Chapter 9

Gene Expression and Regulation

Review

Chromosomes bear the genetic information that is passed from parents to offspring. The genetic information is stored in molecules of DNA. The DNA is used to make RNA, and RNA assembles amino acids into proteins. Some proteins form the basic structure and appearance of cells, while other proteins, those that function as enzymes, regulate chemical reactions that direct metabolism for cell development, growth, and maintenance. Following is the summary of this process that begins with DNA to create a living, functioning organism.

$$\text{DNA} \rightarrow \text{RNA} \rightarrow \text{proteins/enzymes} \rightarrow \text{traits, metabolism, homeostasis}$$

This flow of information is described as the **central dogma** of molecular biology. The underlying molecular mechanisms for this process are the subject of this chapter.

The structure of DNA and RNA was presented earlier in Chapter 2, "Chemistry of Life." As a review, both DNA and RNA are polymers of **nucleotides.** The nucleotide monomer consists of three covalently bonded parts—a nitrogen base, a sugar, and a phosphate. Figure 9-1 reviews the differences in the structures of DNA and RNA and summarizes their functions. Details of the functions of these molecules will be presented in this chapter.

	Nucleotide Components		Function	Structure
	Sugar	Nitrogen Bases		
DNA	deoxyribose	adenine, *thymine*, guanine cytosine	contains hereditary information (genes) of the cell	double helix
RNA (involved in protein synthesis)	ribose	adenine, *uracil*, guanine cytosine	mRNA (messenger RNA) - provides the instructions for assembling amino acids into a polypeptide chain	linear
			tRNA (transfer RNA) - delivers amino acids to a ribosome for their addition into a growing polypeptide chain	upside-down "L" shape
			rRNA (ribosomal RNA) - combines with proteins to form ribosomes	globular
RNA (involved in RNA processing)	ribose	adenine, *uracil*, guanine cytosine	snRNA (small nuclear RNA) - combines with proteins to form small nuclear ribonucleoproteins (snRNPs) which process RNAs before they leave the nucleus	globular
RNA (involved in regulating gene expression)	ribose	adenine, *uracil*, guanine cytosine	miRNA (microRNA) - regulates gene expression by blocking or degrading mRNA	linear
			siRNA (short interfering RNA) - regulates gene expression by blocking or degrading mRNA	linear

Comparison of DNA and RNA

Figure 9-1

Except for asexually reproducing organisms and identical twins, the DNA of every individual is different. These differences, resulting from variations in the sequence of nucleotides, generate different RNA, which produces different proteins.

Early Experiments

Experiments during the first half of the twentieth century lead to the identification of DNA as the hereditary material and that the three-dimensional shape of a DNA molecule is a double helix. Four important research efforts are summarized here:

1. **Griffith discovers that genetic information can be transferred from dead bacteria to living bacteria.** Microbiologist Frederick Griffith experimented with two strains of a bacterium—one that produces a polysaccharide coat and causes pneumonia, and a mutant strain, without the coat, that does not cause pneumonia. Griffith killed the disease-causing bacteria with heat and showed that they could no longer cause pneumonia in mice. He then injected into mice both the dead bacteria that once could cause disease and live bacteria that could not cause disease. These mice died, and Griffith found live bacteria in the mice that had polysaccharide coats. The descendants of these bacteria also had polysaccharide coats and would cause disease. Griffith concluded that genetic information from the dead bacteria with polysaccharide coats *transformed* the bacteria without coats, giving them the ability to make coats and cause disease. Today, the term **transformation** is used to describe the ability of bacteria to absorb and express genetic information (now known to be DNA) obtained from their surroundings.
2. **Avery, MacLeod, and McCarty identify DNA as the heredity information of a cell.** Using the same bacteria used by Griffith, bacteriologists Oswald Avery, Colin MacLeod, and Maclyn McCarty removed the proteins and polysaccharide coats from the dead, disease-causing bacteria. They found that the remaining material was still able to transform bacteria, giving previously harmless bacteria the ability to cause disease. Further tests confirmed that the transforming material was not RNA, but a substance with the same properties as DNA.
3. **The Hershey and Chase experiments establish that DNA was the genetic material of phages.** At the time of these experiments, it was known that **phages,** viruses that infect bacteria, consisted of DNA and protein. In the first part of their experiment, geneticists Alfred Hershey and Martha Chase substituted radioactive sulfur for the sulfur in the amino acids of the phage proteins and mixed these phages with *E. coli* bacteria. After separating the bacteria from the growing medium using a centrifuge, they found the culture media, not the bacteria, were radioactive, indicating that the phage proteins did not enter the bacteria. In the second part of the experiment, they substituted radioactive phosphorous for the phosphorous in the phage DNA. Following the same procedure as in the first part of the experiment, they found that the bacteria, not the growing media, were radioactive, indicating that the phage DNA had entered the bacteria. In a follow-up experiment, the researchers found that infected radioactive bacteria released new phages that were also radioactive. Hershey and Chase concluded that the radioactive DNA from the phages provided the genetic information needed to make new viruses.
4. **Watson, Crick, Wilkins, and Franklin determine the structure of DNA.** Using DNA prepared in the lab of biophysicist Maurice Wilkins, chemist Rosalind Franklin produced an X-ray diffraction photograph of DNA. X-ray diffraction creates a black-and-white pattern of spots that reveal certain structural characteristics of crystals. For DNA, the pattern revealed that the molecule consisted of two strands wrapped around each other (a double helix). Franklin also proposed that sugar-phosphate material formed the outside of the double helix because of its hydrophilic properties, while the hydrophobic nitrogenous bases were located on the inside of the molecule. Using that information, molecular biologists James Watson and Francis Crick proposed a model of DNA resembling a twisted ladder, where the vertical sides of the ladder are sugar-phosphate molecules and its horizontal rungs are pairs of nitrogen bases in which adenine pairs with thymine and guanine pairs with cytosine.

Later research elucidated the flow of information from genes, the hereditary units of DNA, to traits. In the **one-gene-one-enzyme hypothesis,** the gene was defined as the segment of DNA that codes for a particular enzyme. But because many genes code for polypeptides that are not enzymes (such as structural proteins, regulatory proteins, or individual components of enzymes), the gene has since been redefined as the DNA segment that

codes for a particular polypeptide **(one-gene-one-polypeptide hypothesis).** In Chapter 8, "Heredity," the terms **gene** and **genotype** are used to represent the genetic information for a particular trait. From the molecular viewpoint presented in this chapter, *traits are the end products of metabolic processes regulated by enzymes.*

DNA Replication

During interphase of the cell cycle, a second chromatid containing a copy of the DNA molecule is assembled. The process, called **DNA replication,** involves separating (unzipping) the double-stranded DNA molecule into two strands, each of which serves as a template to assemble a new, complementary strand. The result is two identical double-stranded molecules of DNA. Because each of these double-stranded molecules of DNA consists of a single strand of old DNA (the template strand) and a single strand of new, replicated DNA (the complementary strand), the process is called **semiconservative replication.**

During DNA replication, the enzyme **helicase** unwinds the DNA helix, forming a Y-shaped **replication fork. Single-strand binding proteins** attach to each strand of the uncoiled DNA to keep them separate. As helicase unwinds the DNA, it forces the double-helix in front of it to twist. A group of enzymes, called **topoisomerases,** break and rejoin the double helix, allowing the twists to unravel and preventing the formation of knots.

Since a DNA double-helix molecule consists of two opposing DNA strands, the uncoiled DNA consists of a 3′ → 5′ template strand and a 5′ → 3′ template strand. The enzyme that assembles the new DNA strand, DNA polymerase, moves in the 3′ → 5′ direction along each template strand. A new (complement) strand grows in the **antiparallel,** 5′ → 3′ direction.

For the 3′ → 5′ template strand, replication occurs continuously as the DNA polymerase follows the replication fork, assembling a 5′ → 3′ complementary strand. This complementary strand is called the **leading strand.**

For the 5′ → 3′ template strand, however, the DNA polymerase moves away from the uncoiling replication fork. This is because it can assemble nucleotides only as it travels in the 3′ → 5′ direction. As the helix is uncoiled, DNA polymerase assembles short segments of nucleotides along the template strand in the direction away from the replication fork. After each complement segment is assembled, the DNA polymerase must return back to the replication fork to begin assembling the next segment. These short segments of complementary DNA are called **Okazaki fragments.** The Okazaki fragments are connected by **DNA ligase,** producing a single complementary strand. Because this complementary strand requires more time to assemble than the leading strand, it is called the **lagging strand.**

DNA polymerase is able to attach nucleotides only to an already existing complementary strand. Therefore, to initiate a new complementary strand, another enzyme, **primase,** begins replication with a short segment of RNA (not DNA) nucleotides, called an **RNA primer.** The leading strand and every Okazaki fragment on the lagging strand must begin with an RNA primer. When the primer is in place, DNA polymerase can attach succeeding DNA nucleotides to the primer. The RNA nucleotides of the RNA primer are later replaced with DNA nucleotides by DNA polymerase.

