Structure of DNA

The structure of DNA was determined by James Watson and Francis Crick in the early 1950s. The method they used to deduce DNA's structure is reviewed in the Scientific Inquiry feature "Finding the Structure of DNA."

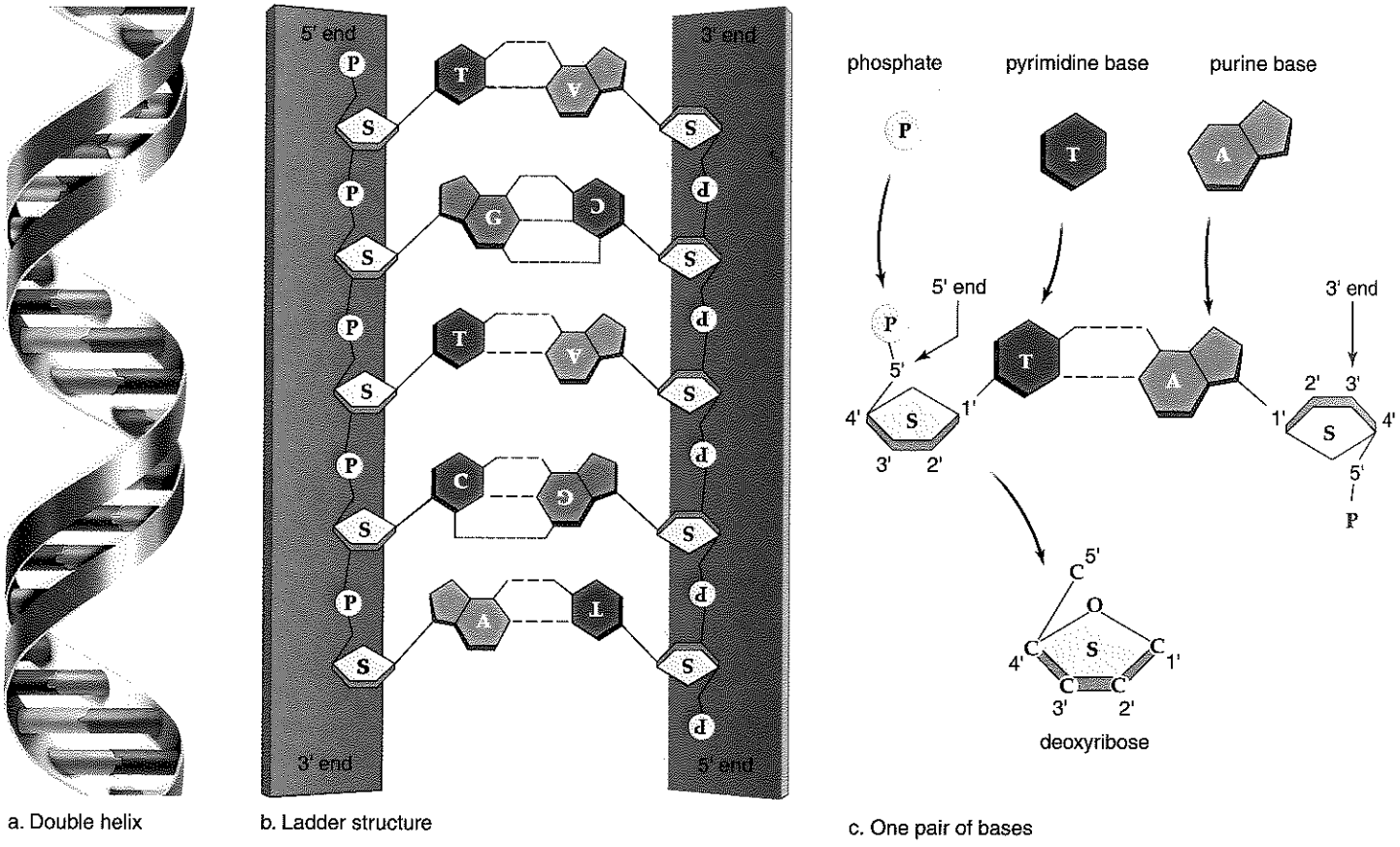
DNA is a chain of nucleotides. Each nucleotide is a complex of three subunits—phosphoric acid (phosphate), a pentose sugar (deoxyribose), and a nitrogen-containing base. There are four possible bases: two are **purines** with a double ring, and two are **pyrimidines** with a single ring. Adenine (A) and guanine (G) are purines; thymine (T) and cytosine (C) are pyrimidines.

A DNA polynucleotide *strand* has a backbone made up of alternating phosphate and sugar molecules. The bases are attached to the sugar but project to one side. DNA has two such strands, and the two strands twist about one another in the form of a **double helix** (Figure 4.3*a*). The strands are held together by hydrogen bonding between the bases: A always pairs with T by forming two hydrogen bonds, and G always pairs with C by forming three hydrogen bonds. Notice that a purine is always bonded to a pyrimidine. This is called **complementary base pairing.** When the DNA helix unwinds, it resembles a ladder (Figure 4.3*b*). The sides of the ladder are the sugar-phosphate backbones, and the rungs of the ladder are the complementary paired bases.

The two DNA strands are *antiparallel,* meaning that they are oriented in opposite directions, which you can verify by noticing that the sugar molecules are oriented differently. The carbon atoms in a sugar molecule are numbered, and the fifth carbon atom (5′) is uppermost in the strand on the left, while the third carbon atom (3′) is uppermost in the strand on the right (Figure 4.3*c*).

Check Your Progress 4.1

1. Summarize the significance of the Griffith and Avery experiments.
2. How did results from the Hershey-Chase experiment suggest that DNA was the genetic material.
3. Describe the structure of the DNA molecule.



a. Double helix b. Ladder structure c. One pair of bases

Figure 4.3 Overview of DNA structure. **a.** DNA double helix. **b.** Unwinding the helix reveals a ladder configuration in which the sides are composed of sugar and phosphate molecules and the rungs are complementary bases. The bases in DNA pair in such a way that the sugar-phosphate backbones are oriented in different directions. **c.** Notice that 3′ and 5′ are part of the system for numbering the carbon atoms that make up the sugar.

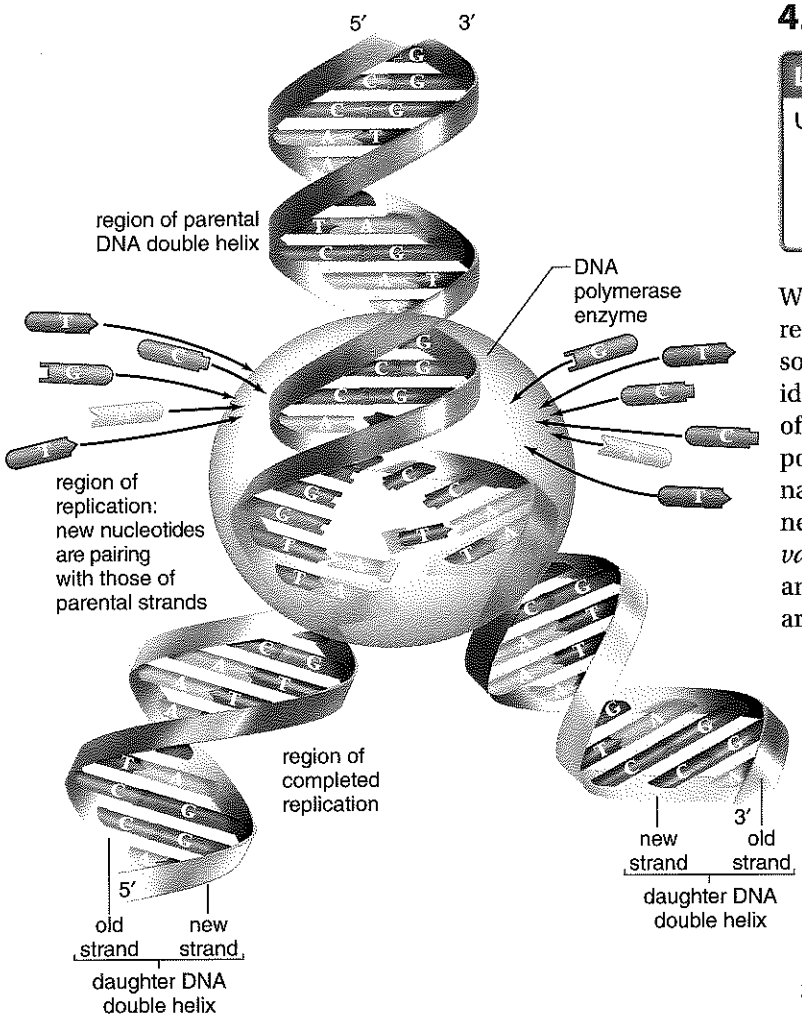


Figure 4.4 Overview of DNA replication. Replication is called semiconservative because each new double helix is composed of an old (parental) strand and a new (daughter) strand.

4.2 DNA Replication

Learning Outcomes

Upon completion of this section, you should be able to

1. Explain why DNA replication is semiconservative.
2. Summarize the events that occur during the process of DNA replication.

When the body grows or heals itself, cells divide. Each new cell requires an exact copy of the DNA contained in the chromosomes. The process of copying one DNA double helix into two identical double helices is called **DNA replication.** The process of DNA replication is carried out by an enzyme called DNA polymerase (Figure 4.4). The DNA polymerase uses each original strand as a template for the formation of a complementary new strand. As a result, DNA replication is termed *semiconservative* because a new double helix has one conserved old strand and one new strand. Replication results in two DNA helices that are identical to each other and to the original molecule.

At the molecular level, several enzymes and proteins participate in the synthesis of the new DNA strands. This process is summarized in Figure 4.5:

1. The enzyme *DNA helicase* unwinds and "unzips" the double-stranded DNA by breaking the weak hydrogen bonds between the paired bases.
2. New complementary DNA nucleotides, always present in the nucleus, fit into place by the process of complementary base pairing. These are positioned and joined by the enzyme *DNA polymerase.*
3. Because the strands of DNA are oriented in an antiparallel configuration, and the DNA polymerase may add new nucleotides only to one end of the chain, DNA synthesis occurs in opposite directions. The *leading strand* follows the helicase enzyme, while synthesis on the *lagging strand* results in the formation of short segments of DNA called *Okazaki fragments.*
4. To complete replication, the enzyme *DNA ligase* connects the Okazaki fragments and seals any breaks in the sugar-phosphate backbone.
5. The two double helix molecules are identical to each other and to the original DNA molecule.

Check Your Progress 4.2

1. Explain why DNA replication is said to be semiconservative.
2. Summarize the sequence of events that occur during DNA replication.
3. Describe the key enzymes involved in DNA replication.

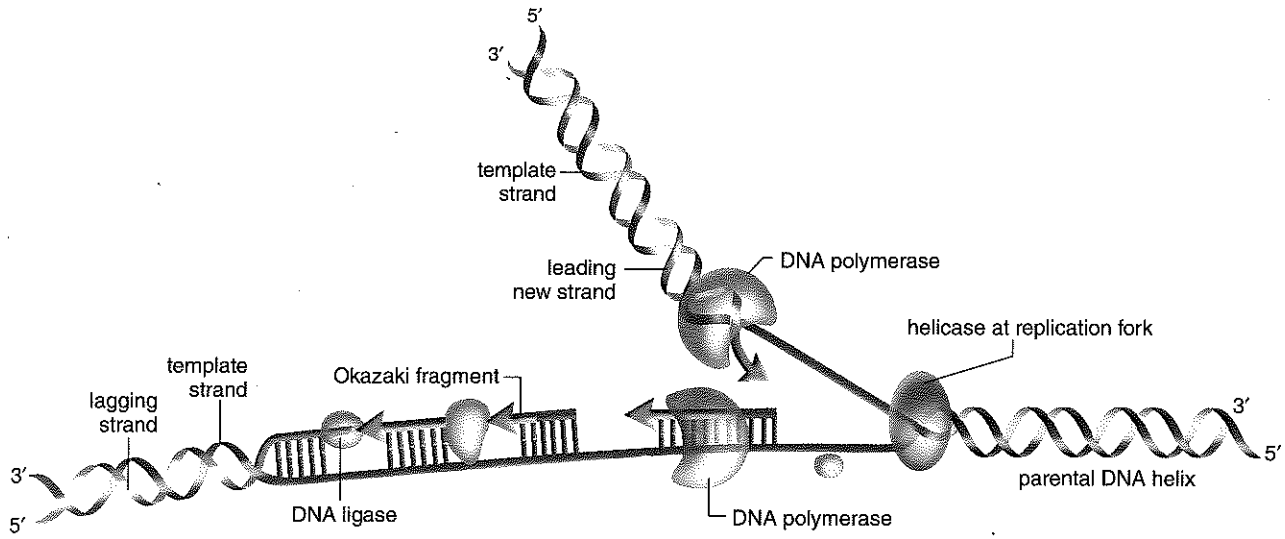


Figure 4.5 Molecular mechanisms of DNA replication. The major enzymes involved in DNA replication. Note that the synthesis of the new DNA molecules occurs in opposite directions due to the orientation of the original DNA strands.

