. Genetic engineers identify and isolate sequences of nucleotides from a living or dead cell and install it into another living cell. This





**Figure 7.16**

Using our understanding of how DNA codes for proteins allows scientists to develop treatments for disease as well as develop crops with greater yield and nutritional value.

has resulted in improved food handling and processing, such as slower ripening in tomatoes. Currently, crops are being genetically manipulated to manufacture large quantities of specialty chemicals such as antibiotics, steroids, and other biologically useful, organic chemicals.

Although some of these chemicals have been produced in small amounts from genetically engineered microorganisms, crops such as turnips, rice, soybeans, potatoes, cotton, corn, and tobacco can generate tens or hundreds of kilograms of specialty chemicals per year. Many of these GM crops also have increased nutritional value and yet can be cultivated using traditional methods. Such crops have the potential of supplying the essential amino acids, fatty acids, and other nutrients now lacking in the diets of people in underdeveloped or developing nations. Researchers have also shown, for example, that turnips can produce interferon (an antiviral agent), tobacco can create antibodies to fight

human disease, oilseed rape plants can serve as a source of human brain hormones, and potatoes can synthesize human serum albumin that is indistinguishable from the genuine human blood protein.

*Gene cloning* is accomplished by using enzymes that are naturally involved in the DNA replication process and other enzymes that are naturally produced by bacteria. When genes are spliced from different organisms into host cells, the host cell replicates these new “foreign” genes and synthesizes proteins encoded by them. Gene splicing begins with the laboratory isolation of DNA from an organism that contains the desired gene—for example, from human cells that contain the gene for the manufacture of insulin. If the gene is short enough and its base sequence is known, it can be synthesized in the laboratory from separate nucleotides. If the gene is too long and complex, it is cut from the chromosome with enzymes called *restriction endonucleases*, so named because

these enzymes (*-ases*) only cut DNA (*nucle-*) at certain base sequences (restricted in their action) and work from the middle or inside of (*endo-*) the DNA. In contrast, a restriction exonuclease digests DNA from the outside (*exo-*) or from the end of the DNA molecule. *Endo* nucleases act like molecular scissors that do not cut the DNA straight across, but in a zig-zag pattern that leaves one strand slightly longer than its complement. The short nucleotide sequence that sticks out and remains unpaired is called a *sticky end* because it can be reattached to another complementary strand.

This isolated gene with its “sticky end” is most frequently spliced into microbial DNA. The host DNA is opened up with the proper restriction endonuclease and ligase (i.e., tied together) enzymes that are used to attach the sticky ends into the host DNA. This gene-splicing procedure can be performed with small loops of bacterial DNA that are not part of the main chromosome. These small DNA loops are called *plasmids*.

Once the splicing is completed, the plasmids can be inserted into the bacterial host cell by treating the cell with special chemicals that encourage it to take in these large chunks of DNA. A more efficient alternative is to splice the desired gene into the DNA of a bacterial virus so that it can carry the new gene into the bacterium as it infects the host cell. Once inside the host cell, the genes can be replicated, along with the rest of the DNA, to clone the “foreign” gene, or they can begin to synthesize the encoded protein.

With another genetic engineering accomplishment, *genetic fingerprinting*, it is possible to show the nucleotide sequence differences between individuals since no two people have the same nucleotide sequences. While this sounds like an easy task, the presence of many millions of base pairs in a person’s chromosomes makes this a lengthy, and sometimes impractical, process. Therefore, scientists don’t really do a complete fingerprint but focus only on certain shorter, repeating patterns in the DNA. By focusing on these shorter, repeating nucleotide sequences, it is possible to determine whether samples from two individuals have these same repeating segments. Scientists use a small number of sequences that are known to vary a great deal among individuals, and compare those to get a certain probability of a match. The more similar the sequences, the more likely the two samples are from the same person. The less similar the sequences, the less likely the two samples are from the same person. In criminal cases, DNA samples from the crime site can be compared to those taken from suspects. If 100% of the short, repeating sequence matches, it is highly probable that the suspect was at the scene of the crime and may be the guilty party. This same procedure can also be used to confirm the identity of a person, as in cases of amnesia, murder, or accidental death (figure 7.17).