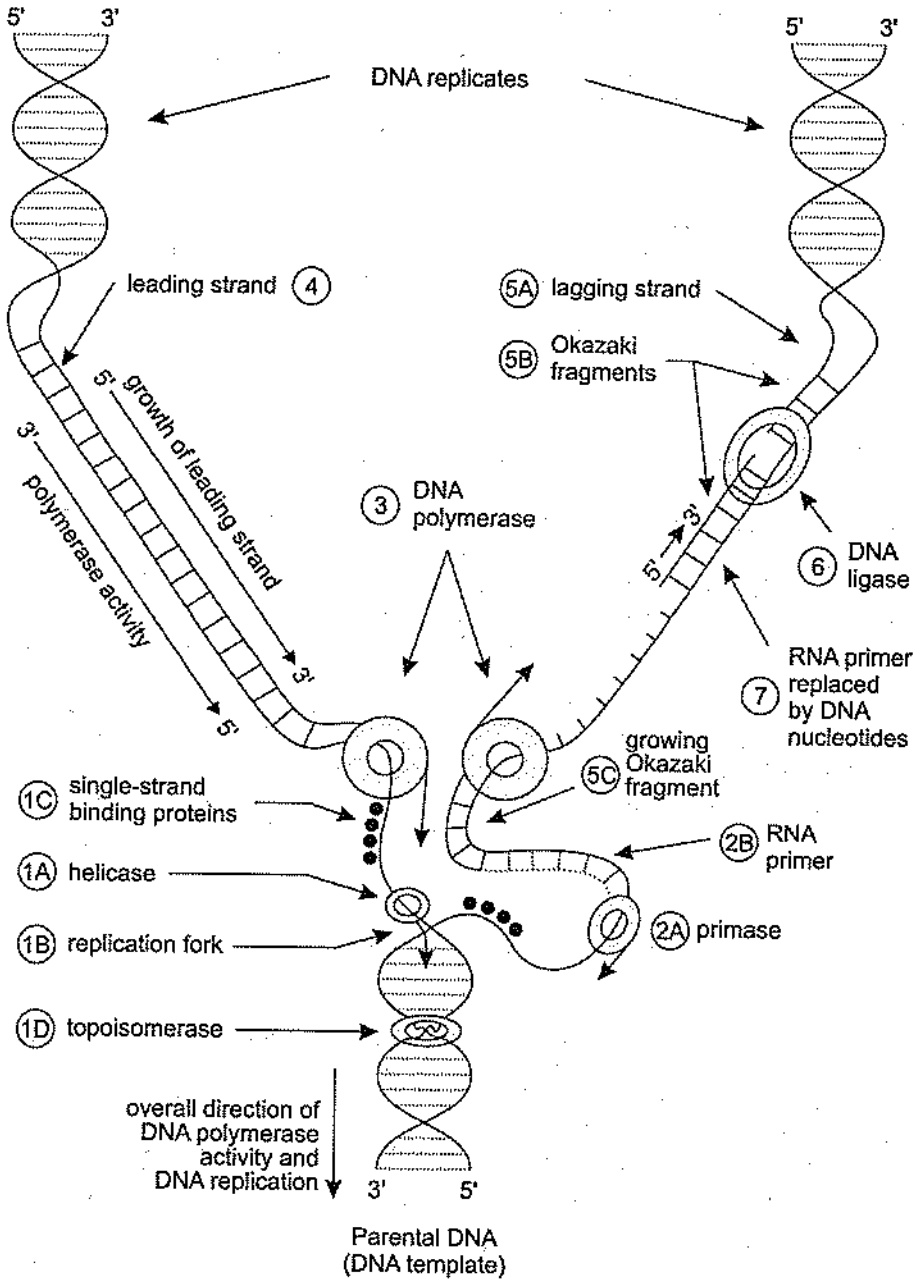
Figure 9-2 illustrates the growth of leading and lagging DNA complements. In the figure, the RNA primer that initiated the leading strand is not shown because it was replaced with DNA nucleotides earlier in its synthesis. The growing Okazaki fragment on the lagging strand, however, still has its RNA primer attached, because a primer must initiate each new fragment.

The details of DNA replication are summarized here. Numbers correspond to events illustrated in Figure 9-2:

1. **Helicase** unwinds the DNA, producing a **replication fork** (1A, 1B). **Single-strand binding proteins** prevent the single strands of DNA from recombining (1C). **Topoisomerase** removes twists and knots that form in the double-stranded template as a result of the unwinding induced by helicase (1D).
2. **Primase** (2A) initiates DNA replication at special nucleotide sequences, called **origins of replication,** with short segments of RNA nucleotides, called **RNA primers** (2B).
3. **DNA polymerase** attaches to the RNA primers and begins **elongation,** the adding of DNA nucleotides to the complementary strand.
4. The **leading complementary strand** is assembled continuously toward the replication fork as the double-helix DNA uncoils.

5. The **lagging complementary strand** (5A) is assembled away from the replication fork in multiple, short **Okazaki fragments** (5B). Each new Okazaki fragment begins when DNA polymerase attaches to an RNA primer (5C).
6. The Okazaki fragments are joined by **DNA ligase.**
7. The RNA primers are replaced with DNA nucleotides.

Energy for elongation is provided by two additional phosphates that are attached to each new nucleotide (making a total of three phosphates attached to the nitrogen base). Breaking the bonds holding the two extra phosphates provides the chemical energy for the process.



DNA Replication

Figure 9-2

DNA replication in prokaryotic and eukaryotic organisms is basically the same, with the following differences:

1. **Chromosome structure.** A prokaryotic chromosome is circular. Eukaryotic chromosomes are linear with ends called **telomeres.**
2. **Origins of replications.** A prokaryotic chromosome has one unique origin of replication. Eukaryotes have multiple origins to accommodate the much larger size of their chromosomes.

DNA Repair

DNA replication is not perfect. Errors happen. Various mechanisms, however, are present to repair the errors, including the following:

1. **Proofreading** of a newly attached base to the growing replicate strand is carried out by DNA polymerase. DNA polymerase checks to make sure that each newly added nucleotide correctly base-pairs with the template strand. If it does not, the nucleotide is removed and replaced with the correct nucleotide.
2. **Mismatch repair proteins** repair errors that escape the proofreading ability of DNA polymerase.
3. **Excision repair proteins** identify and remove damaged nucleotides caused by environmental factors, such as toxins or radiation (UV, X-rays). A polymerase then uses the undamaged complementary strand as a template to repair the damage.

Protein Synthesis

The DNA in chromosomes contains genetic instructions that regulate development, growth, and the metabolic activities of cells. The DNA instructions determine whether a cell will be that of a pea plant, a human, or some other organism, as well as establish specific characteristics of the cell in that organism. For example, the DNA in a cell may establish that it is a human cell. If, during development, it becomes a cell in the iris of an eye, the DNA will direct other information appropriate for its location in the organism, such as the concentration of melanin pigmentation (which influences the appearance of different colors). DNA controls the cell in this manner because it contains codes for polypeptides. Many polypeptides are enzymes that regulate chemical reactions, and these chemical reactions influence the resulting characteristics of the cell.

The process that describes how enzymes and other proteins are made from DNA is called **protein synthesis.** The three steps in protein synthesis are **transcription, RNA processing,** and **translation.** In transcription, RNA molecules are created by using one strand of the DNA molecule as a template. After transcription, RNA processing modifies the RNA molecule with deletions and additions. In translation, the processed RNA molecules are used to assemble amino acids into a polypeptide.

There are three kinds of RNA molecules necessary for protein synthesis that are produced during transcription, as follows:

1. **Messenger RNA (mRNA)** is a single strand of RNA that provides the template used for sequencing amino acids into a polypeptide. A triplet group of three adjacent nucleotides on the mRNA, called a **codon,** codes for one specific amino acid. Since there are 64 possible ways that four nucleotides can be arranged in triplet combinations ($4 \times 4 \times 4 = 64$), there are 64 possible codons. However, there are only 20 amino acids, and thus, some codons code for the same amino acid. The **genetic code,** given in Figure 9-3, provides the decoding for each codon. That is, it identifies the amino acid specified by each of the possible 64 codon combinations. For example, the codon composed of the three nucleotides cytosine-guanine-adenine (CGA) codes for the amino acid arginine. This can be found in Figure 9-3 by aligning the C found in the first column with the G in the center part of the table and the A in the column at the far right. Note that three of the codons in the genetic code are stop codons. They signal an end to translation rather than code for an amino acid. Therefore, only 61 of the codons actually code for amino acids. The codon that codes for the amino acid methionine is also the codon that signals the beginning of translation.

   Note: You do not need to memorize the genetic code (Figure 9-3) for the AP exam.

2. **Transfer RNA (tRNA)** is a short RNA molecule (consisting of about 80 nucleotides) that is used for transporting amino acids to their proper place on the mRNA. Interactions among various parts of the tRNA molecule result in base-pairings between nucleotides, folding the tRNA in such a way that it forms a three-dimensional molecule. (In two dimensions, a tRNA resembles the three leaflets of a clover leaf; in three dimensions, it resembles an upside-down L.) One end of the tRNA attaches to an amino acid. Another portion of the tRNA, specified by a triplet combination of nucleotides, is the **anticodon.** During translation, the anticodon of the tRNA base-pairs with the codon of the mRNA. Exact base-pairing between the third nucleotide of the tRNA anticodon and the third nucleotide of the mRNA codon is often not required. This relaxed base-pairing requirement, called **wobble pairing,** allows the anticodon of some tRNAs to base-pair with more than one kind of codon. As a result, about 45 different tRNAs base-pair with the 61 codons that code for amino acids.
3. **Ribosomal RNA (rRNA)** molecules combine with various proteins to form ribosomes. The rRNA molecules are transcribed in the nucleolus and assembled with proteins imported from the cytoplasm to form a large and a small ribosome subunit. In the cytoplasm, the two subunits join to form a ribosome that coordinates the activities of the mRNA and tRNA during translation.

First Letter ↓	Second Letter U	Second Letter C	Second Letter A	Second Letter G	Third Letter ↓
U	phenylalanine	serine	tyrosine	cysteine	U
	phenylalanine	serine	tyrosine	cysteine	C
	leucine	serine	STOP	STOP	A
	leucine	serine	STOP	tryptophan	G
C	leucine	proline	histidine	arginine	U
	leucine	proline	histidine	arginine	C
	leucine	proline	glutamine	arginine	A
	leucine	proline	glutamine	arginine	G
A	isoleucine	threonine	asparagine	serine	U
	isoleucine	threonine	asparagine	serine	C
	isoleucine	threonine	lysine	arginine	A
	methionine and START	threonine	lysine	arginine	G
G	valine	alanine	aspartate	glycine	U
	valine	alanine	aspartate	glycine	C
	valine	alanine	glutamate	glycine	A
	valine	alanine	glutamate	glycine	G