4.3 Gene Expression

Learning Outcomes

Upon completion of this section, you should be able to

1. Describe the roles of RNA molecules in gene expression.
2. Summarize the sequence of events that occurs during gene expression.
3. Determine the sequence of amino acids in a peptide, given the messenger RNA sequence.
4. Explain the purpose of mRNA processing.

The process of using a gene sequence to synthesize a protein is called gene expression. Gene expression relies on the participation of several different forms of **RNA** (**ribonucleic acid**) molecules, most important of which are messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA). Recall from Section 2.8 that DNA and RNA have several structural differences. These are summarized in Table 4.1.

Overall, gene expression requires two processes called transcription and translation. In eukaryotes, transcription takes place in the nucleus and translation takes place in the cytoplasm. During **transcription,** a portion of DNA serves as a template for mRNA formation. During **translation,** the sequence of mRNA bases (which are complementary to those in the template DNA) determines the sequence of amino acids in a polypeptide. So, in effect, genetic information lies in the sequence of the bases in DNA, which through mRNA determines the sequence of amino acids in a protein. Transfer RNA assists mRNA during protein synthesis by bringing amino acids to the ribosomes. Proteins differ from one another by the sequence of their amino acids, and proteins determine the structure and function of cells and the physical characteristics or phenotype of the organism.

TABLE 4.1 Comparison of DNA and RNA

	DNA	RNA
Sugar	Deoxyribose	Ribose
Bases	Adenine, guanine, thymine, cytosine	Adenine, guanine, uracil, cytosine
Strands	Double stranded	Single stranded
Helix	Yes	No

Transcription

During transcription, a segment of the DNA called a **gene** serves as a template for the production of an RNA molecule. Historically, a gene was considered to be a nucleic acid sequence that codes for the sequence of amino acids in a protein. We now know that not all genes contain instructions for protein formation. Some genes include instructions for the formation of DNA molecules, such as mRNA, tRNA, and rRNA. We also know that protein-coding regions can be interrupted by regions that do not code for a protein. In recognition of these new findings, Mark Gerstein and associates in 2007 suggested a new definition for a gene: "A gene is a genomic sequence (either DNA or RNA) directly encoding functional products, either RNA or protein."[1] Although all three classes of RNA are formed by transcription, we will focus on transcription to form messenger RNA (mRNA), the first step in protein synthesis.

[1]Gerstein, M. B., Bruce, C., and Rozowsky, J. S. et al. "What is a gene, post-ENCODE? History and updated definition," *Genomic Research* 17:669–681 (2007).

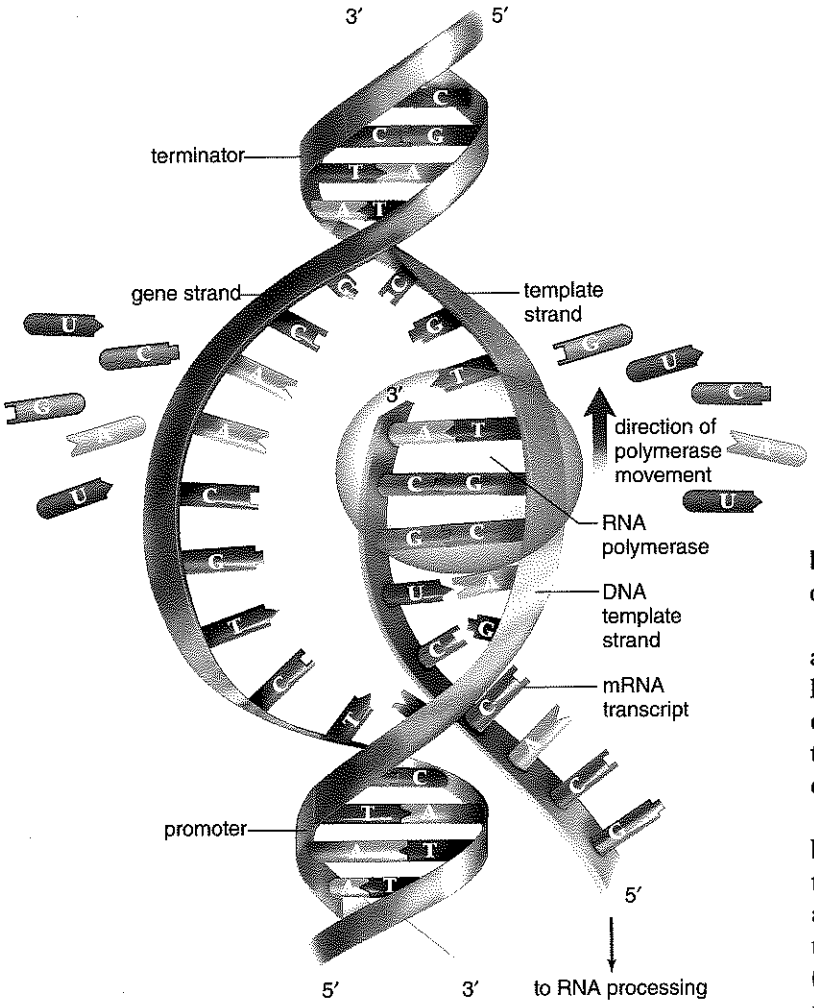


Figure 4.6 Transcription of DNA to form mRNA. During transcription, complementary RNA is made from a DNA template. At the point of attachment of RNA polymerase, the DNA helix unwinds and unzips, and complementary RNA nucleotides are joined together. After RNA polymerase has passed by, the DNA strands rejoin and the mRNA transcript is released.

Messenger RNA

The purpose of **messenger RNA (mRNA)** is to carry genetic information from the DNA to the ribosomes for protein synthesis. Messenger RNA is formed by the process of transcription which, in eukaryotes, occurs in the nucleus. Transcription begins when the enzyme **RNA polymerase** binds tightly to a **promoter,** a region of DNA that contains a special sequence of nucleotides. This enzyme opens up the DNA helix just in front of it so that complementary base pairing can occur in the same way as in DNA replication. Then, RNA polymerase inserts the RNA nucleotides, and an mRNA molecule results. When mRNA forms, it has a sequence of bases complementary to that of the DNA; wherever A, T, G, or C are present in the DNA template, U, A, C, or G, respectively, are incorporated into the mRNA molecule (Figure 4.6). Now, mRNA is a faithful copy of the sequence of bases in DNA.

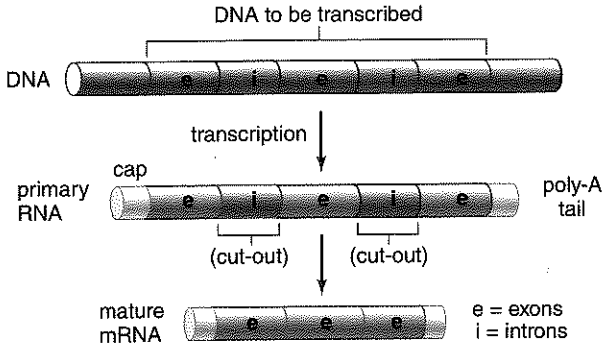


Figure 4.7 mRNA processing. During processing, a cap and tail are added to mRNA, and the introns (i) are removed so that only exons (e) remain.

Processing of mRNA After the mRNA is transcribed in eukaryotic cells, it must be processed before entering the cytoplasm.

The newly synthesized *primary mRNA* molecule becomes a *mature mRNA* molecule after processing. Most genes in humans are interrupted by segments of DNA that are not part of the gene. These portions are called **introns.** The other portions of the gene are called **exons** because they are ultimately expressed (Figure 4.7). Only exons result in a protein product.

Primary mRNA contains bases that are complementary to both exons and introns, but during processing, (1) one end of the mRNA is modified by the addition of a cap, composed of an altered guanine nucleotide, and the other end is modified by the addition of a poly-A tail, a series of adenosine nucleotides. (2) The introns are removed, and the exons are joined to form a mature mRNA molecule consisting of continuous exons. Ordinarily, processing brings together all the exons of a gene. In some instances, cells use only certain exons to form a mature RNA transcript. The result can be a different protein product in each cell.

Translation

Translation is the second process by which gene expression leads to protein synthesis. Translation requires several enzymes, and several different types of RNA molecules, including mRNA, tRNA, and rRNA.

The Genetic Code

The sequence of bases in DNA is transcribed into mRNA, which ultimately codes for a particular sequence of amino acids to form a polypeptide. Can four mRNA bases (A, C, G, U) provide enough combinations to code for 20 amino acids? If only one base stood for an amino acid (i.e., a "singlet code"), then only four amino acids would be possible. If two bases stood for one amino acid, there would only be 16 possible combinations (4×4). If the code is a triplet, then there are 64 possible triplets of the four bases ($4 \times 4 \times 4$). Each triplet of nucleotides is called a **codon.** The code is also degenerate, meaning that most

amino acids are coded for by more than one codon (Figure 4.8). For example, leucine has six codons and serine has four codons. This degeneracy offers some protection against possibly harmful mutations that change the sequence of bases. Of the 64 codons, 61 code for amino acids, whereas the remaining three are *stop codons* (UAA, UGA, UAG), codons that do not code for amino acids but instead termination of polypeptide synthesis.

The genetic code is just about universal in all living organisms. This means that a codon in a fruit fly codes for the same amino acid as in a bird, a fern, or a human. The universal nature of the genetic code suggests that it dates back to the very first organisms on Earth and that all living organisms have a common evolutionary history.

First Base	Second Base				Third Base
	U	C	A	G	
U	UUU phenylalanine	UCU serine	UAU tyrosine	UGU cysteine	U
	UUC phenylalanine	UCC serine	UAC tyrosine	UGC cysteine	C
	UUA leucine	UCA serine	UAA ***stop***	UGA ***stop***	A
	UUG leucine	UCG serine	UAG ***stop***	UGG tryptophan	G
C	CUU leucine	CCU proline	CAU histidine	CGU arginine	U
	CUC leucine	CCC proline	CAC histidine	CGC arginine	C
	CUA leucine	CCA proline	CAA glutamine	CGA arginine	A
	CUG leucine	CCG proline	CAG glutamine	CGG arginine	G
A	AUU isoleucine	ACU threonine	AAU asparagine	AGU serine	U
	AUC isoleucine	ACC threonine	AAC asparagine	AGC serine	C
	AUA isoleucine	ACA threonine	AAA lysine	AGA arginine	A
	AUG ***(start)*** methionine	ACG threonine	AAG lysine	AGG arginine	G
G	GUU valine	GCU alanine	GAU aspartic acid	GGU glycine	U
	GUC valine	GCC alanine	GAC aspartic acid	GGC glycine	C
	GUA valine	GCA alanine	GAA glutamic acid	GGA glycine	A
	GUG valine	GCG alanine	GAG glutamic acid	GGG glycine	G

Figure 4.8 Messenger RNA codons. Notice that in this chart, each of the codons (in boxes) is composed of three letters representing the first base, second base, and third base. For example, find the box where C for the first base and A for the second base intersect. You will see that U, C, A, or G can be the third base. The bases CAU and CAC are codons for histidine.

Transfer RNA

Transfer RNA (tRNA) molecules bring amino acids to the ribosomes, the site of protein synthesis. Each tRNA molecule is a single-stranded polynucleotide that doubles back on itself such that complementary base pairing creates a bootlike shape. On one end is an amino acid, and on the other end is an **anticodon,** a triplet of three bases complementary to a codon of mRNA (Figure 4.9). Although there are 64 possible codons, there are only 40 different tRNA molecules. This is because of the *wobble effect,* which states that for some tRNAs, the third nucleotide in the mRNA codon may vary. This is believed to provide additional degeneracy to the genetic code, and help protect against mutations that may alter the amino acid sequence of a protein.

When a tRNA–amino acid complex comes to the ribosome, its anticodon pairs with an mRNA codon. For example, if the codon is CGG, what is the anticodon, and what amino acid will be attached to the tRNA molecule? Based on Figure 4.8, the answer to this question is as follows:

Codon (mRNA)	Anticodon (tRNA)	Amino Acid (protein)
CGG	GCC	Arginine

The order of the codons of the mRNA determines the order that tRNA–amino acids come to a ribosome and, therefore, the final sequence of amino acids in a protein.

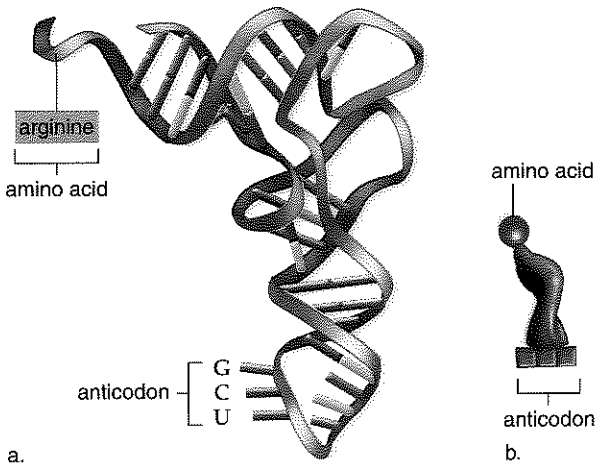


Figure 4.9 Transfer RNA: amino acid carrier. **a.** A tRNA is a polynucleotide that folds into a bootlike shape because of complementary base pairing. At one end of the molecule is its specific anticodon—in this case, GCU (which hybridizes to the codon CGA). At the other end, an amino acid attaches that corresponds to this anticodon—in this case, arginine. **b.** tRNA is represented like this in the illustrations that follow.