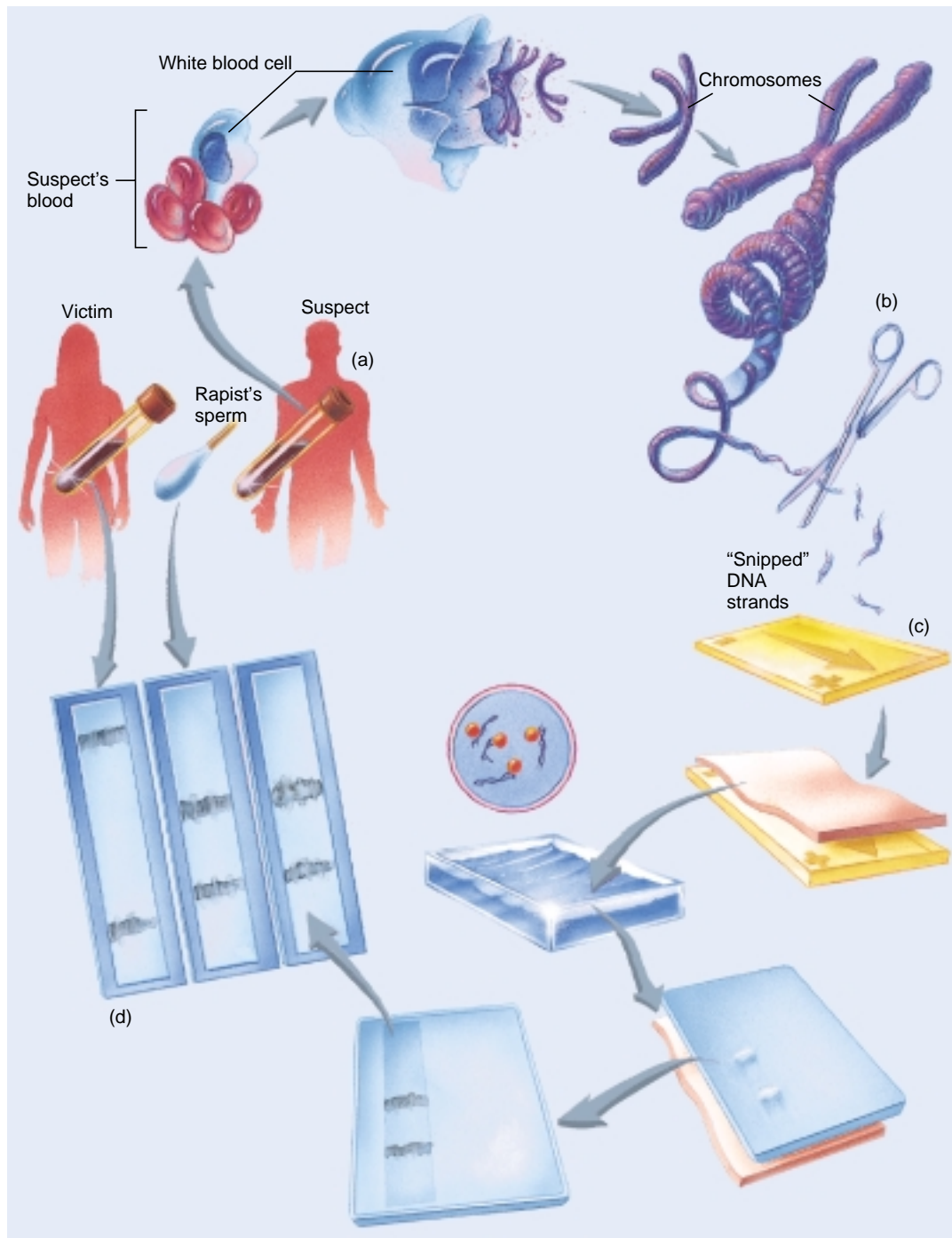
In 1989, the American Association for the Advancement of Science named DNA polymerase “Molecule of the Year.” The value of this enzyme in the *polymerase chain reaction* (PCR) is so great that it could not be ignored. Just what is the PCR, how does it work, and what can you do with it?