The Genetic Code
Figure 9-3

Transcription

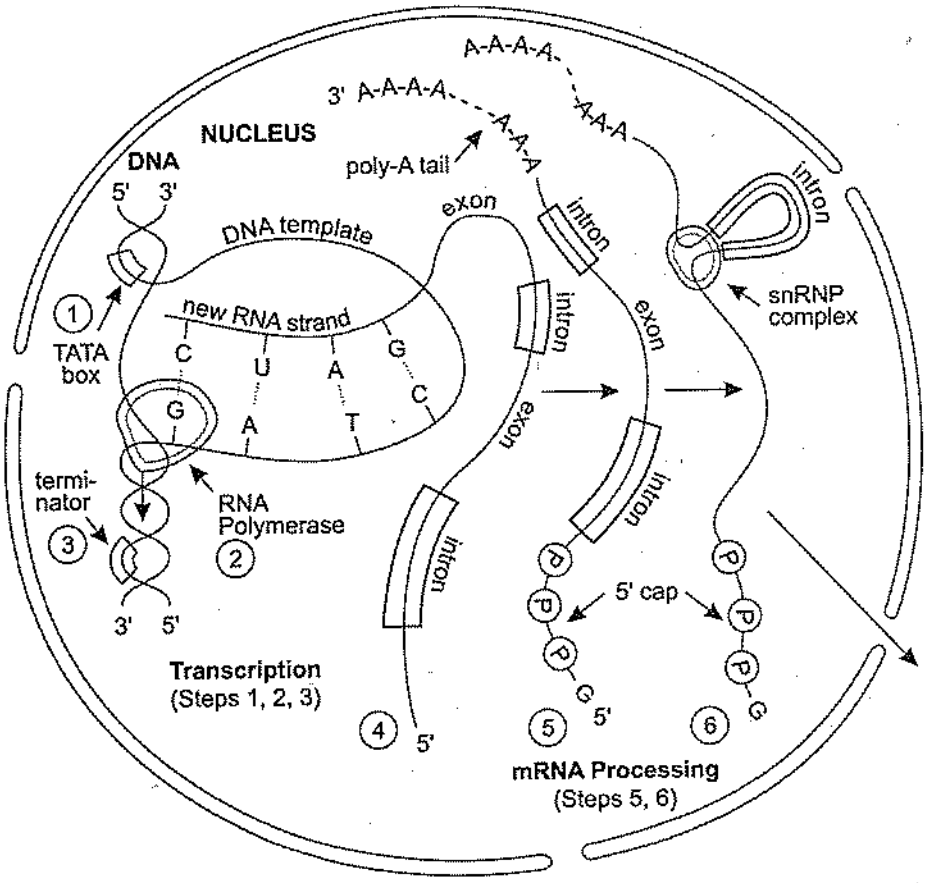
Transcription begins with **initiation,** continues with **elongation,** and ends with **termination.** The details follow, with numbers corresponding to events illustrated in Figure 9-4:

1. In **initiation,** the RNA polymerase attaches to a **promoter region** on the DNA and begins to unzip the DNA into two strands. A promoter region for mRNA transcriptions often contains the sequence T–A–T–A (called the TATA box).
2. **Elongation** occurs as the RNA polymerase unzips the DNA and assembles RNA nucleotides using one strand of the DNA as a template. As in DNA replication, elongation of the RNA molecule occurs in the 5′ → 3′ direction. In contrast to DNA replication, new nucleotides are RNA nucleotides (rather than DNA nucleotides), only one DNA strand is transcribed, and primers are not required.
3. **Termination** occurs when the RNA polymerase reaches a special sequence of nucleotides that serve as a termination point.

mRNA Processing

Before the mRNA molecule leaves the nucleus, it undergoes the following alterations:

1. A **5′ GTP cap** (–P–P–P–G-5′) is added to the 5′ end of the mRNA (see 5 in Figure 9-4). The 5′ GTP cap is a guanine nucleotide with two additional phosphate groups, forming GTP (in the same way that ATP is an adenine nucleotide with two additional phosphates). Capping provides stability to the mRNA and a point of attachment for the small subunit of the ribosome.



Transcription and mRNA Processing

Figure 9-4

2. A **poly-A tail** (–A–A–A . . . A–A-3′) is attached to the 3′ end of the mRNA (see 5 in Figure 9-4). The tail consists of about 200 adenine nucleotides. It provides stability to the mRNA and also appears to control the movement of the mRNA across the nuclear envelope.
3. **RNA splicing** removes nucleotide segments from mRNA. A transcribed DNA segment contains two kinds of sequences—**exons,** which are sequences that *ex*press a code for a polypeptide, and **introns,** *in*tervening sequences that are noncoding. The original unprocessed mRNA transcript contains both the coding and the noncoding sequences (see 4 in Figure 9-4). Before the mRNA moves to the cytoplasm, **small nuclear ribonucleoproteins,** or **snRNPs** (pronounced "snurps"), delete the introns and splice the exons (see 6 in Figure 9-4). A snRNP consists of **small nuclear RNAs (snRNAs)** and various proteins.
4. **Alternative splicing** allows different mRNAs to be generated from the same RNA transcript. By selectively removing different parts of an RNA transcript, different mRNAs can be produced, each coding for a different protein product.

Translation

After transcription, the mRNA, tRNA, and ribosomal subunits are transported across the nuclear envelope and into the cytoplasm. Once in the cytoplasm, a specific amino acid attaches to each of the tRNAs using energy from ATP.

As an introduction, this is what happens during translation:

1. The mRNA attaches to the ribosome.
2. The sequence of codons on the mRNA determines the sequence of amino acids in the polypeptide to be synthesized.
3. One by one, a tRNA brings an amino acid to the ribosome such that the anticodon of the tRNA base-pairs with the codon of the mRNA.
4. The newly arrived amino acid is attached with a peptide bond to other amino acids already present.
5. A tRNA is released from the ribosome.
6. The process is repeated until a "stop" codon on the mRNA has been reached and the completed polypeptide is released.

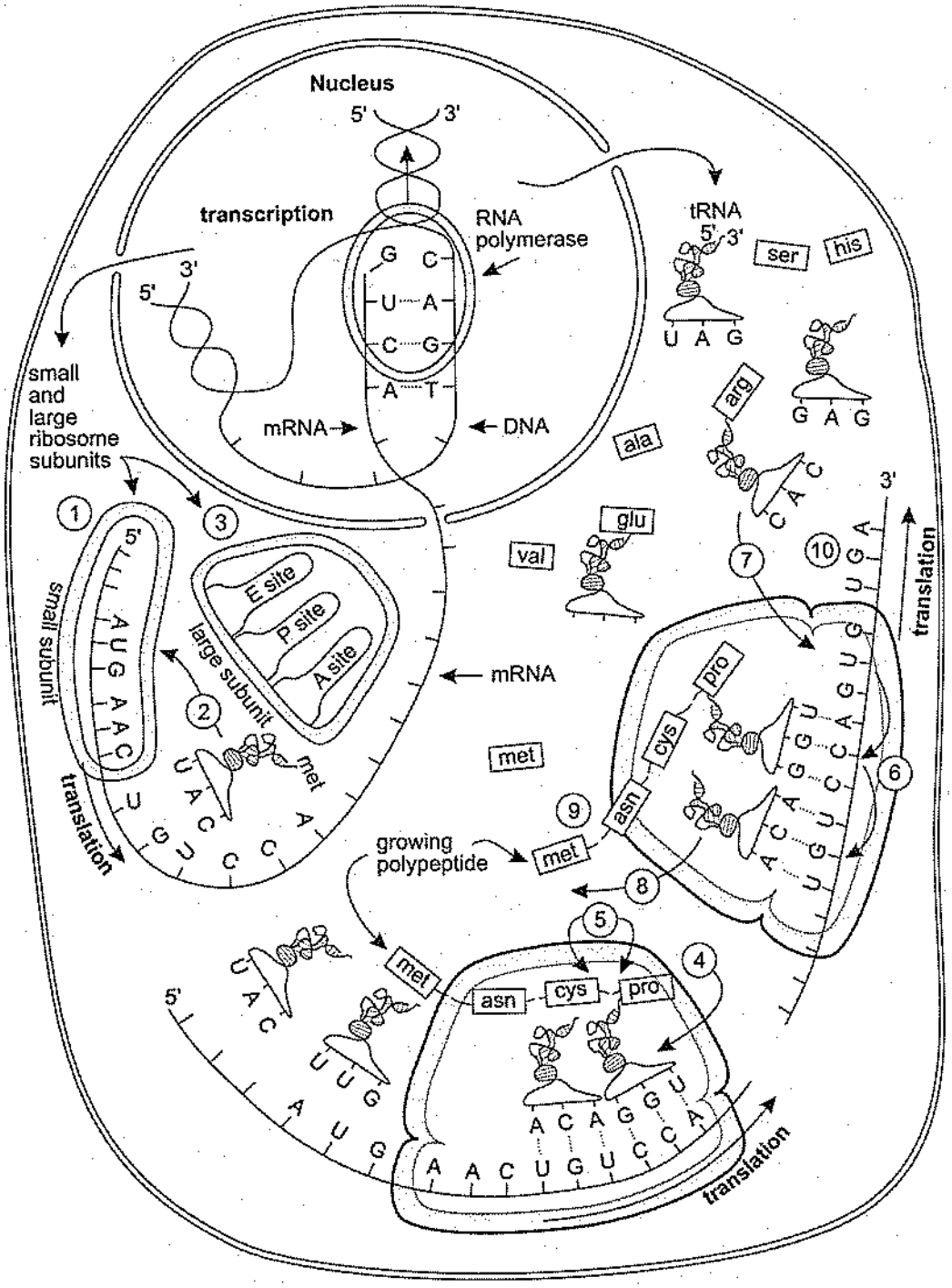
The details are slightly more complicated. As a prelude to the details, note the following:

1. As in transcription, translation is categorized into three steps—initiation, elongation, and termination.
2. Energy for translation is provided by several GTP (guanosine triphosphate) molecules. GTP acts as an energy supplier in the same manner as ATP.
3. A ribosome has three binding sites for tRNA molecules arranged in a series:
   - The A site (for *a*mino acid or "*a*cceptor"), in the *first* position, accepts an incoming tRNA carrying an amino acid. The amino acid is passed on to the tRNA in the second position.
   - The P site (for *p*olypeptide), in the *second* position, holds the tRNA with a growing chain of amino acids (polypeptide).
   - The E site (for *e*xit), in the *third* position, holds the tRNA after it gives up its amino acid.

The details of translation follow, with numbers corresponding to events illustrated in Figure 9-5:

1. **Initiation** begins when the small ribosomal subunit attaches near the beginning of the mRNA.
2. A tRNA (with anticodon UAC) carrying the amino acid methionine attaches to the mRNA at the start codon AUG. (You can remember that the start codon is AUG because school often starts in *Aug*ust.)
3. The large ribosomal subunit attaches to the mRNA with the tRNA (bearing a methionine), occupying the middle of three binding sites (the P site). The ribosome is now completely assembled with the mRNA and one tRNA.

4. **Elongation** occurs as additional tRNAs arrive bearing their amino acids. A newly arriving tRNA attaches to the first binding site (the A site), always with the anticodon of the tRNA appropriately base-pairing with the codon of the mRNA.
5. The amino acid on the tRNA in the second binding site is transferred to the amino acid on the newly arrived tRNA in the first binding site. In Figure 9-5, several amino acids of the growing polypeptide chain are transferred because the figure shows elongation after several tRNAs have delivered amino acids.