Figure 4.10 shows gene expression that results in a protein product. During transcription, the base sequence in DNA is copied into a sequence of bases in mRNA. During translation, tRNAs bring amino acids to the ribosomes in the order dictated by the base sequence of mRNA. The sequence of amino acids forms a polypeptide chain, or complete protein.

Ribosomes and Ribosomal RNA

Ribosomes are small structural bodies found in the cytoplasm and on the endoplasmic reticulum where translation also occurs. Ribosomes are composed of many proteins and several **ribosomal RNAs** (**rRNAs**). In eukaryotic cells, rRNA is produced in a nucleolus within the nucleus. Then the rRNA joins with proteins manufactured in and imported from the cytoplasm to form two ribosomal subunits, one large and one small. The subunits leave the nucleus and join together in the cytoplasm to form a ribosome just as protein synthesis begins.

A ribosome has a binding site for mRNA as well as binding sites for three tRNA molecules. These binding sites facilitate

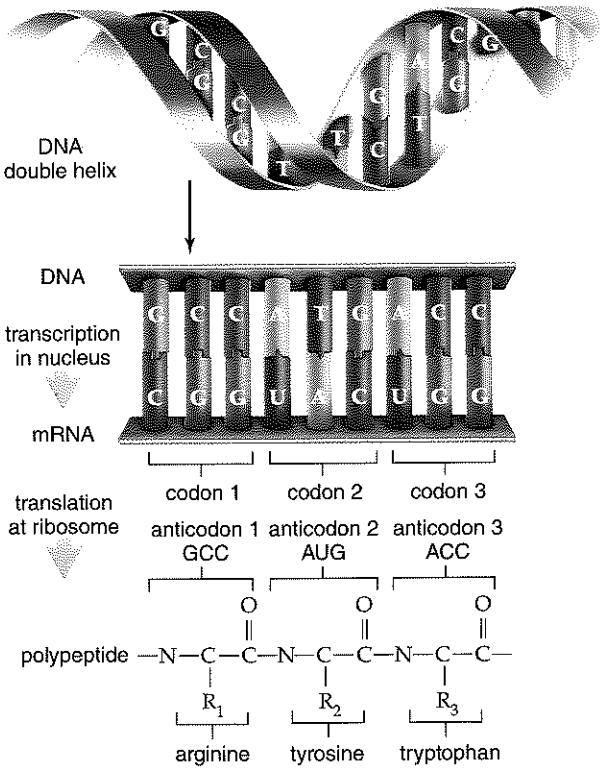


Figure 4.10 Overview of gene expression. One strand of DNA acts as a template for mRNA synthesis, and the sequence of bases in mRNA determines the sequence of amino acids in a polypeptide.

complementary base pairing between tRNA anticodons and mRNA codons. As the ribosome moves down the mRNA molecule, new tRNAs arrive, and a polypeptide forms and grows longer. Translation terminates once the polypeptide is fully formed and an mRNA stop codon is reached. The ribosome then dissociates into its two subunits and falls off the mRNA molecule.

As soon as the initial portion of mRNA has been translated by one ribosome and the ribosome has begun to move down the mRNA, another ribosome attaches to the same mRNA. Therefore, several ribosomes are often attached to and translating a single mRNA, thus forming several copies of a polypeptide simultaneously. The entire complex is called a *polyribosome* or polysome (Figure 4.11).

Translation Requires Three Steps

During translation, the codons of an mRNA base pair with the anticodons of tRNA molecules carrying specific amino acids. The order of the codons determines the order of the tRNA molecules at a ribosome and the sequence of amino acids in a polypeptide. The process of translation must be extremely orderly so that the amino acids of a polypeptide are sequenced correctly.

Protein synthesis involves three steps: initiation, elongation, and termination. Enzymes are required for each of the three steps to function properly. The first two steps, initiation and elongation, require energy.

Initiation

Initiation is the step that brings all the translation components together. Proteins called initiation factors are required to assemble the small ribosomal subunit, as well as mRNA, initiator tRNA, and the large ribosomal subunit for the start of protein synthesis.

Initiation is shown in Figure 4.12. In prokaryotes commonly known as bacteria, a small ribosomal subunit attaches to the mRNA in the vicinity of the *start codon* (AUG). The first or initiator tRNA pairs with this codon because its anticodon is UAC. Then, a large ribosomal subunit joins to the small subunit (Figure 4.12). Although similar in many ways, initiation in eukaryotes is much more complicated and will not be discussed here.

A ribosome has three binding sites for tRNAs. One of these is called the P (for peptide) site, and the other is the A (for amino acid) site. The tRNA exits at the E site. The initiator tRNA is capable of binding to the P site, even though it carries only the amino acid methionine (refer to Figure 4.8). The A site is for tRNA carrying the next amino acid.

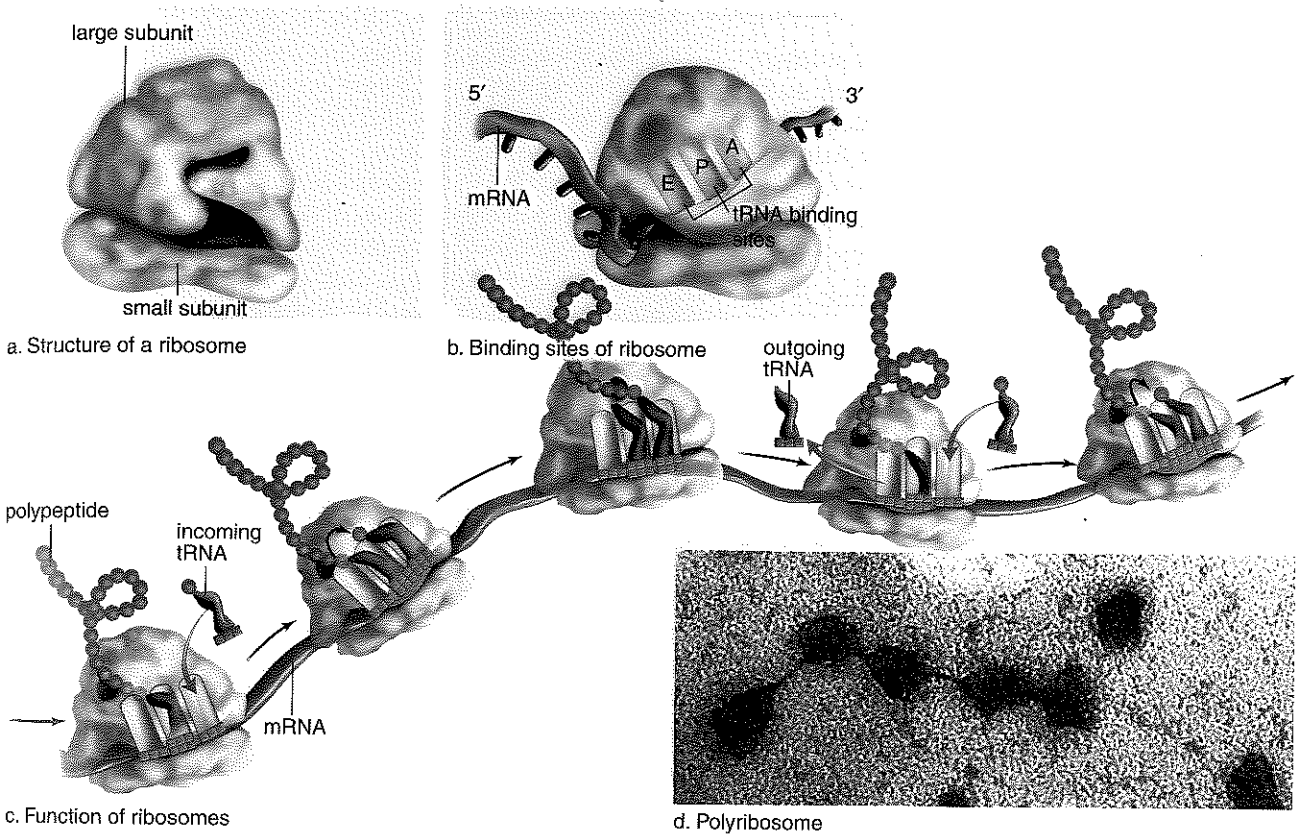


Figure 4.11 Polyribosome structure and function.
a. Structure of a ribosome. **b.** Internal view of a ribosome showing the tRNA binding sites. **c.** Several ribosomes, collectively called a polyribosome, move along an mRNA at one time. Therefore, several polypeptides can be made at the same time. **d.** Electron micrograph of a polyribosome.

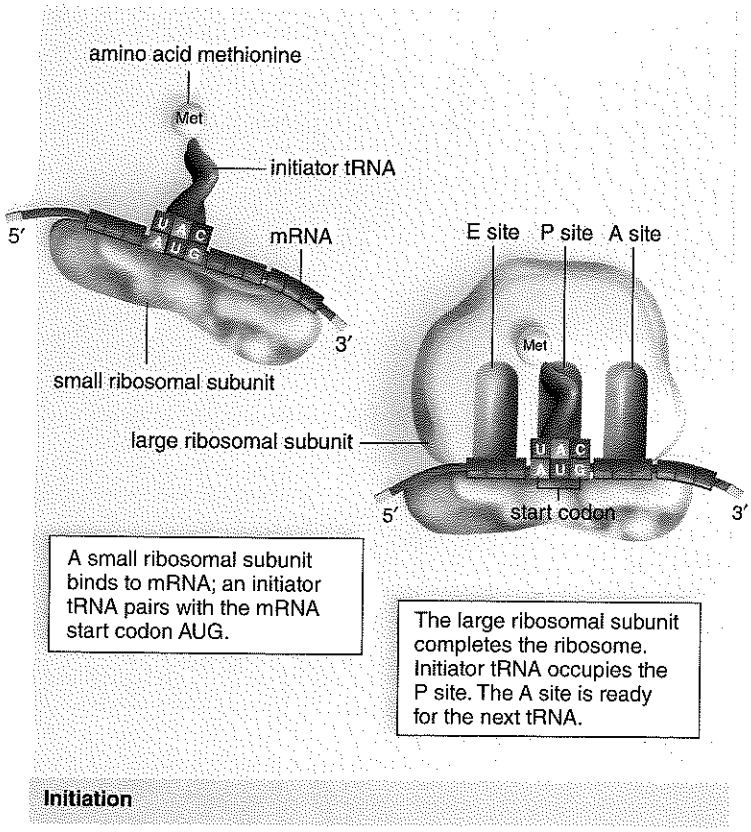


Figure 4.12 Initiation. During initiation, participants in the translation process assemble as shown. The start codon, AUG, codes for the first amino acid, methionine.

Elongation

Elongation is the protein synthesis step in which a polypeptide increases in length one amino acid at a time (Figure 4.13). In addition to the participation of tRNAs, elongation requires elongation factors, which facilitate the binding of tRNA anticodons to mRNA codons at a ribosome.

Elongation consists of a series of four steps (Figure 4.13):

1. A tRNA with an attached peptide is already at the P site, and a tRNA carrying the next amino acid in the chain is just arriving at the A site.
2. Once the next tRNA is in place at the A site, the peptide chain will be transferred to this tRNA.
3. Energy and part of the ribosomal subunit are needed to bring about this transfer. The energy contributes to peptide bond formation, which makes the peptide one amino acid longer by adding the peptide from the A site.
4. Next, translocation occurs—the mRNA moves forward one codon length, and the peptide-bearing tRNA is now at the ribosome P site. The "spent" tRNA now exits. The new codon is at the A site and is ready to receive the next complementary tRNA.

The complete cycle in steps 1 to 4 is repeated at a rapid rate (for example, about 15 times each second in *E. coli* bacteria).

Termination

Termination is the final step in protein synthesis. During termination, as shown in Figure 4.14, the polypeptide and the assembled components that carried out protein synthesis are separated from one another.

Termination of polypeptide synthesis occurs at a stop codon. Termination requires a protein called a release factor, which cleaves the polypeptide from the last tRNA. After this occurs, the polypeptide is set free and begins to take on its three-dimensional shape. The ribosome dissociates into its two subunits.

Properly functioning proteins are of paramount importance to the cell and to the organism. For example, if an organism inherits a faulty gene, the result can be a genetic disorder (such as Huntington disease) caused by a malfunctioning protein or a propensity toward cancer. Proteins are the link between the genetic makeup (genotype) and the physical characteristics (phenotype) of organisms. The DNA sequence underlying these proteins distinguishes different types of organisms. In addition to accounting for the difference between cell types, proteins account for the differences between organisms.