PCR is a laboratory procedure for copying selected segments of DNA. A single cell can provide enough DNA for analysis and identification! Having a large number of copies of a “target sequence” of nucleotides enables biochemists to more easily work with DNA. This is like increasing the one “needle in the haystack” to such large numbers (100 billion in only a matter of hours) that they’re not hard to find, recognize, and work with. The types of specimens that can be used include semen, hair, blood, bacteria, protozoa, viruses, mummified tissue, and frozen cells. The process requires the DNA specimen, free DNA nucleotides, synthetic “primer” DNA, DNA polymerase, and simple lab equipment, such as a test tube and a source of heat.

Having decided which target sequence of nucleotides (which “needle”) is to be replicated, scientists heat the specimen of DNA to separate the coding and non-coding strands. Molecules of synthetic “primer” DNA are added to the specimen. These primer molecules are specifically designed to attach to the ends of the target sequence. Next, a mixture of triphosphorylated nucleotides is added so that they can become the newly replicated DNA. The presence of the primer, attached to the DNA and added nucleotides, serves as the substrate for the **DNA polymerase**. Once added, the polymerase begins making its way down the length of the DNA from one attached primer end to the other. The enzyme bonds the new DNA nucleotides to the strand, replicating the molecule as it goes. It stops when it reaches the other end, having produced a new copy of the target sequence. Because the DNA polymerase will continue to operate as long as enzymes and substrates are available, the process continues, and in a short time there are billions of small pieces of DNA, all replicas of the target sequence.

So what, you say? Well, consider the following. Using the PCR, scientists have been able to:

1. More accurately diagnose such diseases as sickle-cell anemia, cancer, Lyme disease, AIDS, and Legionnaires’ disease;
2. Perform highly accurate tissue typing for matching organ-transplant donors and recipients;
3. Help resolve criminal cases of rape, murder, assault, and robbery by matching suspect DNA to that found at the crime scene;
4. Detect specific bacteria in environmental samples;
5. Monitor the spread of genetically engineered microorganisms in the environment;
6. Check water quality by detecting bacterial contamination from feces;
7. Identify viruses in water samples;
8. Identify disease-causing protozoa in water;
9. Determine specific metabolic pathways and activities occurring in microorganisms;
10. Determine races, distribution patterns, kinships, migration patterns, evolutionary relationships, and rates of evolution of long-extinct species;
11. Accurately settle paternity suits;
12. Confirm identity in amnesia cases;



**Figure 7.17**

### DNA Fingerprinting

Because every person's DNA is unique (a), when samples of an individual's DNA are subjected to restriction enzymes, the cuts will occur in different places and DNA chunks of different sizes will result. (b) Restriction enzymes have the ability to cut DNA at places where specific sequences of nucleotides occur. When the chunks are caused to migrate across an electrophoresis gel (c), the smaller fragments migrate farther than the larger fragments, producing a pattern known as a "DNA fingerprint" (d). Because of individual differences in DNA sequences, these sites vary from person to person. As a result, the DNA fingerprint that separates DNA fragments on the basis of size can appear different from one person to another. Several controls are done in this type of experiment. One of the controls is the presence of the victim's DNA. Why is this important for the interpretation of this test?



13. Identify a person as a relative for immigration purposes;
14. Provide the basis for making human antibodies in specific bacteria;
15. Possibly provide the basis for replicating genes that could be transplanted into individuals suffering from genetic diseases; and
16. Identify nucleotide sequences peculiar to the human genome (an application is currently underway as part of the Human Genome Project).