Protein Synthesis

Figure 9-5

6. Translocation occurs as the ribosome moves over one binding site: The tRNA in the middle moves to the third position (the E site), and the tRNA in the first position (bearing the amino acids) moves to the second position.
7. This leaves the first position binding site empty, allowing for the arrival of a new tRNA.
8. Meanwhile, the tRNA in the third position is released, now free to bind again with its specific amino acid and provide another delivery to the mRNA.
9. Elongation continues as each successive tRNA delivers an amino acid. As each new tRNA arrives, the polypeptide chain is elongated by one new amino acid, growing in sequence and length as prescribed by the sequence of codons on the mRNA.
10. **Termination** occurs when the ribosome encounters one of the three stop codons. At termination, the completed polypeptide, the last tRNA, and the two ribosomal subunits are released. The ribosomal subunits can now attach to the same or another mRNA and repeat the process.

Once the polypeptide is completed, interactions among the amino acids give it its secondary and tertiary structures. Subsequent processing by the endoplasmic reticulum or a Golgi body may make final modifications before the protein attains its final functional structure.

Mutations

A **mutation** is any sequence of nucleotides in a DNA molecule that does not exactly match the original DNA molecule from which it was copied. Mutations can occur as a result of replication errors, or they may result from environmental effects such as radiation (for example, ultraviolet or X-ray) or reactive chemicals. Radiation or chemicals that cause mutations are called **mutagens. Carcinogens** are mutagens that activate uncontrolled cell growth (cancer).

As reviewed earlier in this chapter, there are various mechanisms for repairing damaged DNA. Damaged DNA that is not repaired becomes a mutation. Although some mutations have no phenotypic effect, most are deleterious—that is, they lead to a partial or complete loss of cell functionality. A mutation that occurs in a sex cell is passed on to the next generation and introduces new allelic variation into the population and the potential for evolutionary change.

There are various kinds of mutations. A **point mutation** is a single nucleotide error and includes the following:

1. A **substitution** occurs when the DNA sequence contains an incorrect nucleotide in place of the correct nucleotide.
2. A **deletion** occurs when a nucleotide is omitted from the nucleotide sequence.
3. An **insertion** occurs when a nucleotide is added to the nucleotide sequence.
4. A **frameshift** mutation occurs as a result of a nucleotide deletion or insertion. Such mutations cause all subsequent nucleotides to be displaced one position. If a frameshift mutation occurs in a DNA segment whose transcription produces the mRNA, all codons following the transcribed mutation will change.

A point mutation may or may not have a significant phenotypic effect. If the mRNA is produced from a DNA segment that contains a point mutation, one of the following results:

1. A **silent mutation** occurs when the new codon still codes for the *same* amino acid. This occurs most often when the nucleotide substitution results in a change of the last of the three nucleotides in a codon. This relaxed requirement for the nucleotide in the third position is called **wobble pairing.** Consult Figure 9-3 for examples of codons that differ by their third nucleotide but code for the same amino acid.
2. A **missense mutation** occurs when the new codon codes for a *new* amino acid. The effect can be minor, or it may result in the production of a protein that is unable to fold into its proper three-dimensional shape and, therefore, is unable to carry out its normal function. The hemoglobin protein that causes sickle-cell disease is caused by a missense mutation.
3. A **nonsense mutation** occurs when the new codon codes for a stop codon. The hemoglobin protein that causes some forms of thalassemia is caused by a nonsense mutation.

Chromosomal aberrations are changes in the chromosome structure or in the makeup of the genome:

1. **Deletions** occur when segments of a chromosome are lost. Most chromosome deletions result in a significant loss of important DNA and are fatal.
2. **Duplications** occur when segments of a chromosome are repeated. If the duplication occurs within a gene segment, it is likely to cause a frameshift mutation with deleterious consequences. On the other hand, a duplicated gene can have beneficial effects by providing additional gene products for processes that are in high demand. For example, most species, including bacteria, have gene redundancy for the gene that codes for rRNA, a product in high demand. Also, extra copies of a gene provide the opportunity for subsequent mutations to create novel gene variations without interfering with the normal operation of the original gene. Some examples of this follow.
   - **Globin genes.** The great similarities among the various globin chains of hemoglobin suggest that they each evolved from a common gene. In humans, multiple variations of the gene occur on two separate chromosomes.
   - **Antifreeze genes.** In certain arctic fish, glycoproteins in the blood provide resistance to freezing. The genes for these glycoproteins appear to be the result of multiple duplications and divergence of the gene that codes for trypsinogen, an enzyme whose activated form, trypsin, digests proteins in the small intestine. This illustrates how novel genes can originate from gene duplication of genes originally used for a totally different purpose.
3. **Inversions** occur when a DNA segment is reversed. Depending upon where the chromosome breaks occur, the mutation may or may not have a significant effect.
4. **Translocations** occur when segments of the chromosome are deleted or copied and then inserted elsewhere, either within the same chromosome or in another chromosome. For example, one form of Down syndrome occurs when a piece of chromosome 21 is translocated to chromosome 14. Offspring who inherit this chromosome 14, along with the two normal copies of chromosome 21, effectively received three copies of the translocated segment.
5. **Transposons** (transposable elements, mobile genetic elements, or "jumping genes") are naturally occurring mutations. They are DNA segments that insert themselves throughout the genome after copying or deleting themselves from another area.
   - In corn (maize), transposons are responsible for mutant strains whose kernels lack pigmentation or have spotted pigmentation.
   - The human genome has as much as 50% of its DNA derived from transposons, although most transposons appear to be inactively sitting within introns.

Viruses

Viruses are parasites of cells. A typical virus penetrates a cell, commandeers its metabolic machinery, assembles hundreds of new viruses that are copies of itself, and then leaves the cell to infect other cells. In the process, the host cell is usually destroyed.

Viruses are specific for the kinds of cells they will parasitize. Some viruses attack only one kind of cell within a single host species, and others attack similar cells from a range of closely related species. **Bacteriophages,** or **phages,** for example, are viruses that attack only bacteria.

Viruses consist of the following structures:

1. A **nucleic acid,** either RNA or DNA (but not both), contains the hereditary information of the virus. The DNA or RNA may be double-stranded (**dsDNA** or **dsRNA**) or single-stranded (**ssDNA** or **ssRNA**).
2. A **capsid,** or **protein coat,** encloses the nucleic acid.
3. An **envelope** surrounds the capsid of some viruses. Envelopes incorporate phospholipids and proteins obtained from the cell membrane of the host cell.

Replication of viruses occurs by one of the following two cycles:

1. In the **lytic cycle,** a virus penetrates the cell membrane of the host and uses the enzymes of the host to produce viral nucleic acids and viral proteins. The nucleic acids and proteins are then assembled into new viruses, which subsequently erupt from the host cell, destroying the cell in the process. The new viruses then infect other cells, and the process repeats. There are variations of this theme, depending upon whether the nucleic acid of the virus is DNA or RNA, double-stranded or single-stranded. Some important variations follow.
   - For most DNA viruses, the DNA is replicated to generate new viral DNA and the DNA is transcribed to produce viral mRNA. Viral mRNA is then translated to produce viral proteins. The DNA and proteins are assembled into new viruses.
   - For some RNA viruses, the RNA serves as mRNA or as a template to make mRNA. The mRNA is translated to make proteins, and these proteins are assembled with RNA to make new viruses.
2. In the **lysogenic cycle,** the viral DNA is temporarily incorporated into the DNA of the host cell. A virus in this dormant state is called a **provirus** (or, if a bacteriophage, a **prophage**). The virus remains inactive until some trigger, often an external environmental stimulus (such as radiation or certain chemicals), causes the virus to begin the destructive lytic cycle.

Retroviruses are ssRNA viruses that use an enzyme called **reverse transcriptase** to make a DNA complement of their RNA. The DNA complement can then be transcribed immediately to manufacture mRNA (lytic cycle), or it can begin the lysogenic cycle by becoming incorporated into the DNA of the host. Because there is often little specificity to where the viral DNA is inserted into the host genome, retroviruses are a special kind of transposon. Human immunodeficiency virus (HIV, the cause of AIDS) is a retrovirus.

Because viruses produce many copies over very short intervals, their potential for rapid evolution is great. RNA viruses, in particular, have much higher rates of replication errors because RNA replication lacks the repair mechanisms associated with DNA replication. As a result, mutations in RNA viruses are more frequent than in DNA viruses.

Evolution of viruses is also augmented when the genetic material of several different, but related, viruses recombine when present together in a host cell. There is evidence that two strains of simian immunodeficiency virus (SIV), an HIV-like virus in monkeys, recombined to create HIV.

The high rates of mutation in viruses intensify their pathogenicity because host populations do not evolve immune-system defenses as fast as the viruses evolve. Note that HIV, influenza (flu) viruses, and most of the viruses that cause the common cold are all RNA viruses. These viruses mutate and evolve quickly, producing new strains on a regular (or seasonal) basis and, as a result, remain infectious throughout the human lifetime.

Prokaryotes

Archaea and bacteria are prokaryotes. They do not contain a nucleus, nor do they possess any of the specialized organelles of eukaryotes. The primary genetic material of prokaryotes is a chromosome consisting of a single, circular DNA molecule without the proteins associated with the DNA of eukaryotic chromosomes. A prokaryotic cell reproduces by **binary fission.** In binary fission, the chromosome replicates and the cell divides into two cells, each cell bearing one chromosome. The spindle apparatus, microtubules, and centrioles found in eukaryotic cell divisions are lacking, since in prokaryotes, there is no nucleus to divide.

Prokaryotes may also contain **plasmids,** short, circular dsDNA molecules outside the chromosome. Plasmids carry genes that are beneficial but not normally essential to the survival of the prokaryote. A group of plasmids, called **R plasmids,** provide bacteria with resistance against antibiotics. Plasmids replicate independently of the chromosome. Some plasmids, called **episomes,** can become incorporated into the prokaryotic chromosome.

DNA replication in prokaryotic and eukaryotic organisms is basically the same. But because prokaryotic chromosomes are circular, replication begins at a single, unique origin, progressing in both directions until they meet at the termination site. In contrast, eukaryotes, with much larger chromosomes, have multiple points of origin.

Transcription in prokaryotes is also similar to that of eukaryotes. However, prokaryotic genes do not have introns. But a single mRNA may contain transcripts of multiple genes whose enzyme products function sequentially as part of a metabolic pathway. Such gene clusters are called **operons.** Also, translation is coupled with transcription—that is, the ribosome attaches to one end of the mRNA transcript and begins translation while the mRNA transcript is still being produced at the other end.