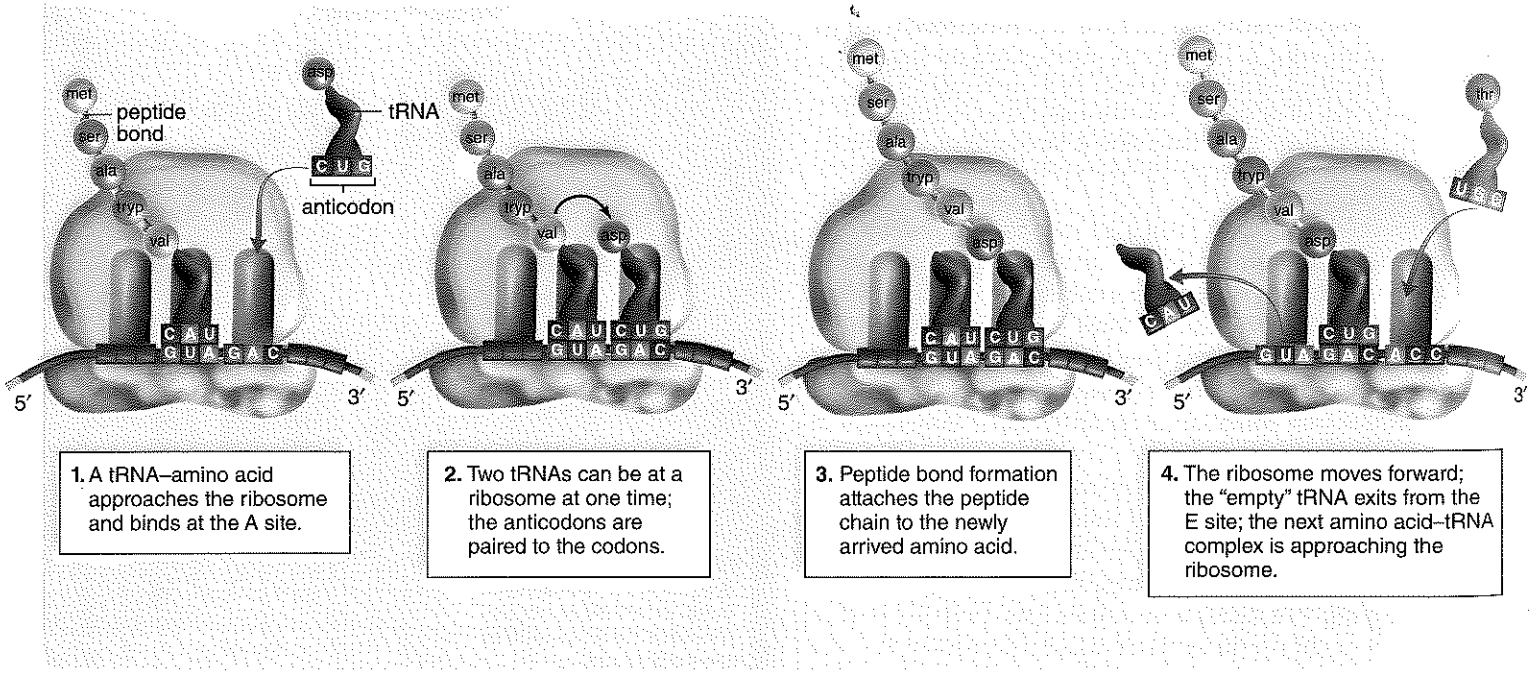


Figure 4.13 Elongation. Note that a polypeptide is already at the P site. During elongation, polypeptide synthesis occurs as amino acids are added one at a time to the growing chain.

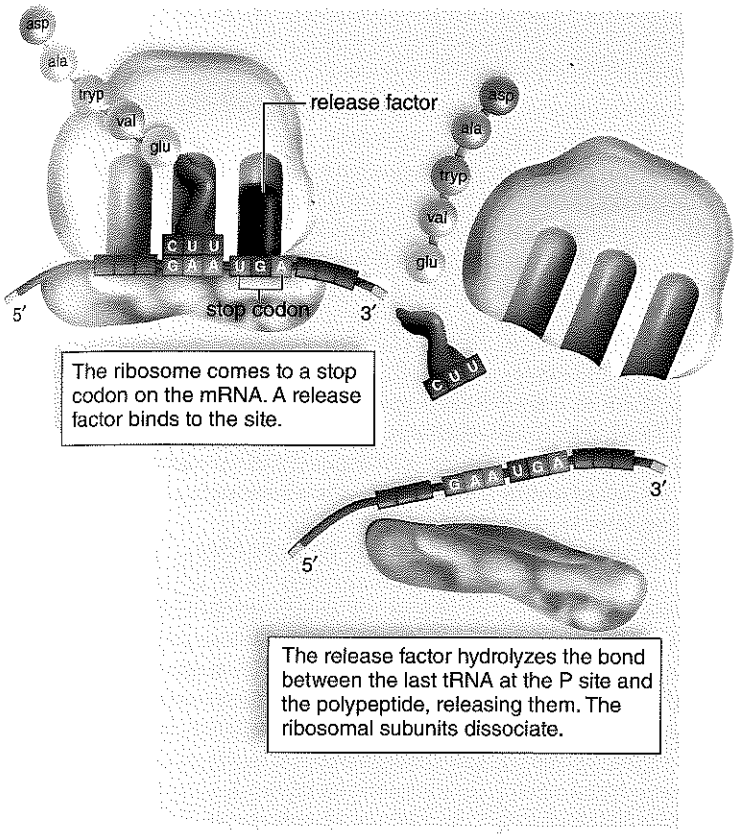


Figure 4.14 Termination. During termination, the finished polypeptide is released, as are the mRNA and the last tRNA.

Review of Gene Expression

A gene is expressed when its protein product has been synthesized. Protein synthesis requires the process of transcription and translation (Figure 4.15). During transcription, a segment of a DNA strand serves as a template for the formation of messenger RNA (mRNA). The bases in mRNA are complementary to those in DNA. Every three mRNA bases is a *codon* (a triplet code) for a certain amino acid. Messenger RNA is processed before it leaves the nucleus, during which time the introns are removed and the ends are modified. Messenger RNA carries a sequence of codons to the *ribosomes.* During translation, tRNAs bring attached amino acids to the ribosomes. Because tRNA anticodons pair with codons, the amino acids become sequenced in the order originally specified by DNA. The genes we receive from our parents determine the proteins in our cells and these proteins are responsible for our inherited traits.

Check Your Progress 4.3

1. Explain the role of mRNA, tRNA, and rRNA in gene expression.
2. Describe the movement of information from the nucleus to the formation of a functional protein.
3. Discuss why the genetic code is said to be degenerate.

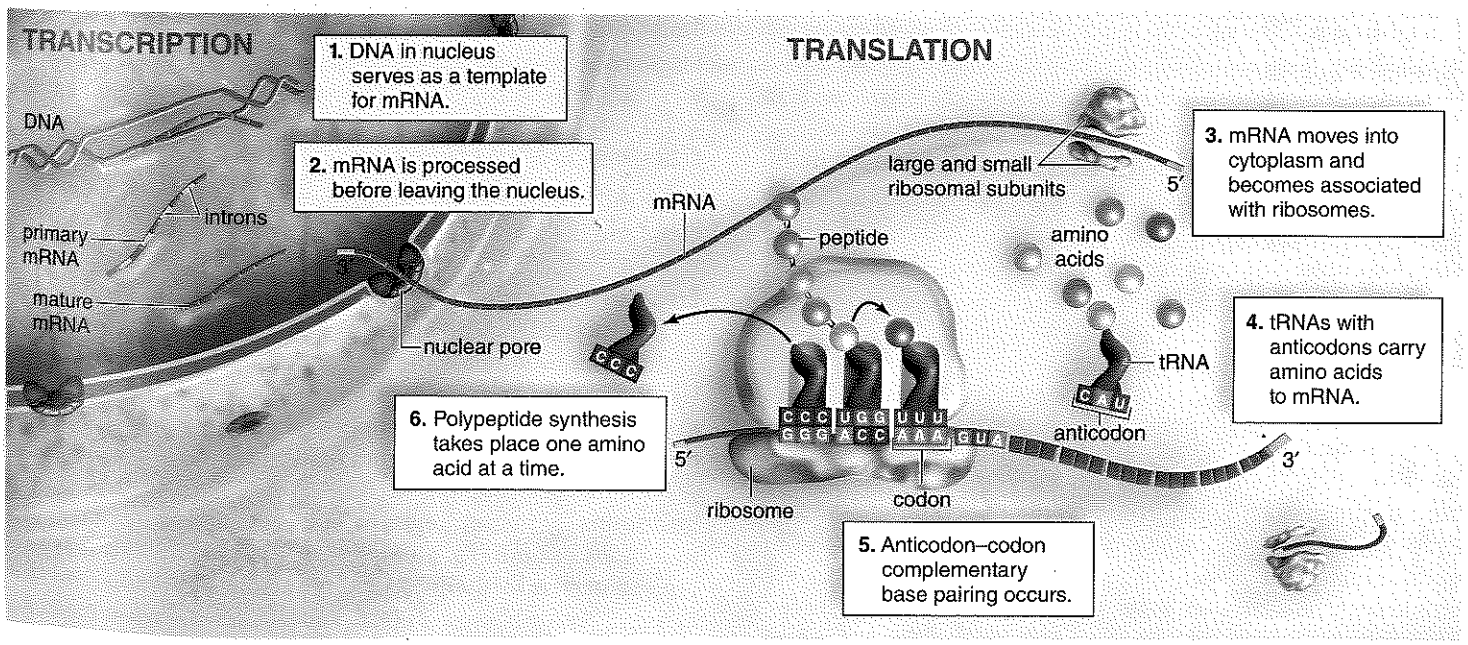


Figure 4.15 Review of gene expression. Messenger RNA is produced and processed in the nucleus during transcription, and protein synthesis occurs at the ribosomes (in cytoplasm and rough ER) during translation.

SCIENCE IN YOUR LIFE ▶ **CASE STUDY**

Mitochondrial DNA: Our Maternal Heritage

If we trace our history back far enough, we can see that we all have a common ancestry. The universal nature of the genetic code suggests that the heritage of all life can be traced back to the first organism. DNA studies can also help us understand how people came to live in different places on Earth.

Mitochondrial DNA (mtDNA) is made from the same components as nuclear DNA but it is circular, thousands of times smaller, and located in mitochondria. It is also more abundant. There are hundreds to thousands of mitochondria in a cell, depending on a particular cell's energy requirements.

In humans, mtDNA is considered the smallest chromosome, coding for only 37 genes and containing about 16 600 base pairs. The genes are essential for normal mitochondrial function. Mitochondrial DNA provides instructions for making the enzymes used in cellular respiration and some of the RNA used in protein synthesis.

Unlike nuclear DNA, which is inherited from both parents, mtDNA is inherited solely from the mother. An egg contains both nuclear DNA and mitochondria with mtDNA. When a sperm fertilizes an egg, only the head of the sperm containing nuclear DNA penetrates the egg. The tail of the sperm, and the mitochondria that power it, do not combine with the egg. Thus, the fertilized egg contains only maternal mitochondria. This means that our mtDNA is identical to the mtDNA of our mothers, grandmothers, and great grandmothers on our mothers' side.

To visualize how we can trace our maternal heritage, imagine the following: if we put all the mothers on Earth in a line, and then put their mothers in another line, and then put the mothers of these mothers in another line, and so on, each line would be smaller than the one before. If we kept on going, back thousands of generations, eventually we would trace back to one mother. This woman has been named Mitochondrial Eve. Mitochondrial Eve is considered to be our matrilineal most recent common ancestor (MRCA). This woman did live amongst other women (and men), but only her mtDNA has survived in an unbroken line and is found in individuals living today. Geneticists estimate that Mitochondrial Eve existed about 190 000 to 200 000 years ago.

Like nuclear DNA, mtDNA can mutate and has been mutating since Mitochondrial Eve. The mutation rate is suggested to be about once every 3500 years. Understanding that mtDNA mutates at a known rate over time, scientists can use this information to determine the relatedness of any two individuals. If their mtDNA is identical, they are very closely related. If their mtDNA is very different, then their last common ancestor lived a long time ago.

The story of our genetic divergence is really a story of our collective shared heritage. Studies of mtDNA fascinate many different kinds of researchers because they allow us to look into our collective past, explore the relatedness of human populations and trace our ancestor's paths across Earth.

Questions to Consider

1. Approximately what population numbers are suggested around the time of Mitochondrial Eve's existence? What do studies of mtDNA suggest about patterns of human migration?
2. Research on mtDNA is sometimes complemented by research that studies mutations of the Y-chromosome, also known as studies of Y-Chromosomal Adam. However, this other research is not quite as clear or easy to define as research on Mitochondrial Eve. How might researchers investigate our patrilineal most recent common ancestor? What would finding information about this ancestor tell us?

4.4 Gene Mutations and Cancer

Learning Outcomes

Upon completion of this section; you should be able to

1. Summarize the causes of gene mutations.
2. Describe why cancer is a failure of genetic control.
3. Describe the characteristics of cancer cells.

A **gene mutation** is a permanent change in the sequence of bases in DNA. The effect of a DNA base sequence change on protein activity can range from no effect to complete inactivity. Germ-line mutations are those that originally occurred in sex cells and can be passed to subsequent generations. We now know that some germ-line mutations can result in cancer. Also, somatic mutations in body cells that are not passed on to future generations can sometimes lead to the development of cancer.