## Gene Therapy

The fields of biotechnology and bioengineering allow scientists and medical doctors to work together and potentially cure genetic disorders. Unlike contagious diseases, genetic diseases cannot be caught or transmitted because they are caused by a genetic predisposition for a particular disorder—not a separate, disease-causing organism like a bacterium. Gene therapy involves inserting genes, deleting genes, or manipulating the action of genes in order to cure or lessen the effect of genetic diseases. These activities are very new and experimental. While these lines of investigation create hope for many, there are many problems that must be addressed before gene therapy becomes a reliable treatment for many disorders.

The strategy for treating someone with gene therapy varies depending on the individual's disorder. When designing a gene therapy treatment, scientists have to ask themselves, "Exactly what is the problem?" Is the mutant gene not working at all? Is it working normally, but is there too little activity? Is there too much protein being made? Or possibly, is the problem that the gene is acting in a unique and new manner? If there is no gene activity or too little gene activity, the scientists need to somehow introduce a more active version of the gene. If there is too much activity or if the problem is caused by the gene having a new activity, this excess activity must first be stopped and then the normal activity restored.

To stop a mutant gene from working, scientists must change it. This typically involves inserting a mutation into the protein-coding region of the gene or the promoter region that is necessary to activate the gene. Scientists have used **transposons** and some types of viruses to do this in organisms other than humans for some time now. The difficulty in this technique is to mutate only that one gene without disturbing the other genes of the cells and creating more mutations in other genes. Once the mutant gene is silenced, the scientists begin the work of introducing a "good" copy of the gene. Again there are many difficulties in this process:

- Scientists must find a way of returning the corrected DNA to the cell.
- The corrected DNA must be made a part of the cell's DNA so that it is passed on with each cell division, it doesn't interfere with other genes, and can be transcribed by the cell as needed.
- Finally, cells containing the corrected DNA must be reintroduced to the patient.

Many of these techniques are experimental and the medical community is still evaluating their usefulness in treating many disorders as well as the risks that these techniques pose to the patient.

Recently, the first efforts to determine the **human genome**, the entire human DNA sequence, were completed. Many scientists feel that advances in medical treatments will occur more quickly by having this information available. Three new fields of biology have grown out of these efforts: **genomics**, **transcriptomics**, and **proteomics**. Genomics is the study of the DNA sequence and looks at the significance of how different genes and DNA sequences are related to each other. When genes are identified from the DNA sequence, transcriptomics looks at when, where, and how much a gene is expressed. Finally, proteomics examines the proteins that are predicted from the DNA sequence. From these types of studies, we are able to identify gene families that can be used to determine how humans have evolved on a molecular level, how genes are used in an organism throughout its body and over its lifespan, and how to identify common themes from one protein to the next.

## SUMMARY

The successful operation of a living cell depends on its ability to accurately use the genetic information found in its DNA. The enzymes that can be synthesized using the information in DNA are responsible for the efficient control of a cell's metabolism. However, the production of protein molecules is under the control of the nucleic acids, the primary control molecules of the cell. The structure of the nucleic acids, DNA and RNA, determines the structure of the proteins, whereas the structure of the proteins determines their function in the cell's life cycle. Protein synthesis involves the decoding of the DNA into specific protein molecules and the use of the intermediate molecules, mRNA and tRNA, at the ribosome. Errors in any of the codons of these molecules may produce observable changes in the cell's functioning and can lead to cell death.

DNA replication results in an exact doubling of the genetic material. The process virtually guarantees that identical strands of DNA will be passed on to the next generation of cells.

Methods of manipulating DNA have led to the controlled transfer of genes from one kind of organism to another. This has made it possible for bacteria to produce a number of human gene products.

## THINKING CRITICALLY



*For guidelines to answer this question, visit the Online Learning Center.*

An 18-year-old college student reported that she had been raped by someone she identified as a "large, tanned white man." A student in her biology class fitting that description was said by eyewitnesses to have been, without a doubt, in the area at approximately the

time of the crime. The suspect was apprehended and upon investigation was found to look very much like someone who lived in the area and who had a previous record of criminal sexual assaults. Samples of semen from the woman's vagina were taken during a physical exam after the rape. Cells were also taken from the suspect. He was brought to trial but was found to be innocent of the crime based on evidence from the criminal investigations laboratory. His alibi—that he had been working alone on a research project in the biology lab—held up. Without PCR genetic fingerprinting, the suspect would surely have been wrongly convicted, based solely on circumstantial evidence provided by the victim and the “eyewitnesses.”