Genetic variation is introduced into prokaryotes through **horizontal gene transfer.** Although it occurs among both archaea and bacteria, it is best understood among bacteria. It occurs in the following ways:

1. **Conjugation** is a process of DNA exchange between bacteria. A donor bacterium produces a tube, or **pilus** (plural, **pili**), that connects to a recipient bacterium. Through the pilus, the donor bacterium sends chromosomal or plasmid DNA to the recipient. In some cases, copies of large portions of a donor's chromosome are sent, allowing recombination with the recipient's chromosome. One plasmid, called the **F plasmid,** contains the genes that enable a bacterium to produce pili. When a recipient bacterium receives the F plasmid, it, too, can become a donor cell.
2. **Transduction** occurs when new DNA is introduced into the genome of a bacterium by a virus. When a virus is assembled during a lytic cycle, it is sometimes assembled with some bacterial DNA in place of some of the viral DNA. When this aberrant virus infects another cell, the bacterial DNA that it delivers can recombine with the resident DNA. Like conjugation, transduction can transfer bacterial resistance or other pathogenic traits to host cells.
3. **Transformation** occurs when bacteria absorb DNA from their surroundings and incorporate it into their genome. Specialized proteins on the cell membranes of some bacteria facilitate this kind of DNA uptake.

Regulation of Gene Expression

Although some genes, like those that code for ribosomal proteins, are always turned on, most are not. That is because a cell is constantly modifying biochemical activities to respond to varying internal and external conditions. In order to accomplish this feat, mechanisms exist to turn genes on and off at appropriate times.

Regulatory mechanisms in prokaryotes occur as part of **operons.** An operon is a unit of DNA that contains multiple genes whose products work together to direct a single metabolic pathway. It contains the following components:

1. The **promoter** region is a sequence of DNA to which the RNA polymerase attaches to begin transcription.
2. The **operator** region is engaged by a regulatory protein to either block or promote the action of the RNA polymerase.
3. The **structural genes** contain coding DNA, that is, DNA sequences that code for various related enzymes that direct the production of some particular end product.
4. A **regulatory gene,** lying outside the operon region, produces a regulatory **protein** that engages the operator region and governs whether RNA polymerase can attach to the promoter region and begin transcription. Regulatory proteins are allosteric, that is, they become active (or inactive) only when they bind to some specific substrate molecule. A regulatory protein can be one of two kinds:
   - A **repressor protein** *blocks* the attachment of RNA polymerase to the promoter region. Repressor proteins characterize **negative regulation** because they must be *inactive* in order for transcription to occur.
   - An **activator protein** *promotes* the attachment of RNA polymerase to the promoter region. Activator proteins characterize **positive regulation** because they must be *active* in order for transcription to occur.

Three examples of gene regulation in bacteria follow (summarized in Figure 9-6):

1. The ***trp* operon** in *E. coli* produces enzymes for the synthesis of the amino acid tryptophan. A regulatory gene produces an inactive repressor that does not bind to the operator. As a result, the RNA polymerase proceeds to transcribe the structural genes necessary to produce enzymes that synthesize tryptophan. When tryptophan is available to *E. coli* from the surrounding environment, the bacterium no longer needs to manufacture its own tryptophan. In this case, rising levels of tryptophan induce some tryptophan to react

with the inactive repressor and make it active. Here, tryptophan is acting as a **corepressor.** The active repressor now binds to the operator region, which, in turn, prevents the transcription of the structural genes. Since these structural genes stop producing enzymes only in the presence of an active repressor, they are called **repressible enzymes,** and the operon is a **repressible operon.** Because a repressor protein is involved, it is an example of **negative regulation.**

2. The ***lac* operon** in *E. coli* controls the breakdown of lactose. A regulatory gene produces an active repressor that binds to the operator region. When the operator region is occupied by the repressor, RNA polymerase is unable to transcribe several structural genes that code for enzymes that control the uptake and subsequent breakdown of lactose. When lactose is available, however, some of the lactose (in a converted form) combines with the repressor to make it inactive. When the repressor is inactivated, RNA polymerase is able to transcribe the genes that code for the enzymes that break down lactose. Since a substance (lactose, in this case) is required to induce (turn on) the operon, the enzymes that the operon produces are said to be **inducible enzymes,** and the operon is an **inducible operon.** Because a repressor protein is involved, it is an example of **negative regulation.**

Kind of Regulation	Operon Description	Regulatory Protein	Regulatory Protein Status	Regulatory Activity	RNA Polymerase Access to Operator	Action
Negative Regulation	Repressible Operon (trp operon)	repressor	repressor + corepressor (tryptophan)	repressor active	blocked	do not produce tryptophan
			repressor only	repressor inactive	promoted	produce tryptophan
	Inducible Operon (lac operon)	repressor	repressor + inducer (lactose)	repressor inactive	promoted	break down lactose
			repressor only	repressor active	blocked	do not break down lactose
Positive Regulation	Glucose Repression	activator	activator (CAP) only	activator inactive	blocked	do not break down lactose
			activator (CAP) + effector (lactose)	activator active	promoted	break down lactose

Summary of Operon Activity

Figure 9-6

3. **Glucose repression** is a second regulatory process that influences the *lac* operon. Glucose is a preferential source of energy when both glucose and lactose are present. But when only lactose is present, this process enhances the breakdown of lactose (already permitted by the inactive repressor of the *lac* operon). This is accomplished by an **activator** regulatory protein, CAP, that is activated by cyclic AMP (cAMP). When glucose levels are up, cAMP levels are down, and the CAP activator is inactive. But when glucose is absent, cAMP levels are up, CAP is activated and binds to the operator, promoting RNA polymerase transcription of the enzymes that break down lactose. Because an activator protein is involved (CAP), this is an example of **positive regulation.**

In general, repressible operons are associated with genes that regulate *anabolic* biochemical pathways (pathways that consume energy to synthesize new molecules). Inducible operons are associated with *catabolic* pathways (pathways that release energy when they break down molecules). It should also be noted that these regulatory processes are negative feedback mechanisms: They turn on remedial processes in response to changes in environmental conditions (too much or too little tryptophan or lactose, for example) and turn off the processes when suitable conditions return.

The regulation of gene expression in eukaryotic cells is more complicated than in prokaryotes. There are several reasons for this:

1. **Multicellularity.** Many eukaryotic organisms are multicellular. This requires different gene regulation programs for different cell types.
2. **Chromosome complexity.** The chromosomes of eukaryotic organisms are more complex than those of prokaryotes due to their much larger size and elaborate organization with histone proteins. Also, some metabolic processes require the activation of multiple genes, each of which is located on a different chromosome. In these cases, coordinated expression of these genes requires a more sophisticated system of regulation than that which is present in prokaryotes.
3. **Uncoupling of transcription and translation.** In prokaryotes, translation begins while transcription is still in progress. In contrast, eukaryotic transcription occurs in the nucleus isolated from translation, which takes place in the cytoplasm. This allows for a greater range of mechanisms to control gene expression.

In line with the complexity of eukaryotic cells, the mechanisms that regulate eukaryotic transcription are similarly complex. Every stage that contributes to the final protein product, from accessing the gene in the chromatin to the final folding of the translated protein, can be subject to some kind of regulation. Following are the more prominent gene-regulating mechanisms that occur:

1. **DNA methylation** occurs when methyl groups ($-CH_3$) attach to DNA bases. This makes it more difficult for transcription factors to access the DNA. DNA methylation seems to be associated with long-term inactivation of genes.
2. **Histone modification** refers to changes in the organization of histone proteins with DNA. The DNA double helix in chromatin is wrapped around a bundle of eight histone molecules to form tightly knit complexes called **nucleosomes.** Access to DNA for transcription can be increased or decreased by the following:
   - **Acetylation.** Histone molecules loosen their grip on the DNA molecule when they are acetylated— that is, when an acetyl group ($-COCH_3$) is attached. Acetylated histones are associated with *activated* transcription.
   - **Methylation.** Histones are methylated when a methyl group ($-CH_3$) is attached. In most cases, methylated histones are associated with *repressed* transcription.
3. **X inactivation** is a special case of chromosome inactivation in female mammals. A few days after fertilization, one of the two X sex chromosomes in each cell of female embryos is randomly inactivated. A gene on the chromosome that is silenced produces a noncoding RNA transcript that is associated with a loss of acetylation in the histone proteins of nucleosomes. The descendants of each cell maintain the same inactivated X chromosome. The purpose of X inactivation is to equalize the gene dosage that both males (who have only one X chromosome) and females express. (See "X-Inactivation" in Chapter 8, "Heredity.")
4. **Transcription initiation** is regulated by a **transcription complex,** a group of various proteins that are associated with RNA polymerase activity. The makeup of the transcription complex determines the degree to which transcription is activated or repressed. There are several components to the complex:
   - **General transcription factors** are proteins that are required by *all* transcription events to successfully initiate transcription by RNA polymerase. General transcription factors attach with the RNA

polymerase to the *promoter region* upstream and adjacent to the gene to be transcribed. Some general transcription factors target the TATA box sequence associated with the promoter region.

- **Specific transcription factors** are additional proteins associated with regulating *specific* transcription activities—specific to cell type, specific to the particular genes, or specific to the timing of the transcription. There are two kinds of specific transcription factors: activators and repressors. Specific transcription factors attach to **enhancers,** DNA binding sites that can be thousands of nucleotides upstream or downstream from the gene. There may be one or more enhancers that are unique to a particular gene, and each of those enhancers may be specific to a different timing of transcription or to a specific cell type. Because an enhancer may be quite a distance away from the gene it influences, the DNA segment containing the enhancer (and bearing its *specific* transcription factor) folds such that it can join the *general* transcription factors and RNA polymerase on the promoter.
- **Coactivators** and **mediators** are additional proteins that contribute to the binding of transcription complex components.

5. **RNA processing,** as discussed earlier in this chapter, can produce different mRNAs by slicing the primary RNA transcript in different ways. This allows a single gene to encode proteins that are specific to the cell type or to its developmental stage.
6. **RNA interference (RNAi)** refers to gene silencing caused by short RNA molecules. They can do this in three ways:
   - They bind to complementary sequences of mRNAs in the cytoplasm and block their translation.
   - They bind to, cleave, and degrade complementary sequences of mRNA.
   - They bind to chromatin in the nucleus, preventing transcription of genes.