Causes of Mutations

Gene mutations may be caused by errors in replication, mutagens, and the activity of transposons.

Errors in Replication DNA replication errors are a rare source of mutations. DNA polymerase, the enzyme that carries out replication, proofreads the new strand against the old strand. Usually, mismatched pairs are then replaced with the correct nucleotides. In the end, there is typically only one mistake for every 1 billion nucleotide pairs replicated.

Mutagens Environmental influences called **mutagens** cause mutations in humans. Mutagens include radiation (for example, radioactive elements, X rays, ultraviolet [UV] radiation) and certain organic chemicals (for example, chemicals in cigarette smoke and certain pesticides). The rate of mutations resulting from mutagens is generally low because DNA repair enzymes constantly monitor and repair any irregularities.

Transposons **Transposons** are specific DNA sequences that have the remarkable ability to move within and between chromosomes. Their movement to a new location sometimes alters neighbouring genes, particularly by increasing or decreasing their expression. Although "movable elements" in corn were described over 50 years ago by Barbara McClintock, their significance was only realized recently. Also called jumping genes, transposons have now been discovered in almost every species, including bacteria, fruit flies, and humans. McClintock described how the presence of white kernels in corn is due to a transposon located within a gene coding for a pigment-producing enzyme (Figure 4.16*a*, *b*). Maize, a type of corn, displays a variety of colours and patterns because of transposons (Figure 4.16*c*). In a rare human neurological disorder called Charcot-Marie-Tooth disease, a transposon called Mariner gradually causes nerve damage and muscle weakness.

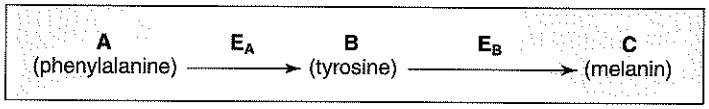
Effect of Mutations on Protein Activity

Point mutations involve a change in a single DNA nucleotide and, therefore, a possible change in a specific amino acid. The base change in the second row of Figure 4.17*a* has no effect on the resulting amino acid in hemoglobin. The change in the third row, however, codes for the amino acid glutamic acid instead of valine. This base change accounts for the genetic disorder sickle cell disease because the incorporation of valine, instead of glutamic acid, causes hemoglobin molecules to form semirigid rods, and the red blood cells become sickle-shaped. Compare Figure 4.17*b* to Figure 4.17*c*. Sickle-shaped cells clog blood vessels and die off more quickly than normal-shaped cells. The base change in the fourth row of Figure 4.17*a* may also have drastic results because the DNA now codes for a stop codon.

Frameshift mutations occur most often because one or more nucleotides are either inserted or deleted from DNA. The result of a frameshift mutation can be a completely new sequence of codons and nonfunctional protein. Here is how this occurs: The sequence of codons is read from a specific starting point, as in the sentence THE CAT ATE THE RAT. If the letter C is deleted from this sentence and the reading frame is shifted, we read THE ATA TET HER AT—something that does not make sense.

Nonfunctional Proteins

A single nonfunctioning protein can have a dramatic effect on phenotype. For example, enzymes are proteins that control reactions in the cell. Cell reactions that involve the synthesis or break down of a biological molecule operate together in a sequential series. This series of reactions is often called a metabolic pathway. Each step or reaction in the pathway is regulated by the activity of a particular enzyme. One metabolic pathway in cells is as follows:



If a faulty code for enzyme E_A is inherited, a person is unable to convert the molecule A (phenylalanine) to B (tyrosine). Phenylalanine builds up in the system, and the excess causes mental impairment and the other symptoms of the genetic disorder phenylketonuria (PKU). In the same pathway, if a person inherits a faulty code for enzyme E_B, then B cannot be converted to C, and the individual has a genetic disorder called albinism, which results in a lack of pigmentation in the hair, eyes, and skin.

A rare condition called androgen insensitivity is due to a faulty receptor for androgens, which are male sex hormones such as testosterone. Although there is plenty of testosterone in the blood, the cells are unable to respond to it. Female instead of male external genitals form, and female instead of male secondary sex characteristics occur. The arrangement of chromosomes is that of a male rather than a female, and the individual does not have the internal sexual organs of a female.

Mutations Can Cause Cancer

It is estimated that one-third of us will develop cancer at some time in our lives. Although cancer death rates decreased by 21% in men and 9% in women between 1988 and 2007, more than 75 000 people in Canada still die from the disease each year. In Canada, the three deadliest forms of cancer are lung cancer, colorectal cancer, and breast cancer.

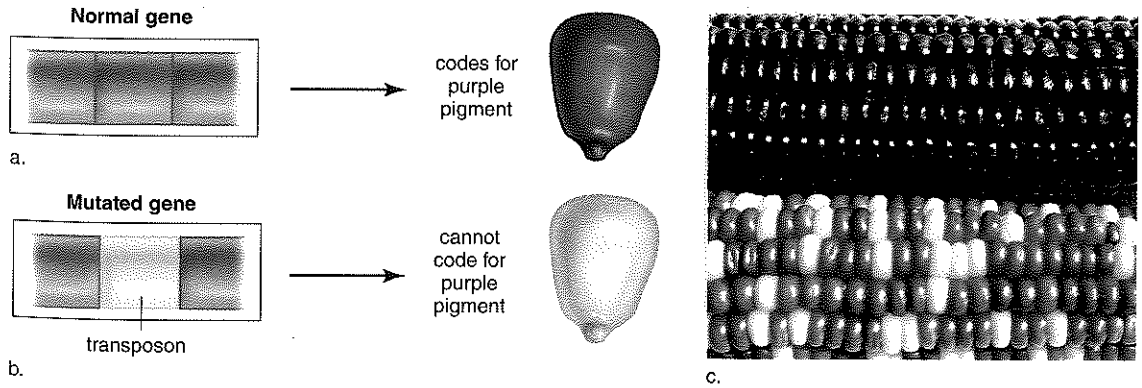
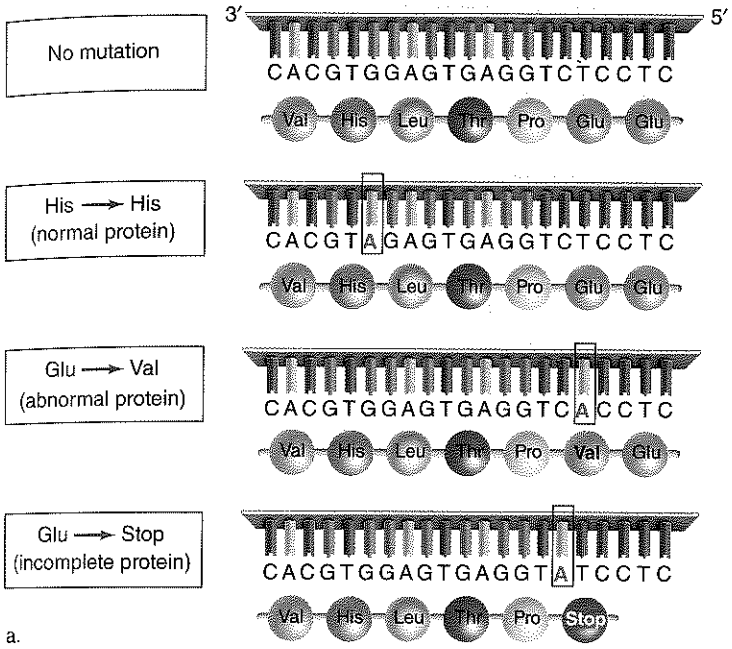
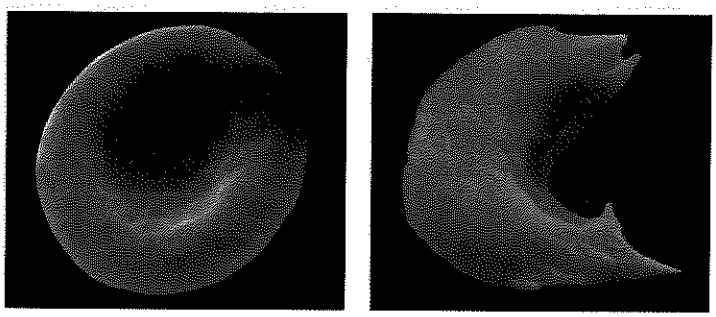


Figure 4.16 Transposon. **a.** A purple coding gene ordinarily codes for a purple pigment. **b.** A transposon "jumps" into the purple-coding gene. This mutated gene is unable to code for purple pigment and a white kernel results. **c.** Maize displays a variety of colours and patterns due to transposon activity.



a.



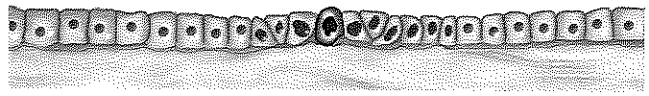
b. Normal red blood cell 7000× c. Sickled red blood cell 7000×

Figure 4.17 Point mutations in hemoglobin. The effect of a point mutation can vary. **a.** Starting at the *top:* Normal sequence of bases in hemoglobin; next, the base change has no effect; next, due to base change, DNA now codes for valine instead of glutamic acid, and the result is that normal red blood cells (**b**) become sickle-shaped (**c**); next, base change will cause DNA to code for termination and the protein will be incomplete.

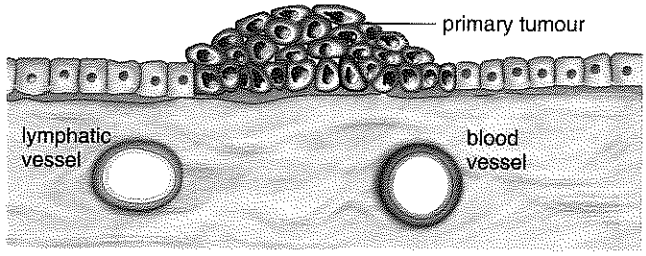
The development of cancer involves a series of accumulating mutations that can be different for each type of cancer. Tumour suppressor genes ordinarily act as brakes on cell division, especially when it begins to occur abnormally. Proto-oncogenes stimulate cell division but are usually turned off in fully differentiated nondividing cells. When proto-oncogenes mutate, they become oncogenes that are active all the time. Carcinogenesis begins with the loss of tumour suppressor gene activity and/or the gain of oncogene activity. When tumour suppressor genes are inactive and oncogenes are active, cell division occurs uncontrollably because a cell signalling pathway that reaches from the plasma membrane to the nucleus no longer functions as it should

Mutations in many genes can contribute to the development of cancer. Several proto-oncogenes code for ras proteins, which are needed for cells to grow, to make new DNA, and to not grow out of control. A point mutation is sufficient to turn a normally functioning *ras* proto-oncogene into an oncogene. Abnormal growth results.

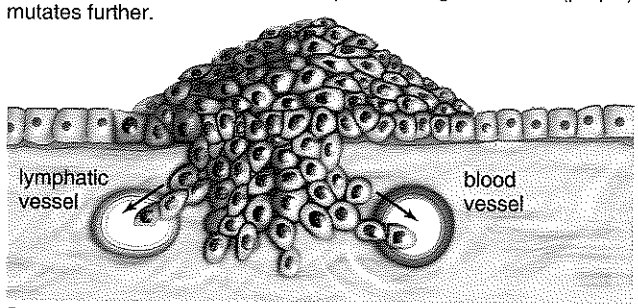
Although cancers vary greatly, they usually follow a common multistep progression (Figure 4.18). Most cancers begin as an abnormal cell growth that is **benign,** or not cancerous, and usually does not grow larger. However, additional mutations may occur, causing the abnormal cells to fail to respond to inhibiting signals that control the cell cycle. When this occurs, the growth becomes **malignant,** meaning that it is cancerous and possesses the ability to spread.



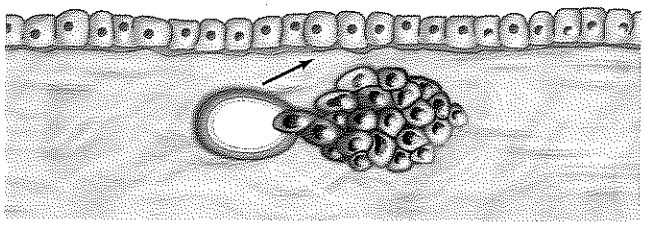
New mutations arise, and one cell (brown) has the ability to start a tumour.



Cancer in situ. The tumour is at its place of origin. One cell (purple) mutates further.



Cancer cells now have the ability to invade lymphatic and blood vessels and travel throughout the body.



New metastatic tumours are found some distance from the primary tumour.

Figure 4.18 Progression of cancer. A single abnormal cell begins the process, and the most aggressive cell, thereafter, becomes the one that divides the most and forms the tumour. Eventually, cancer cells gain the ability to invade underlying tissue and travel to other parts of the body, where they develop new tumours.

Characteristics of Cancer Cells

The primary characteristics of cancer cells are as follows:

Cancer cells are genetically unstable. Generation of cancer cells appears to be linked to mutagenesis. A cell acquires a mutation that allows it to continue to divide. Eventually one of the progeny cells will acquire another mutation and gain the ability to form a tumour. Further mutations occur, and the most aggressive cell becomes the dominant cell of the tumour. Tumour cells undergo multiple mutations and also tend to have chromosomal aberrations and rearrangements.