Place yourself in the position of the expert witness from the criminal laboratory who performed the PCR genetic fingerprinting tests on the two specimens. The prosecuting attorney has just asked you to explain to the jury what led you to the conclusion that the suspect could not have been responsible for this crime. Remember, you must explain this to a jury of 12 men and women who, in all likelihood, have little or no background in the biological sciences. Please, tell the whole truth and nothing but the truth.

## CONCEPT MAP TERMINOLOGY

Construct a concept map using these terms.

base-pairing	mutation
complementary bases	replication
DNA polymerase	template
DNA repair	



Visit the Online Learning Center to use an interactive format to construct another concept map.

## KEY TERMS



Use interactive flash cards, on the Online Learning Center, to help you learn the meanings of these terms.

adenine	gene
anticodon	gene expression
bioengineering	genetically modified (GM)
biotechnology	genome
central dogma	genomics
chromatin fibers	guanine
chromosomal aberrations	human genome
chromosome	initiator codon
coding strand	messenger RNA (mRNA)
codon	mutagenic agent
complementary base	mutation
cytosine	nitrogenous base
deoxyribonucleic acid (DNA)	nucleic acids
deoxyribose	nucleoproteins
DNA code	nucleosomes
DNA polymerase	nucleotide
DNA replication	point mutation
expressed	promoter region

protein-coding region	termination codon
protein synthesis	termination sequence
proteomics	thymine
recombinant DNA	transcription
regulatory region	transcription factors
ribonucleic acid (RNA)	transcriptomics
ribose	transfer RNA (tRNA)
ribosomal RNA (rRNA)	transgenic organisms
RNA polymerase	translation
telomere	transposons
template	uracil

## LEARNING CONNECTIONS



Visit the Online Learning Center to find additional study resources including those highlighted below.

### 7.1 The Central Dogma

#### Interactive Concept Maps

- The central dogma

1. What is the product of transcription? Translation?
2. What is a gene?

### 7.2 Nucleic Acid Structure and Function

#### Animations and Review

- DNA structure
- Transcription
- Translation

3. What are the differences among a nucleotide, a nitrogenous base, and a codon?
4. What are the differences between DNA and RNA?
5. List the sequence of events that takes place when a DNA message is translated into protein.
6. If a DNA nucleotide sequence is CATAAAGCA, what is the mRNA nucleotide sequence that would base-pair with it?
7. What amino acids would occur in the protein chemically coded by the sequence of nucleotides in question 6?
8. How do tRNA, rRNA, and mRNA differ in function?

### 7.3 Mutations

#### Interactive Concept Map

- Effects of mutations

9. Both chromosomal and point mutations occur in DNA. In what ways do they differ? How is this related to recombinant DNA?
10. What is a silent mutation? Provide an example.

### 7.4 Controlling Gene Expression

#### Animations and Review

- Gene activity

11. Provide two examples of why it is advantageous for a cell to control gene expression.
12. Provide two examples of how a cell uses transcription to control gene expression.



### 7.5 Nucleotide Sequences—The Cell's Legacy

#### Animations and Review

- DNA replication

13. What is polymerase and how does it function?
14. Why is DNA replication necessary?
15. How does DNA replication differ from the manufacture of an RNA molecule?

### 7.6 Using DNA to Our Advantage

#### Experience This!

- Engineering agriculture
- New careers

#### Case Study

- *The New York Times*: Iceland Sells Genes for Profit

#### Food for Thought

- DNA dragnet

#### Interactive Concept Map

- Text Concept Map

16. Provide two examples of agricultural benefits developed from DNA technology.
17. Provide two examples of medical benefits developed from DNA technology.

Test yourself with multiple choice questions that provide immediate feedback on the Online Learning Center with PowerWeb at

[www.mhhe.com/enger11](http://www.mhhe.com/enger11)