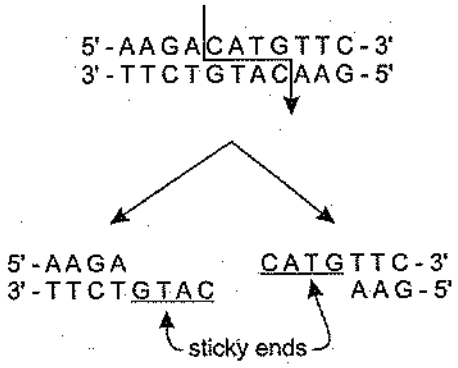
   The two best understood kinds of these molecules are both single-stranded (ssRNA) and about 20 nucleotides long. They differ mostly by their origin as described here:
   - microRNAs (miRNAs) originate from mRNAs that have been transcribed from regulatory genes. The mRNAs are subsequently truncated in the cytoplasm to form miRNAs.
   - Short interfering RNAs (siRNAs) originate from double-stranded RNAs (dsRNAs) that have formed in the cytoplasm from ssRNAs or are dsRNAs that have been introduced into the cell experimentally. In either case, the dsRNAs are subsequently truncated in the cytoplasm to form siRNAs.
7. **mRNA degradation** occurs as a result of RNAi (above), but also because mRNAs are unstable molecules. The poly-A tail and the 5′ GTP cap help maintain mRNA stability on a scale of hours. But degradation of the tail occurs as the mRNA ages and degrading enzymes, targeting the tail and cap, quickly follow. Meanwhile, sequences rich in adenine and uracil in untranslated regions of the mRNA are recognition sites for other degrading enzymes.
8. **Protein degradation** is the final stage in the life of proteins. As proteins age, they lose functionality as their three-dimensional shape changes. Nonfunctional proteins are marked for destruction with the protein **ubiquitin** (so called because it is ubiquitous, present in all eukaryotic cells).

Cells in the early stages of embryonic development are **stem cells.** When stem cells divide, daughter cells have the potential to become any kind of fetal or adult cell. But as development continues, cells differentiate, become specialized, and subsequent cell divisions produce cells that are similarly specialized. Cells become specialized because transcription factors activate some genes while repressing others. The process is often self-reinforcing—once certain genes are turned on or off, they remain so. Genes that are permanently turned off are often associated with DNA methylation and histone modification. The process that fixes a cell's fate is called **cell determination.**

Reproductive cloning is the process of making an individual with the same nuclear DNA as another animal. The first mammal to be cloned was Dolly, a sheep. The process for cloning Dolly and many mammals since was "somatic cell nuclear transfer." In this process, the nucleus of an unfertilized egg cell is replaced by the nucleus from a fully differentiated adult cell (for Dolly, a cell taken from her udder). This demonstrated that cell determination can be reversed, that is, a nucleus from a cell whose fate was fixed was unfixed by implanting it in the cytoplasm of a stem cell, where transcription factors and other regulatory factors were absent.

Biotechnology

Recombinant DNA contains DNA segments or genes from different sources. DNA transferred from one part of a DNA molecule to another, from one chromosome to another chromosome, or from one organism to another, all constitute recombinant DNA. The transfer of DNA segments can occur naturally through viral transduction, bacterial conjugation, or transposons. It is also a regular event in eukaryotes as a result of crossing over during prophase of meiosis. Recombinant DNA can also be produced artificially with biotechnology. **Biotechnology** is the use of biological systems to modify organisms or produce desired products. This definition addresses a range of activities, from selected breeding of plants or animals to manipulating bacteria to produce human insulin.



Restriction Enzyme *PciI* Activity
Figure 9-7

Recombinant DNA technology uses **restriction enzymes (restriction endonucleases)** to cut up DNA. Restriction enzymes are obtained from bacteria that manufacture these enzymes to combat invading viruses. Restriction enzymes are very specific, cutting DNA at specific recognition sequences of nucleotides called **restriction sites** (Figure 9-7). The cut across a double-stranded DNA is usually staggered, producing fragments that have one strand of the DNA extending beyond the complementary strand. The unpaired extension is called a **sticky end.**

DNA Cloning

DNA cloning is a procedure that allows DNA fragments or genes to be copied. The major features of the process follow:

1. **Use a restriction enzyme to cut up the foreign DNA that contains a gene to be copied.** The restriction enzyme produces multiple fragments of the foreign DNA with sticky ends. Some of these fragments will contain the gene to be copied.
2. **Use the same restriction enzyme to cut up the DNA of a cloning vector.** A cloning vector is a vehicle used to transfer DNA between cells. Plasmids are commonly used as vectors because they can subsequently be introduced into bacteria by transformation. Using the same restriction enzyme produces the same sticky ends as that produced in the foreign DNA. To help in the identification of the copied gene, use a plasmid that has one restriction site for the restriction enzyme and is engineered to contain the following:
   - ***amp*^{R} gene.** This gene gives a bacterium resistance against the antibiotic ampicillin.
   - ***GFP* gene.** This gene, originally obtained from jellyfish, presents a bright green fluorescence.
   - ***lacZ* gene.** This gene codes for an enzyme that naturally breaks down lactose. Conveniently, the enzyme also breaks down a related, but artificially made, molecule called **X-gal.** X-gal is colorless, but it forms a blue product when broken down by the *lac*Z enzyme. *The one restriction site for the restriction enzyme occurs within the* lac*Z gene.*
3. **Mix cut foreign DNA with cut plasmids.** As fragments reattach by base-pairing at their sticky ends, foreign fragments, some of which contain the gene to be copied, will fuse with plasmid fragments.
4. **Apply DNA ligase to stabilize attachments.** Some of the plasmids, by chance, are now recombinant plasmids containing the foreign gene. Other plasmids will not contain the foreign gene.
5. **Mix plasmids with bacteria to allow transformation.** Some of the bacteria will absorb the plasmids (transformation). Some bacteria will absorb recombinant plasmids. Others will absorb nonrecombinant plasmids.
6. **Grow the transformed bacteria in the presence of ampicillin and X-gal.** Only bacteria that have absorbed a plasmid (transformed bacteria) will grow in the presence of ampicillin because only the introduced plasmids contain the *amp*^{R} gene. Of those bacteria, only bacteria that have absorbed a *recombinant* plasmid will be white because they lack a functioning *lac*Z gene. For these bacteria, the foreign DNA was inserted within

the *lacZ* gene of the plasmid, making the gene dysfunctional. Bacteria that absorbed a *non*recombinant plasmid have a functioning *lacZ* gene and turn X-gal blue.

The end product of this process is a **genomic library,** a collection of bacteria, each of which contains a fragment of the genome of the foreign DNA but together contain the entire genome of the foreign DNA.

When foreign genes are inserted into the genome of a bacterium, introns often prevent their transcription. To avoid this problem, the DNA fragment bearing the required gene is obtained directly from the mRNA that codes for the desired polypeptide. Reverse transcriptase (obtained from retroviruses) is used to make a DNA molecule directly from the mRNA. DNA obtained in this manner is called **complementary DNA (cDNA)** and comprises cDNA libraries that lack the introns that suppress transcription.

As a result of DNA cloning, the human gene for insulin was inserted into *E. coli.* The transformed *E. coli* produces insulin, which is isolated and used to treat diabetes.

The Polymerase Chain Reaction (PCR)

The **polymerase chain reaction (PCR)** is a technique that makes large numbers of DNA copies faster than the DNA cloning process previously described. Steps of the process are outlined here:

1. **DNA is heated.** Heating denatures (separates) the hydrogen bonding holding the double-stranded DNA (dsDNA) together and forms two single-stranded DNA (ssDNA) molecules.
2. **DNA is cooled and single-stranded DNA primers are added.** Two primers are added, each complementary to the 3′ end of the single strands of DNA. These single-stranded DNA primers serve the same purpose as RNA primers during normal DNA replication.
3. **DNA polymerase is added.** A special heat-tolerant DNA polymerase derived from bacteria adapted to living in hot springs is added. The DNA polymerase attaches to the primers at each end of a single strand of DNA and synthesizes a complementary DNA strand. At the end of this step, the one initial dsDNA molecule becomes two dsDNA molecules.
4. **Repeat the above steps.** Each repetition of the sequence doubles the number of DNA molecules, and the total number of molecules increases exponentially. PCR can amplify DNA to billions of copies in hours.

Gel Electrophoresis and DNA Fingerprinting

Gel electrophoresis is a procedure that separates restriction fragments. In this procedure, DNA fragments of different lengths are separated as they diffuse through a gelatinous material under the influence of an electric field. Since DNA is negatively charged (because of the phosphate groups), it moves toward the positive electrode. Shorter fragments migrate further through the gel than longer, heavier fragments. Gel electrophoresis is often used to compare DNA fragments of closely related species in an effort to determine evolutionary relationships.

When restriction fragments between individuals of the same species are compared, the fragments differ in length because of polymorphisms, which are slight differences in DNA sequences. These fragments are called **restriction fragment length polymorphisms (RFLPs).** In **DNA fingerprinting,** RFLPs produced from DNA left at a crime scene are compared to RFLPs from the DNA of suspects. Areas of the human genome that are particularly polymorphic contain **short tandem repeats (STRs).** STRs are short sequences of nucleotides (two to five base pairs) that repeat multiple times, with the number of repeats varying markedly among individuals.

Concerns about Biotechnology

The advent of biotechnology has introduced new capabilities to human civilization. Improvements in the identification and treatment of disease and advances in forensic science are clear benefits. But some benefits can also be abused while the benefits of others are debated. In some cases, social and ethical questions arise. Here are some of the issues that biotechnology presents:

1. **Pharmaceuticals.** DNA cloning allows quick and inexpensive production of many pharmaceuticals. For example, human insulin (previously isolated from animals) and human growth hormone, or HGH

(previously obtained from human cadavers), are now readily available as products of DNA cloning. Insulin is used to treat diabetes, and HGH is used to treat certain forms of dwarfism. But HGH is also used by some athletes to increase athletic performance and by celebrities to supposedly reduce the effects of aging.