Cancer cells do not correctly regulate the cell cycle. Cancer cells continue to cycle through the cell cycle. The normal controls of the cell cycle do not operate to stop the cycle and allow the cells to differentiate. Because of that, cancer cells tend to be nonspecialized. Both the rate of cell division and the number of cells increase.

Cancer cells escape the signals for cell death. A cell that has genetic damage or problems with the cell cycle will initiate apoptosis, or programmed cell death. However, cancer cells do not respond to internal signals to die, and they continue to divide even with genetic damage. Cells from the immune system, when they detect an abnormal cell, will send signals to that cell, inducing apoptosis. Cancer cells also ignore these signals.

Most normal cells have a built-in limit to the number of times they can divide before they die. One of the reasons normal cells stop entering the cell cycle is that the telomeres become shortened. **Telomeres** are sequences at the ends of the chromosomes that keep them from fusing with each other. With each cell division, the telomeres shorten, eventually becoming short enough to signal apoptosis. Cancer cells turn on the gene that encodes the enzyme telomerase, which is capable of rebuilding and lengthening the telomeres. Cancer cells thus show characteristics of "immortality" in that they can enter the cell cycle repeatedly.

Cancer cells can survive and proliferate elsewhere in the body. Many of the changes that must occur for cancer cells to form tumours elsewhere in the body are not understood. The cells apparently disrupt the normal adhesive mechanism and move to another place within the body. They travel through the blood and lymphatic vessels and then invade new tissues, where they form tumours. This process is known as **metastasis.** As a tumour grows, it must increase its blood supply by forming new blood vessels, a process called **angiogenesis.** Tumour cells switch on genes that code for the production of growth factors that encourage blood vessel formation. These new vessels supply the tumour cells with the nutrients and oxygen they require for rapid growth and rob normal tissues of nutrients and oxygen.

Check Your Progress 4.4

1. Explain how gene mutations occur.
2. Distinguish between a point mutation and a frameshift mutation.

4.5 DNA Cloning

Learning Outcomes

Upon completion of this section, you should be able to

1. Describe the steps in forming recombinant DNA.
2. Discuss how the polymerase chain reaction works.
3. Explain what is meant by a DNA "fingerprint."

Knowledge of DNA biology has led to an ability to manipulate the genes of organisms. We can clone genes and then use them to alter the **genome** (the complete genetic makeup of an organism) of viruses and cells. This practice, called **genetic engineering,** has innumerable uses, from producing a product to treating genetic disorders.

The Cloning of a Gene

We often think of **cloning** as the production of identical copies of an organism through some asexual means. The members of a bacterial colony on a petri dish are clones because they all came from the division of the same original cell. Human identical twins are also considered clones, because the first two cells of the embryo separated and each became a complete individual.

Another major biological application of cloning is **gene cloning,** which is the production of many identical copies of a single gene. Biologists clone genes for a number of reasons. They might want to produce large quantities of the gene's protein product, such as human insulin, learn how a cloned gene codes for a particular protein, or use the genes to alter the phenotypes of other organisms in a beneficial way. When cloned genes are used to modify a human, the process is called gene therapy. Otherwise, organisms with foreign DNA or genes inserted into them are called **transgenic organisms,** which are frequently used to produce a product desired by humans. While a variety of techniques now exist to produce cloned DNA, most processes rely on recombinant DNA technology and the polymerase chain reaction (PCR).

Recombinant DNA Technology

Recombinant DNA (rDNA) contains DNA from two or more different sources, such as the human cell and the bacterial cell in Figure 4.19. To make rDNA, a researcher needs a **vector,** a piece of DNA that can be manipulated such that foreign DNA can be added to it. One common vector is a plasmid. **Plasmids** are small accessory rings of DNA from bacteria that are not part of the bacterial chromosome and are capable of self-replicating. Plasmids were discovered by investigators studying the bacterium *Escherichia coli.*

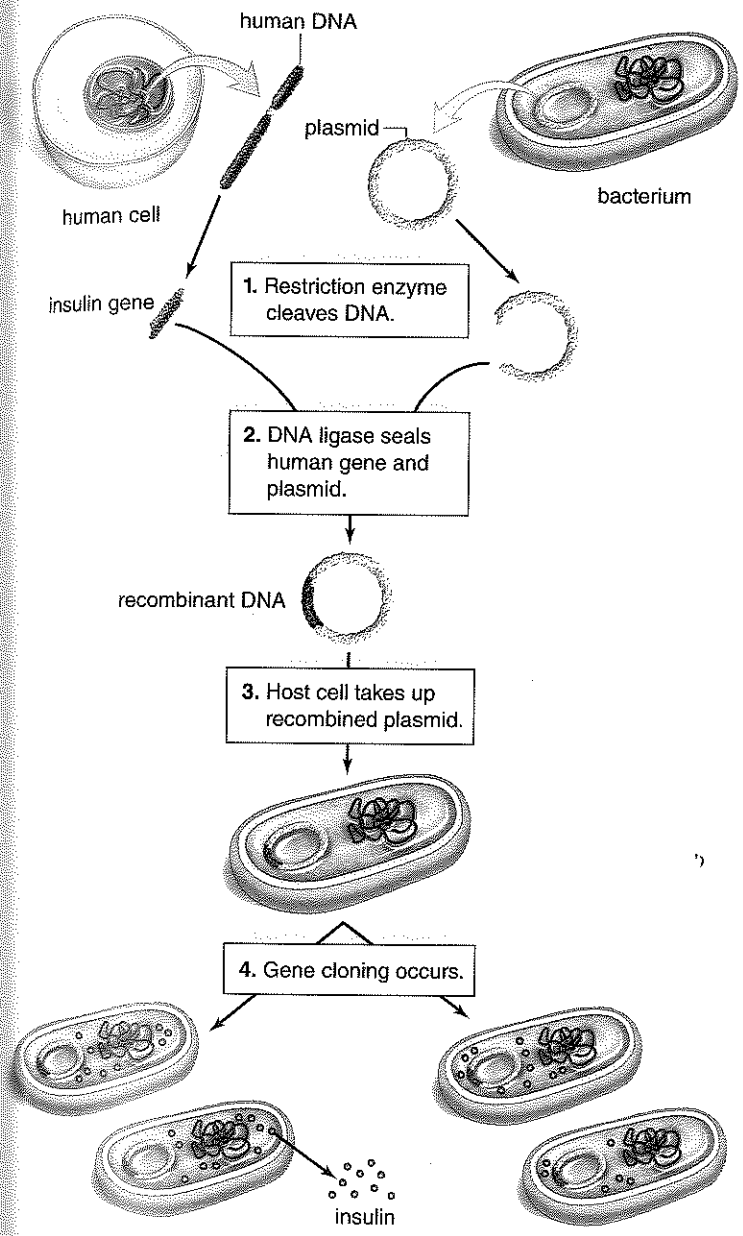
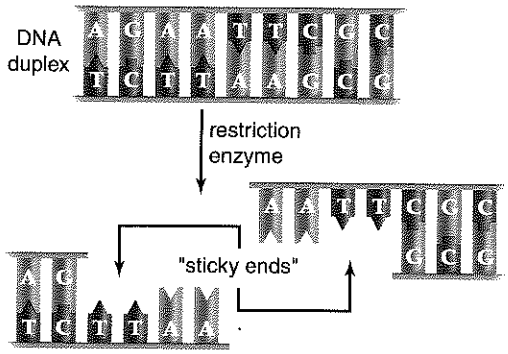


Figure 4.19 Cloning a human gene. Human DNA and bacterial plasmid DNA are cleaved by a specific type of restriction enzyme. For example, human DNA containing the insulin gene is spliced into a plasmid by the enzyme DNA ligase. Gene cloning is achieved after a bacterium takes up the plasmid. If the gene functions normally as expected, the product (for example, insulin) may also be retrieved.

Two types of enzymes that are used to introduce foreign DNA into vector DNA are: (1) a **restriction enzyme** to cleave the vector DNA and (2) DNA ligase to seal DNA fragments. Both DNAs are cleaved with the restriction enzyme so they have complementary ends that can anneal. Hundreds of restriction enzymes occur naturally in bacteria, where they act as a primitive immune system by cutting up any viral DNA that enters the cell. They are called restriction enzymes because they *restrict* the growth of viruses, but they can also be used as molecular scissors to cut double-stranded DNA at a specific site. For example, the restriction enzyme called *Eco*RI always recognizes and cuts double-stranded DNA in the following manner when DNA has the sequence of bases GAATTC:



Notice that a gap now exists into which a piece of foreign DNA can be placed if it ends in bases complementary to those exposed by the restriction enzyme. That is why both the foreign DNA and the vector DNA are cleaved with the same restriction enzyme. The single-stranded, but complementary, ends of the two DNA molecules are called "sticky ends" because they can bind a piece of foreign DNA by complementary base pairing. Sticky ends facilitate the insertion of foreign DNA into vector DNA.

DNA ligase, an enzyme that functions in DNA replication, is then used to seal the foreign piece of DNA into the vector. Bacterial cells take up recombinant plasmids, especially if the cells are treated to make them more permeable. Thereafter, as the plasmid replicates, so does the foreign DNA and thus the gene is cloned.

The Polymerase Chain Reaction

The **polymerase chain reaction (PCR)** can create billions of copies of a segment of DNA in a test tube in a matter of hours. PCR is very specific—it *amplifies* (makes copies of) a targeted DNA sequence, usually a few hundred bases in length. PCR requires the use of DNA polymerase, the enzyme that carries out DNA replication, and a supply of nucleotides for the new DNA strands. PCR involves three basic steps that occur repeatedly, usually for about 35 to 40 cycles: (1) a denaturation step at 95°C, where DNA is heated to become single stranded; (2) an annealing step at a temperature usually between 50° and 60°C, where an oligonucleotide primer hybridizes to each of the single DNA strands; and (3) an extension step at 72°C, where an engineered DNA polymerase adds complementary bases to each of the primers, creating double-stranded DNA.

PCR is a chain reaction because the targeted DNA is repeatedly replicated, much in the same way natural DNA replication occurs, if it continuously occurred. Figure 4.20 uses colour to distinguish the old strand from the new DNA strand. Notice that the amount of DNA doubles with each replication cycle. Thus, assuming you start with only one copy of DNA, after one cycle