2. **Human disease profiles.** Some diseases are inherited and can be identified before symptoms appear by evaluating the genes of the individual. Sometimes, treatments can ensue and symptoms, including early death, can be circumvented. Other times, there are no available treatments. In this case, it is not clear whether affected individuals should (or want to) know that they will suffer from a debilitating disease in the near future, whether medical insurance has a right to know that they are high-risk patients, and whether society should control the reproductive potential of the individual. These are issues of individual self-determination, medical privacy, and reproductive rights.
3. **Transgenic** organisms are those that possess genes that, through genetic engineering, have been taken from other organisms, including other species.
   - **Genetic engineering in plants.** Genes have been inserted into agricultural plants that provide resistance to pests and herbicides and tolerance to drought and other extreme environmental conditions. In other cases, genetic engineering has changed fruit color, increased crop yields, or extended shelf life. For example, several genetically modified (GM) crops, including corn (Bt corn), carry the Bt gene that gives plants insecticide properties. The gene, whose origin is the plasmid of a soil bacterium (*Bacillus thuringiensis*, or Bt), produces a chemical that is toxic only to certain insects. This reduces the need for pesticides, which often kill insects indiscriminately as well as their natural predators. Often, however, other insects not killed by the Bt toxin increase in numbers, requiring, once again, the application of insecticides. In addition, insects once susceptible to Bt toxin evolve resistance to the toxin. Also, in some genetically engineered crops, there are uncertain consequences, should the transgenic gene, through gene flow via pollen, spread to wild plant populations.
   - **Genetic engineering in animals.** Genes have been inserted into domestic animals to produce desirable products or to produce animals that are more vigorous or convenient to rear. Salmon have been modified with a growth hormone gene (from a different salmon species) to make them grow faster. Like plants, a major concern for the transgenic salmon is gene flow into wild populations. Goats have been genetically modified to produce milk that contains spider fibers (for use as sutures or industrial products). Predictably, there are often unexpected results with genetically modified organisms (GMOs). A GM breed of featherless chickens, developed to simplify marketing preparation, was more prone to insect bites and more sensitive to UV radiation. Also, mating success was impaired because courtship requires feather displays.
   - **GMOs in the food chain.** Whether animal or plant, GMOs in the food chain are controversial. A major concern is one of health: Unidentified genes, accidently and unknowingly inserted into the GM plant or animal, may produce products that create allergies.
4. **Reproductive cloning.** Traditional selective breeding of animals, where two animals with the desired traits are bred, is a slow process. Each new generation must reach reproductive age before another round of selective breeding can occur. Reproductive cloning, however, promises to produce, effectively within a single generation, copies of any desirable individual. Many identical copies of exceptional individuals could be created using the best available individuals—a cow that produces the most milk, a prize-winning racehorse, or your favorite pet. So far, however, reproductive cloning has had mediocre success. Problems include organ failure, a high susceptibility to disease, shorter-than-normal life spans, and low success rates. (Hundreds of trials are sometimes required before one successful clone is produced.)

Review Questions

Multiple-Choice Questions

The questions that follow provide a review of the material presented in this chapter. Use them to evaluate how well you understand the terms, concepts, and processes presented. Actual AP multiple-choice questions are often more general, covering a broad range of concepts, and often more lengthy. For multiple-choice questions typical of the exam, take the two practice exams in this book.

Directions: Each of the following questions or statements is followed by four possible answers or sentence completions. Choose the one best answer or sentence completion.

1. The two strands of a DNA molecule are connected by

   A. hydrogen bonds between the codons and anticodons
   B. hydrogen bonds between the bases of one strand and the bases of the second strand
   C. covalent bonds between phosphate groups
   D. covalent bonds between the nitrogen bases

2. All of the following combinations of nucleotides are examples of normal base-pairing EXCEPT:

   A. an adenine DNA nucleotide to a thymine DNA nucleotide
   B. a guanine DNA nucleotide to a cytosine DNA nucleotide
   C. a thymine RNA nucleotide to an adenine DNA nucleotide
   D. a uracil RNA nucleotide to an adenine DNA nucleotide

3. Which of the following is true?

   A. Ribosomes contain RNA nucleotides and amino acids.
   B. The uracil nucleotide consists of the uracil nitrogen base, a deoxyribose sugar, and a phosphate group.
   C. When tRNA attaches to mRNA during translation, cytosine nucleotides base-pair with guanine nucleotides and adenine nucleotides base-pair with thymine nucleotides.
   D. In eukaryotes, DNA is manufactured in the nucleus and RNA is manufactured in the cytoplasm.

4. All of the following enzymes are involved in DNA replication EXCEPT:

   A. helicase
   B. DNA ligase
   C. RNA polymerases
   D. primase

5. ATP, the common energy-carrying molecule, most resembles the

   A. adenine DNA nucleotide
   B. adenine RNA nucleotide
   C. adenine DNA nucleotide with two extra phosphates
   D. adenine RNA nucleotide with two extra phosphates

6. The end products of translation are

   A. amino acids
   B. polypeptides
   C. lipids
   D. RNA molecules

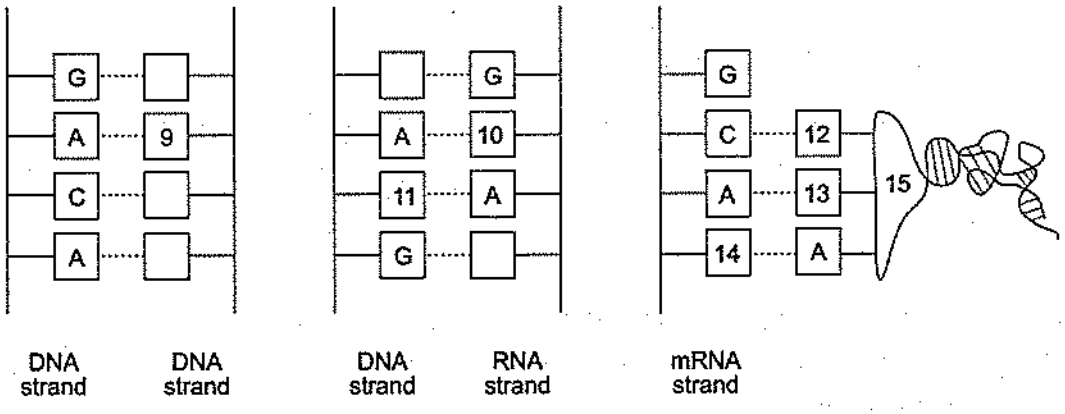
7. Which of the following contains a code for a protein?

- **A.** DNA polymerase
- **B.** RNA polymerase
- **C.** rRNA
- **D.** mRNA

8. Which of the following changes following the start codon in the mRNA would most likely have the greatest deleterious effect?

- **A.** a deletion of a single nucleotide
- **B.** a deletion of a nucleotide triplet
- **C.** a single nucleotide substitution of the nucleotide occupying the first codon position
- **D.** a single nucleotide substitution of the nucleotide occupying the third codon position

Questions 9–15 refer to the following diagram of DNA and RNA segments. Boxes represent nucleotides. The letters A, G, and C refer to the names of the nucleotides that occupy a particular position.



Use the following key for questions 9–14. Each answer in the key may be used once, more than once, or not at all.

- **A.** cytosine nucleotide
- **B.** guanine nucleotide
- **C.** thymine nucleotide
- **D.** uracil nucleotide

9. Which nucleotide would occupy box 9?

10. Which nucleotide would occupy box 10?

11. Which nucleotide would occupy box 11?

12. Which nucleotide would occupy box 12?

13. Which nucleotide would occupy box 13?

14. Which nucleotide would occupy box 14?

15. The nucleotide strand represented by 15 is

- **A.** tRNA
- **B.** rRNA
- **C.** mRNA
- **D.** ATP

16. The DNA of an elephant and the DNA of a cherry tree will most likely differ in all of the following respects EXCEPT:

- **A.** the kinds of genes for which the DNA codes
- **B.** the kinds of nucleotides utilized in forming DNA
- **C.** the number of DNA molecules
- **D.** the length of DNA molecules

17. Protein synthesis consists of all of the following steps EXCEPT:

- **A.** replication
- **B.** transcription
- **C.** translation
- **D.** elongation

18. The genetic instructions for forming a polypeptide chain are carried to the ribosome by the

- **A.** tRNA
- **B.** rRNA
- **C.** mRNA
- **D.** DNA

19. In bacteria, a small circle of DNA found outside the main chromosome is called a

- **A.** plasmid
- **B.** cDNA
- **C.** RFLP
- **D.** PCR

20. Genetic variation can be introduced into bacteria by all of the following methods EXCEPT:

- **A.** transfer of DNA between bacteria through pili
- **B.** DNA amplification
- **C.** mutation
- **D.** transformation

21. All viruses consist of

- **A.** DNA and a protein coat
- **B.** RNA and a protein coat
- **C.** a nucleic acid and a protein coat
- **D.** proteins and polysaccharides

22. The mRNA actively being translated in the cytoplasm would have all of the following EXCEPT:

- **A.** a poly-A tail
- **B.** a 5′ GTP cap
- **C.** exons
- **D.** introns

23. The *lac* operon in *E. coli* is involved in

- **A.** regulating the expression of a gene
- **B.** regulating the translation of mRNA
- **C.** controlling the formation of ribosomes
- **D.** controlling DNA replication

Free-Response Questions

The AP exam has long and short free-response questions. The long questions have considerable descriptive information that may include tables, graphs, or figures. The short questions are brief but may also include figures. Both kinds of questions have four parts and generally require that you bring together concepts from multiple areas of biology.

The questions that follow are designed to further your understanding of the concepts presented in this chapter. Unlike the free-response questions on the exam, they are narrowly focused on the material in this chapter. For free-response questions typical of the exam, take the two practice exams in this book.

Directions: The best way to prepare for the AP exam is to write out your answers as if you were taking the exam. Use complete sentences for all your answers and do *not* use outline form or bullets. You may use diagrams to supplement your answers, but be sure to describe the importance or relevance of your diagrams.

---

1. Gene regulation in eukaryotic cells is considerably more complicated than gene regulation in prokaryotes. Describe two reasons eukaryotic organisms require a more complex approach to gene regulation.
2. In some cases, a single nucleotide mutation does not lead to the creation of a different protein. Explain how this can happen.
3. Traditional fingerprints and DNA fingerprints are both used to identify suspected criminals. Can either of these two techniques distinguish identical twins? Justify your answer.
4. Genes store the hereditary information of the cell. Describe how each of the following contributes to the process of transforming the information stored in a gene into the expression of a physical trait.
   a. transcription
   b. RNA processing
   c. translation
   d. protein folding
5. Both of the following cause disease. Describe how their disease-causing activities differ.
   a. viruses
   b. bacteria
6. Describe how gene regulation occurs in
   a. bacterial cells
   b. eukaryotic cells

Answers and Explanations

Multiple-Choice Questions

1. **B.** Weak hydrogen bonds form between bases of the two strands. In particular, a pyrimidine (a base with one nitrogen ring) in one strand bonds to a purine (a base with two nitrogen rings) in the second strand.
2. **C.** Thymine is not used as a base in any RNA nucleotide. Instead, RNA uses the uracil base, which base-pairs with the adenine DNA nucleotide during transcription.
3. **A.** Ribosome is a complex of rRNA molecules (RNA nucleotides) and proteins (amino acids). The remaining answer choices are incorrect because the uracil nucleotide contains a ribose sugar (not a deoxyribose sugar), adenine nucleotides base-pair with uracil nucleotides, and both DNA and RNA are produced in the nucleus.
4. **C.** RNA polymerases are involved in the transcription of DNA into RNA, not the replication of DNA. The major enzyme for DNA replication is DNA polymerase (not one of the answer choices). Primase directs the attachment of RNA nucleotides to a DNA template, but those RNA primers must be in place before DNA polymerase can attach and begin attaching DNA nucleotides.
5. **D.** Since ATP contains the adenine nitrogen base, the sugar ribose (not deoxyribose), and three phosphate groups, it is equivalent to the adenine RNA nucleotide with two extra phosphate groups. In contrast, an adenine DNA nucleotide with two extra phosphates contains a deoxyribose sugar and is written as dATP.
6. **B.** Translation is the process in which ribosomes conduct the matching of tRNA with mRNA, producing an amino acid chain, or polypeptide.
7. **D.** The mRNA is a sequence of nucleotides. Each triplet of nucleotides codes for a particular amino acid. The sequence of triplets on the mRNA corresponds to the sequence of amino acids in an entire polypeptide, or protein.
8. **A.** A deletion of a nucleotide in the mRNA produces a change in the triplet-codon reading frame and creates a frameshift mutation. For example, if the first nucleotide of a codon is deleted, then its second nucleotide becomes its first, its third nucleotide becomes its second, and the first nucleotide of the next codon becomes its third. This, then, is repeated in every subsequent codon. Such an arrangement is likely to change many of the amino acids in the sequence (depending upon where in the sequence the frameshift begins) and, thus, affect the final sequence of the polypeptide considerably. Answer choice B will result in a missing amino acid, and answer choice C will very likely change one amino acid to a different amino acid. These changes may alter the effectiveness of the polypeptide, but not as severely as changing many amino acids, as would occur in answer choice A. Answer choice D may have no effect at all because a change in the third position of a codon will often code for the same amino acid. (This results from the wobble of the third position of the tRNA anticodon.) The inherited disorder sickle-cell disease is caused by the replacement of one amino acid by another in two chains of the hemoglobin protein, severely reducing the effectiveness of hemoglobin in carrying oxygen. However, a frameshift in the mRNA coding for hemoglobin would certainly make it entirely ineffective.
9. **C.** In a DNA double helix, thymine base-pairs with adenine.
10. **D.** During transcription of DNA, the uracil RNA nucleotide base-pairs with the adenine DNA nucleotide.
11. **C.** During transcription of DNA, the adenine RNA nucleotide base-pairs with the thymine DNA nucleotide.
12. **B.** Guanine base-pairs with cytosine.
13. **D.** When base-pairing occurs between the anticodon nucleotides of the tRNA and the codon nucleotides of the mRNA, a uracil nucleotide base-pairs with an adenine nucleotide.
14. **D.** This is the same base-pair as in question 13.

15. **A.** The process illustrated here is translation. In particular, the anticodon of a tRNA is shown base-pairing with the codon of the mRNA.

16. **B.** The DNA of all cells uses the same DNA nucleotides—adenine, cytosine, guanine, and thymine nucleotides. On the other hand, the DNA of two unrelated species is likely to differ considerably in the genes produced, as well as in the number of DNA molecules (that is, the number of chromosomes), the DNA lengths, and the DNA nucleotide sequences.

17. **A.** Replication is the process of copying DNA. Protein synthesis involves transcription and translation. Translation begins with initiation, continues with elongation, and ends with termination.

18. **C.** The genetic instructions originate on the DNA and are carried to the cytoplasm by the mRNA. The rRNA and tRNA operate together to translate the code on the mRNA into a polypeptide.

19. **A.** Plasmids are small circular DNA molecules that a bacterium contains in addition to its primary chromosome.

20. **B.** DNA amplification refers to a laboratory process, such as PCR, that generates multiple copies of DNA. Answer choice A is a description of conjugation, which, together with mutation, transformation, and transduction, is a naturally occurring process that can introduce genetic variation into the genome of bacteria.

21. **C.** Viruses consist of a nucleic acid (DNA or RNA) surrounded by a protein coat. Some viruses contain an envelope made from lipids or glycoproteins obtained from the membranes of their hosts, but they do not have the phospholipid bilayer membrane typical of cells.

22. **D.** Introns are intervening sequences in the mRNA that are cleaved from the mRNA by snRNPs before export to the cytoplasm and subsequent translation.

23. **A.** Operons are DNA segments that include a promoter region, an operator region, and a series of structural genes. Together with a regulatory gene lying outside the operon, the three parts of an operon work collectively to control transcription, which results in the regulation of gene expression.

Free-Response Questions

1. Eukaryotic organisms are often multicellular. As a result, gene regulation must be able to regulate gene expression specific to cell type. Eukaryotes also have more than one chromosome, requiring simultaneous regulation of multiple genes on different chromosomes when they all contribute to a metabolic pathway.

2. In many cases, no new amino acid is designated if the mutation is a substitution in the third position of the mRNA codon. More than one codon often codes for the same amino acid, and the third position is the most variable (wobble pairing).

3. Yes. The traditional fingerprint can distinguish identical twins because the expression of dermal ridges (the ridges on the skin that make fingerprints) are the product of genes whose expression has been modified by gene regulation. Gene regulation is often influenced by environmental factors that differ between identical twins during fetal development. In contrast, a DNA fingerprint is a snapshot of the DNA, the "raw" genetic material, uninfluenced by gene regulation; it cannot distinguish identical twins. *(Rare exceptions to this occur as a result of somatic mutations—mutations in cells other than those that produce gametes. Somatic mutations originate from mitotic errors during development and create cell lines within an individual that are genetically different.)*

4. a. Genes are segments of the DNA that contain instructions for producing a specific polypeptide. Many polypeptides are enzymes that regulate cellular reactions, which, in turn, produce chemical end products that appear as traits. The process by which information is transferred from gene to enzyme is called protein synthesis. Transcription, the first step in the process, describes how RNA is synthesized from DNA. RNA polymerase, in association with various regulating transcription factors, attaches to DNA at a promoter region on the DNA. RNA polymerase directs RNA nucleotides to base-pair with the DNA fragment that represents the gene. If a fragment of DNA contained the nucleotide sequence adenine, cytosine, guanine, and thymine, the RNA nucleotides that would base-pair with it are uracil, guanine, cytosine, and adenine, respectively. The products of transcription are three kinds of RNA—mRNA, tRNA, and rRNA.

b. The mRNA contains the code for the polypeptide. Noncoding intervening sequences, called introns, are removed and the mRNA is stabilized with a 5′ GTP cap and a poly-A tail. The mRNA then moves to the cytoplasm.

c. Translation occurs in the cytoplasm and describes the actual assembly of amino acids into proteins. It requires all three kinds of RNA—rRNA, tRNA, and mRNA.

First, ribosomes, consisting of rRNA and proteins, attach to the mRNA. Then a tRNA attaches to the ribosome. There are various kinds of tRNA molecules. Each kind is specific for a particular amino acid that attaches to one end of the tRNA. Each tRNA has a special region of three nucleotides, called an anticodon.

During the process of translation, ribosomes direct the pairing of the anticodons of tRNAs with appropriate triplet regions of the mRNA, called codons. Each mRNA codon specifies a particular amino acid. The genetic code describes which amino acid is indicated by each of the 64 different mRNA codons. Some codons indicate a stop code, signaling that translation of the mRNA is complete. Another codon indicates methionine, the start amino acid.

During translation, the ribosome provides binding sites to incoming tRNAs. Each tRNA brings the appropriate amino acid as dictated by the codon sequence on the mRNA. As each new tRNA arrives, the growing polypeptide chain that is attached to the previous tRNA is transferred to the new tRNA. The old tRNA is released; the ribosome moves over one binding site; and the process repeats until the stop codon is encountered. At this point the ribosome separates into two subunits, and the polypeptide is released.

d. Once released, the amino acids in the polypeptide may interact with one another, giving the polypeptide secondary and tertiary protein structures. Secondary structure results from hydrogen bonding and produces two kinds of structures—a helix or a pleated sheet. Tertiary structure results from additional bonding between R groups of the amino acids and disulfide bonds between cysteine amino acids. In its final form, the polypeptide may be an enzyme that can regulate a reaction that will produce some end product, or trait.

In the limited time you have available to answer this free-response question, your goal should not be to describe a single facet of protein synthesis in extreme detail. Rather, you should try to describe each step in the entire process, pursuing detail only when you have available time. When your essay is evaluated, points are given for each step in the process. There is a maximum number of points given for each step, so pursuing excess detail in one step does not improve your score. Providing a few pieces of information for every part optimizes your time and maximizes your score. For example, you would not improve your score by describing, in detail, the structure of the poly-A tail or 5' GTP cap of the mRNA. Also, details about the regulation of transcription would provide more than the question asks. However, if you omit an explanation of the genetic code, you will probably lose 1 or 2 points out of a maximum of 10.

5. a. Viruses cause disease by destroying cells. Viruses consist of a nucleic acid core (either DNA or RNA) and a protein coat. In the lytic cycle of reproduction, a DNA virus enters a cell and uses the metabolic machinery and raw materials of the cell to manufacture more viral DNA and viral protein coats. The viral DNA and protein assemble into hundreds of new viruses that burst from the cell, killing the cell in the process. In the lysogenic cycle, the virus may temporarily remain dormant as part of the host's genome, to become active only when exposed to radiation or other environmental disturbance. When activated, the viral DNA begins the lytic cycle.

Some viruses are RNA retroviruses. These RNA viruses produce the enzyme reverse transcriptase to first manufacture DNA, which, in turn, enters a lytic or lysogenic cycle.

b. Bacteria, unlike viruses, do not usually cause disease by direct destruction of host cells. Rather, most bacteria cause disease by producing toxins, usually waste products of their normal metabolism. When the toxins affect the normal metabolism of the host, disease results.

Some bacteria also cause disease by competing for the same resources as do the host cells. In other cases, the symptoms of a disease are the result of the host's response to invasion by foreign bodies. For example, in pneumonia, the mucus that accumulates in the lungs is produced by the lung cells in response to the presence of the bacteria.

6. a. *To answer this question, describe one (or more, if time permits) of these operons: lac operon, trp operon, or how CAP regulates lactose metabolism.*

b. *For this part of the question, you should describe how transcription factors work. If time permits, describe other mechanisms such as DNA methylation, histone modification, RNA interference, and mRNA degradation.*