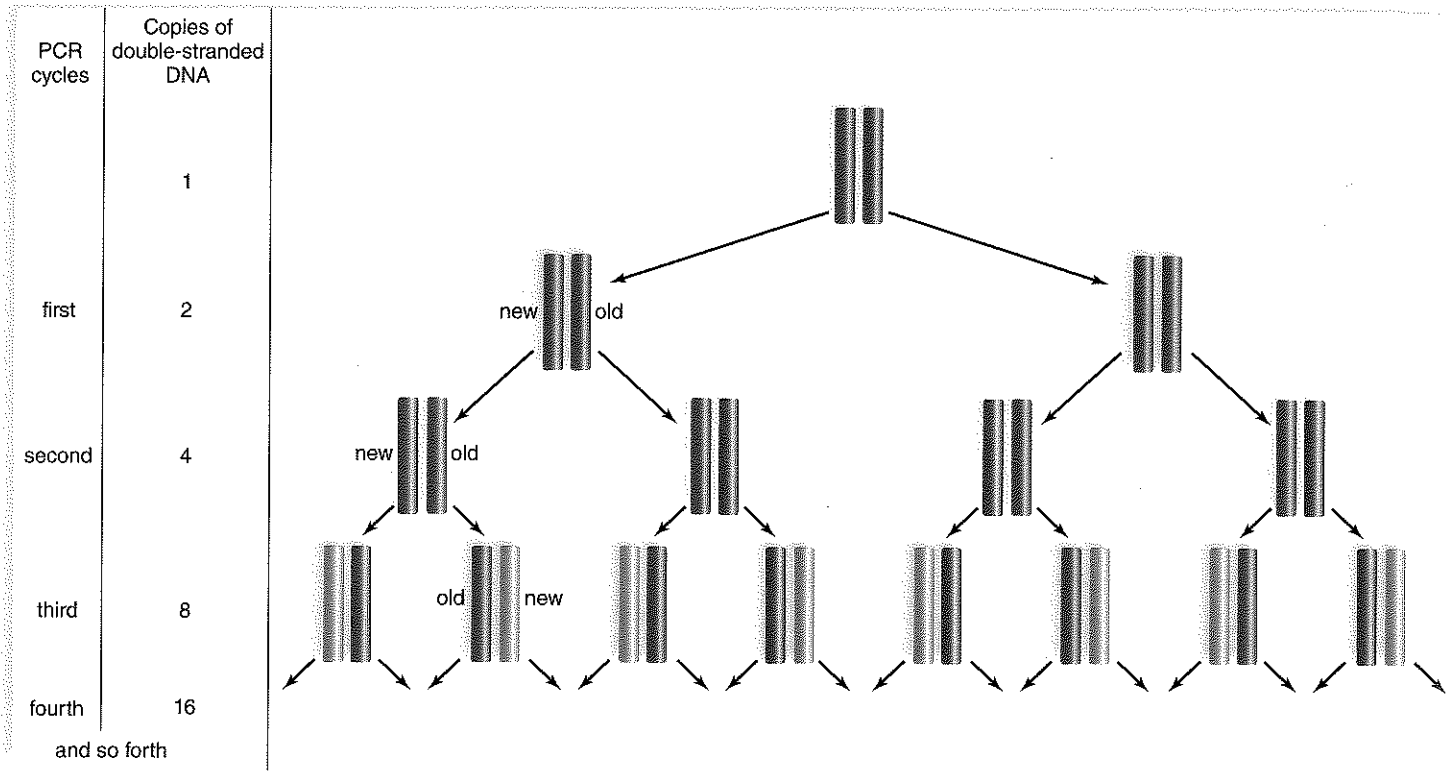


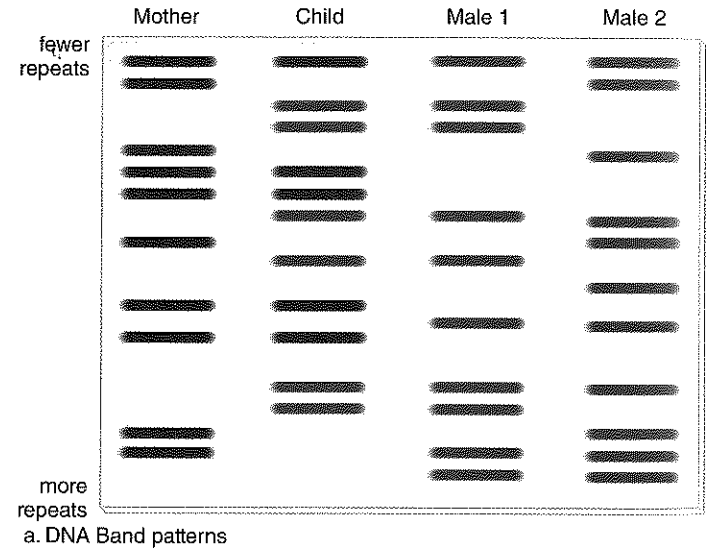
Figure 4.20 Polymerase chain reaction (PCR). PCR allows the production of many identical copies of DNA in a laboratory setting. Assuming you start with only one copy, you can create millions of copies with only 20 to 25 cycles. For this reason, only a tiny DNA sample is necessary for forensic genetics.

you will have two copies, after two cycles four copies, and so on. PCR has been in use since its development in 1985 by Kary Banks Mullis, and now almost every laboratory has automated PCR machines to carry out the procedure. Automation became possible after a temperature-insensitive (thermostable) DNA polymerase was extracted from the bacterium *Thermus aquaticus,* which lives in hot springs. The enzyme can withstand the high temperature used to denature double-stranded DNA.

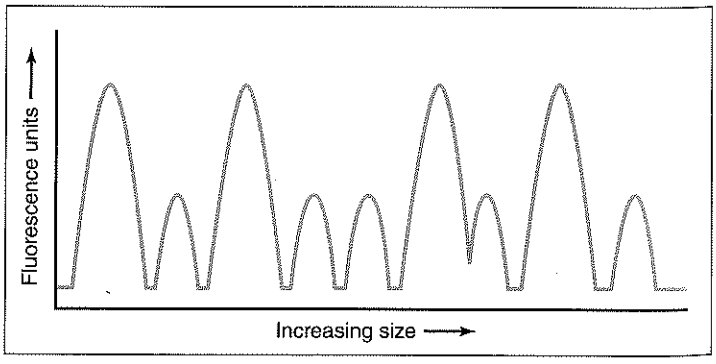
DNA Analysis

There are many reasons for analyzing DNA. One method is called DNA fingerprinting, which is a technique that involves treating the entire genome with restriction enzymes. Because each person has their own restriction enzyme sites, they would have a unique collection of DNA fragment sizes. During a process called *gel electrophoresis,* whereby an electrical current is used to force DNA through a porous gel material, these fragments are separated according to their size. Smaller fragments move farther through the gel than larger fragments, and result in a pattern of distinctive bands, called a **DNA fingerprint.**

Now, **short tandem repeat (STR)** profiling is the method of choice. STRs are the same short sequence of DNA bases that recur several times, as in GATAGATAGATA. STR profiling is



a. DNA Band patterns



b. Automated DNA fingerprinting

Figure 4.21 The use of STR profiling to establish paternity. **a.** In this method, DNA fragments containing STRs are separated by gel electrophoresis. Male 1 is the father. **b.** Each person's profile (only one shown) represents a unique pattern.

advantageous because it does not require the use of restriction enzymes. Instead, PCR is used to amplify target sequences of DNA, which are fluorescently labelled. The PCR products are placed in an automated DNA sequencer. As the sequences move through the sequencer, the fluorescent labels are picked up by a laser. A detector then records the length of each DNA fragment. The fragments are different lengths because each person has their own number of repeats at the particular location of the STR on the chromosome (i.e., each STR locus). That is, the greater the number of STRs at a locus, the longer the DNA fragment amplified by PCR. If individuals are homozygotes, they will have a single fragment, and heterozygotes will have two fragments of different lengths (Figure 4.21*a*). The more STR loci employed, the more confident scientists can be of distinctive results for each person (Figure 4.21*b*).

Applications of PCR are limited only by our imaginations. When the DNA matches that of a virus or mutated gene, it is known that a viral infection, genetic disorder, or cancer is present. DNA fingerprints from blood or tissues at a crime scene have been successfully used in convicting criminals. DNA fingerprinting through STR profiling was extensively used to identify the victims of recent tsunamis in Indonesia and Japan. Relatives can be found, paternity suits can be settled (Figure 4.21), and genetic disorders can be detected. PCR has also shed new light on evolutionary studies by comparing DNA extracted from human mummies thousands of years old or animal fossils millions of years old.

Check Your Progress 4.5

1. Summarize the two required steps for producing recombinant DNA.
2. Explain why STRs may be used for identification.

4.6 Biotechnology Products and Gene Therapy

Learning Outcomes

Upon completion of this section, you should be able to

1. Describe ways in which society benefits from genetically modified bacteria, plants, and animals.
2. Describe the steps involved in the production of a transgenic animal.
3. Compare and contrast ex vivo and in vivo gene therapy.

Today, transgenic bacteria, plants, and animals are often called **genetically modified organisms (GMOs)**, and the products they produce are called **biotechnology** products (Figure 4.22).

Transgenic Bacteria

Recombinant DNA technology is used to produce transgenic bacteria, which are grown in huge vats called bioreactors. The bacteria express the cloned gene, and the gene product is usually collected from the medium in which the bacteria are grown. Biotechnology products produced by bacteria include insulin, human growth hormone, tPA (tissue plasminogen activator), and hepatitis B vaccine.

Figure 4.22 Biotechnology products. Products such as the biodegradable plastic poncho shown here, or fuels like bioethanol, are becoming increasingly common.

Transgenic bacteria have many other uses as well. Some have been produced to promote the health of plants. For example, bacteria that normally live on plants and encourage the formation of ice crystals have been changed from frost-plus to frost-minus bacteria. As a result, new crops such as frost-resistant strawberries are being developed. Also, a bacterium that normally colonizes the roots of corn plants has now been endowed with genes (from another bacterium) that code for an insect toxin. The toxin protects the roots from insects.

Bacteria can be selected for their ability to degrade a particular substance, and this ability can then be enhanced by bioengineering. For instance, naturally occurring bacteria that eat oil can be genetically engineered to do an even better job of cleaning up beaches after oil spills (Figure 4.23), such as the 2010 *Deep Water Horizon* spill in the Gulf of Mexico. Bacteria can also

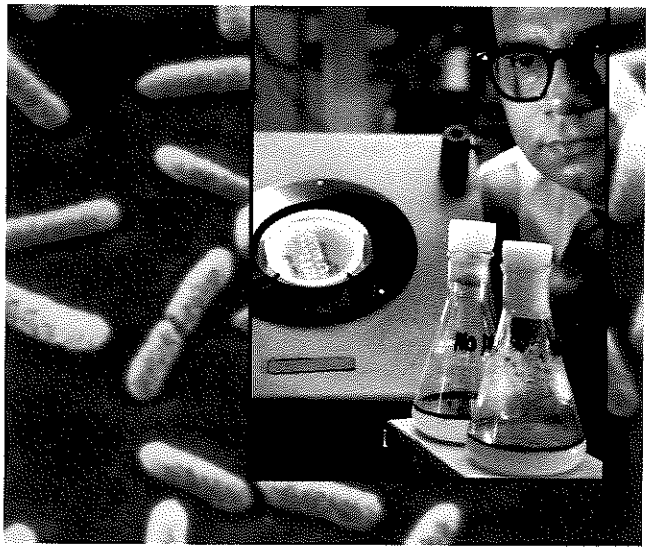


Figure 4.23 Bioremediation. These bacteria, which are capable of decomposing oil, were engineered and patented by the investigator Dr. Ananda Chakrabarty. The flask toward the rear contains oil and no bacteria. The flask toward the front contains the engineered bacteria and is almost clear of oil.

remove sulfur from coal before it is burned, resulting in cleaner emissions. One bacterial strain was given genes that allowed it to clean up levels of toxins that would have killed other bacterial strains. Further, these bacteria were given "suicide" genes that caused them to self-destruct when their job was done.

Organic chemicals are often synthesized by having catalysts act on precursor molecules or by using bacteria to carry out the synthesis. Today, it is possible to go one step further and manipulate the genes that code for these enzymes. For instance, biochemists discovered a strain of bacteria that is especially good at producing phenylalanine, an organic chemical needed to make aspartame. They isolated, altered, and cloned the appropriate genes so that various bacteria could be genetically engineered to produce phenylalanine.

Transgenic Plants

Techniques have been developed to introduce foreign genes into immature plant embryos or into plant cells called *protoplasts* that have had their cell wall removed. It is possible to treat protoplasts with an electric current while they are suspended in a liquid containing foreign DNA. The electric current makes tiny, self-sealing holes in the plasma membrane through which the desired genetic material can enter. Protoplasts go on to develop into mature plants containing and expressing the foreign DNA.

One altered plant known as the pomato is the result of these technologies. This plant produces potatoes below ground and tomatoes above ground. Foreign genes transferred to cotton, corn, and potato strains have made these plants resistant to pests because their cells now produce an insect toxin. Similarly, soybeans have been made resistant to a common herbicide that is sprayed to kill weeds that compete with soybean growth. Some corn and cotton plants are both pest- and herbicide-resistant. These and other genetically engineered crops that are expected to have increased yields are now commonly sold commercially. However, the public is concerned about the possible effect of genetically modified organisms on human health, as well as the environment.

Like bacteria, plants are also being engineered to produce human proteins, such as hormones, clotting factors, and antibodies, in their seeds. One type of antibody made by corn can deliver radioisotopes to tumour cells, and another made by soybeans can be used to treat genital herpes.

Transgenic Animals

Techniques have been developed to insert genes into the eggs of animals. It is possible to microinject foreign genes into eggs by hand, but another method uses vortex mixing. The eggs are placed in an agitator with DNA and silicon-carbide needles. The needles make tiny holes in the eggs through which the DNA can enter. When these eggs are fertilized, the resulting offspring are transgenic animals. Using this technique, many types of animal eggs have acquired the gene for bovine growth hormone (BGH). The procedure has been used to produce larger fishes, cows, pigs, rabbits, and sheep.

Gene pharming, the use of transgenic farm animals to produce pharmaceuticals, is being pursued by a number of firms. Genes that code for therapeutic and diagnostic proteins are incorporated into an animal's DNA, and the proteins appear in the animal's milk. Plans are under way to produce drugs for the treatment of cystic fibrosis, cancer, blood diseases, and other disorders by this method. Figure 4.24 outlines the procedure for producing transgenic animals: DNA containing the gene of interest is injected into donor eggs. Following in vitro fertilization, the zygotes are placed in host females, where they

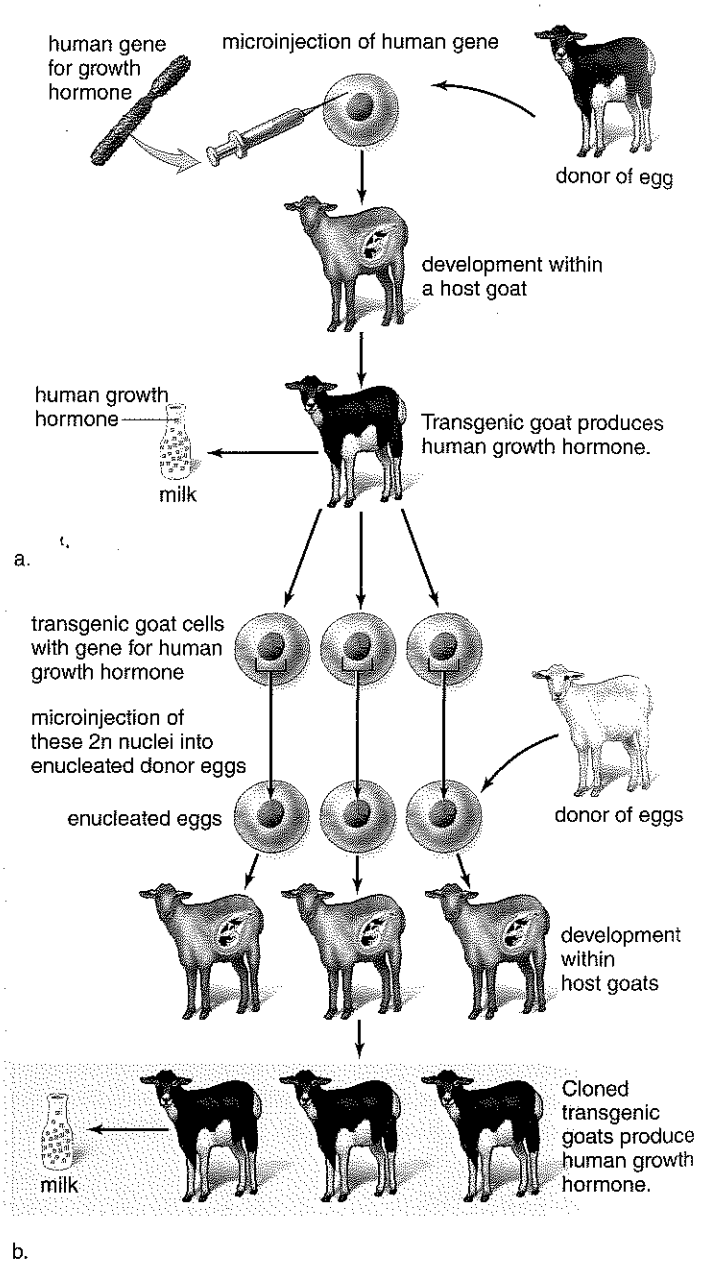


Figure 4.24 Production of transgenic animals.
a. A genetically engineered egg develops in a host to create a transgenic goat that produces a biotechnology product in its milk. **b.** Nuclei from the transgenic goat are transferred into donor eggs, which develop into cloned transgenic goats.

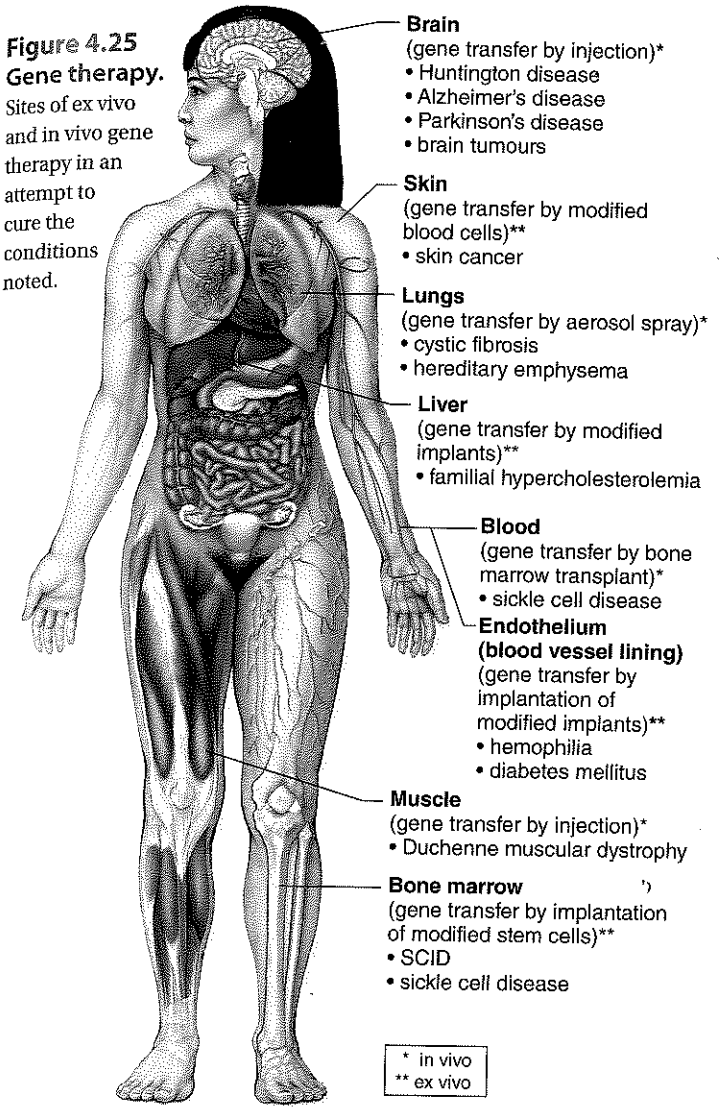


Figure 4.25 Gene therapy. Sites of ex vivo and in vivo gene therapy in an attempt to cure the conditions noted.

develop. After female offspring mature, the product is secreted in their milk.

Eliminating a gene is another way to study a gene's function. Transgenic animals enable researchers to study gene function. A knockout mouse has had both alleles of a gene removed or made nonfunctional. For example, scientists have constructed a knockout mouse lacking the *CFTR* gene, the same gene mutated in cystic fibrosis patients. The mutant mouse has a phenotype similar to a human with cystic fibrosis and can be used to test new drugs for the treatment of the disease.

Gene Therapy

Once a genetic disorder is detected, gene therapy is a potential course of treatment in a few cases. **Gene therapy** is the insertion of genetic material into human cells for the treatment of genetic disorders and various other human illnesses, such as cardiovascular disease and cancer. Figure 4.25 shows regions of the body that have received copies of normal genes by various methods of gene transfer. Viruses genetically modified to be safe can be used to ferry a normal gene into the body, and so can liposomes, which are microscopic globules of lipids specially prepared to enclose the normal gene (i.e., ex vivo gene therapy). On the other hand, sometimes the gene is injected directly into a particular region of the body (i.e., in vivo gene therapy).

Note that despite its promise for treating disorders, gene therapy is still in its infancy and has had detrimental side effects for many patients, such as causing leukemia. Nonetheless, some strides are being made with particular diseases, such as those described in the following sections.

Ex Vivo Gene Therapy

Figure 4.26 describes an ex vivo methodology for treating children who have SCID (severe combined immunodeficiency). These children lack the enzyme ADA (adenosine deaminase),

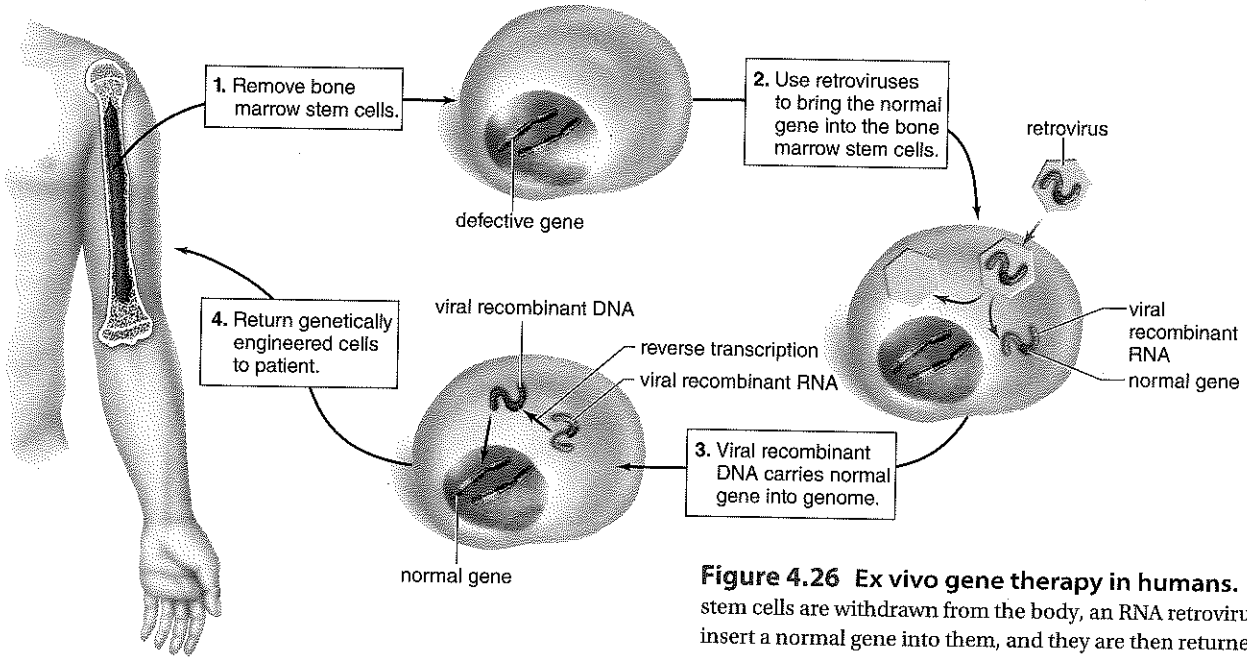


Figure 4.26 Ex vivo gene therapy in humans. Bone marrow stem cells are withdrawn from the body, an RNA retrovirus is used to insert a normal gene into them, and they are then returned to the body.

which is involved in the maturation of T and B cells. Therefore, these children are prone to constant infections and may die without treatment. To carry out gene therapy, bone marrow stem cells are removed from the bone marrow of the patient and infected with a virus that carries a normal gene for the enzyme into their DNA. Then the cells are returned to the patient, where it is hoped they will divide to produce more blood cells with the same genes. Patients who have undergone this procedure show significantly improved immune function associated with a sustained rise in the level of ADA enzyme activity in the blood.

Another example of ex vivo gene therapy has been used to treat familial hypercholesterolemia, a condition that develops when liver cells lack a receptor protein for removing cholesterol from the blood. The high levels of blood cholesterol make the patient subject to fatal heart attacks at a young age. A small portion of the liver is surgically excised and then infected with a virus containing a normal gene for the receptor before being returned to the patient. Patients are expected to experience lowered serum cholesterol levels following this procedure.

In Vivo Gene Therapy

Cystic fibrosis patients lack a gene that codes for the transmembrane carrier of the chloride ion. They often die due to numerous infections of the respiratory tract because a thick mucus forms in the lungs and attracts bacteria and other antigens. In gene therapy trials, the gene needed to cure cystic fibrosis is sprayed into the nose or delivered to the lower respiratory tract by an adenovirus vector or by using liposomes. So far, these treatments have met with limited success, but investigators are trying to improve uptake by using a combination of different vectors.

Gene therapy is increasingly relied upon as a part of cancer treatment. Genes are being used to make healthy cells more tolerant of chemotherapy, while making tumour cells more sensitive. Knowing that the tumour suppressor gene *p53* brings about apoptosis (cell death), researchers are interested in finding a way to selectively introduce *p53* into cancer cells, and in that way, kill them.

Check Your Progress 4.6

1. List some of the beneficial applications of transgenic bacteria, plants, and animals.
2. Distinguish between a transgenic organism and a cloned organism.
3. Summarize the methods that are being used to introduce genes into humans for gene therapy.
4. Provide examples of ex vivo and of in vivo gene therapy.

4.7 Genomics, Proteomics, and Bioinformatics

Learning Outcomes

Upon completion of this section, you should be able to

1. Discuss the implications of knowing the human genome sequence.
2. Describe a major insight from comparative genomics.
3. Compare functional genomics and proteomics.

In the preceding century, researchers discovered the structure of DNA, how DNA replicates, and how DNA and RNA are involved in the process of protein synthesis. Genetics in the 21st century largely concerns **genomics,** the study of the complete genetic sequences of humans and other organisms. Knowing the sequence of bases in genomes is the first step, and mapping their location on the chromosomes is the next step. The immensity of the task can be appreciated by knowing not only that we have approximately 25 000 genes that code for proteins, but also that nearly 99% of the 3.2 billion bases of our genome are noncoding and contain many repetitive sequences of unknown function. Many other organisms have an even larger number of protein-coding genes but fewer noncoding regions when compared to the human genome.

Sequencing the Genome

We now know the sequence of the roughly 3.2 billion pairs of DNA bases in our genome. This feat was accomplished by the **Human Genome Project (HGP)**, a 13-year effort that involved both laboratories around the world. How did they do it? First, investigators developed a procedure that would allow them to decipher a short sequence of base pairs, and then instruments became available that could carry out sequencing automatically. Over the 13-year span, DNA sequencers were constantly improved, and today's instruments can automatically analyze up to 120 million base pairs of DNA in a 24-hour period. So, new genomes are being sequenced all the time, and at a much faster rate than the human genome. The genome of the African clawed frog, *Xenopus laevis,* which is roughly the same size as the human genome, was sequenced in under a year.

Completion of the human genome sequence has opened up great possibilities for biomedical research and treatment. These methods have been used to screen for particular diseases. The HGP also led to the discovery of many small regions of DNA that vary among individuals (polymorphisms). Most of these are *single nucleotide polymorphisms* (*SNPs*), meaning that they have a difference of only one nucleotide. Many SNPs have no effect. Others may contribute to protein-coding differences affecting the phenotype. It is possible that certain