2.2 DNA and Protein Synthesis

Learning Objectives

- Compare genotype and phenotype
- Describe the components of a nucleotide
- Name the different nucleotides that comprise DNA and RNA
- Describe how the nucleotides are connected to form single-stranded DNA or RNA
- Describe how hydrogen bonding is useful in combining antiparallel nucleotide strands to form the double helix
- Explain why antiparallel strands require special arrangement for DNA replication
- List the ways RNA differs from DNA
- List the various kinds of RNA and their function in the cell
- Know the origin of the RNA polymerases responsible for formation of the various RNA classes
- Describe what is meant by "the genetic code"
- Define and contrast "transcription" and "translation"
- Distinguish among "response elements," "intron," and "exon"
- Describe histones and their postulated function in the structure of chromosomes
- Describe what is meant by the "histone code"
- Describe how DNA is methylated and how this methylation can be passed from parent to daughter cells

DNA MAKES UP THE GENOME

As described in Chapter 1.1, almost all cells in the body have the same amount and kind of DNA. The diversity of human cell forms derives from their **expression** of different parts of the DNA. The total DNA with its division into units, called **genes**, constitutes the **genome**. **Expression** of a gene means that the DNA that makes up the gene is used to direct the synthesis of a specific protein. As we will see in this chapter, genes associate with a host of proteins that regulate the expression of the genes. The human genome refers to the set of genes that are normally present in humans.

The DNA in human cells is organized into compact units called **chromosomes**, meaning "colored body," which refers to their appearance in fixed and stained preparations. Each chromosome carries a defined set of genes that carries the instructions for making a set of proteins. Because each chromosome is paired, nearly every gene comes in pairs, but the two pairs are usually not identical. Paired genes carry the instructions for the synthesis of analogous materials, but they differ in the details. These alternate forms of the genes in a single person are called **alleles**. The set of alleles of a particular person is called the **genotype**. The set of proteins and other materials that the person actually makes, and which determine their outward appearance and behavior, is called the **phenotype**. Humans have 23 pairs of chromosomes.

Two chromosomes determine the sex of the individual. These are the X and Y chromosomes. Persons with two X chromosomes are genotypic females; having one X and one Y makes a genotypic male. Because the Y chromosome is smaller than the X, some of the genes carried on the X chromosome are not paired. This is the one exception to the rule that all genes are paired.

DNA CONSISTS OF TWO INTERTWINED SEQUENCES OF *NUCLEOTIDES*

DNA IS BUILT FROM NUCLEOTIDES

DNA stands for **deoxyribonucleic acid**. It is located primarily in the nucleus of cells but important parts are also present in the mitochondria. It is composed of a sequence of building blocks called **nucleotides**. These nucleotides come in two different types and four varieties. The types are the purines and pyrimidines. The purines in DNA are **adenine** and **guanine**, and the pyrimidines are **thymine** and **cytosine**. Each of these nucleotides consists of the base (adenine, guanine, thymine, and cytosine) linked to a sugar, **deoxyribose**, and **phosphate**. The chemical structures of the nucleotides are shown in Figure 2.2.1.

NUCLEOTIDES ARE LINKED TOGETHER TO FORM A CHAIN

These four bases are linked together to form a long sequence of nucleotides. The DNA is elongated by reacting a nucleotide triphosphate (with two more phosphates linked to the phosphate shown in Figure 2.2.1) on the 3′ end of an existing chain. This reaction is catalyzed by an enzyme called **DNA polymerase**. This enzyme is involved in the **replication** of DNA, where two complete DNA strands are made from

DOI: http://dx.doi.org/10.1016/B978-0-12-800883-6.00011-2

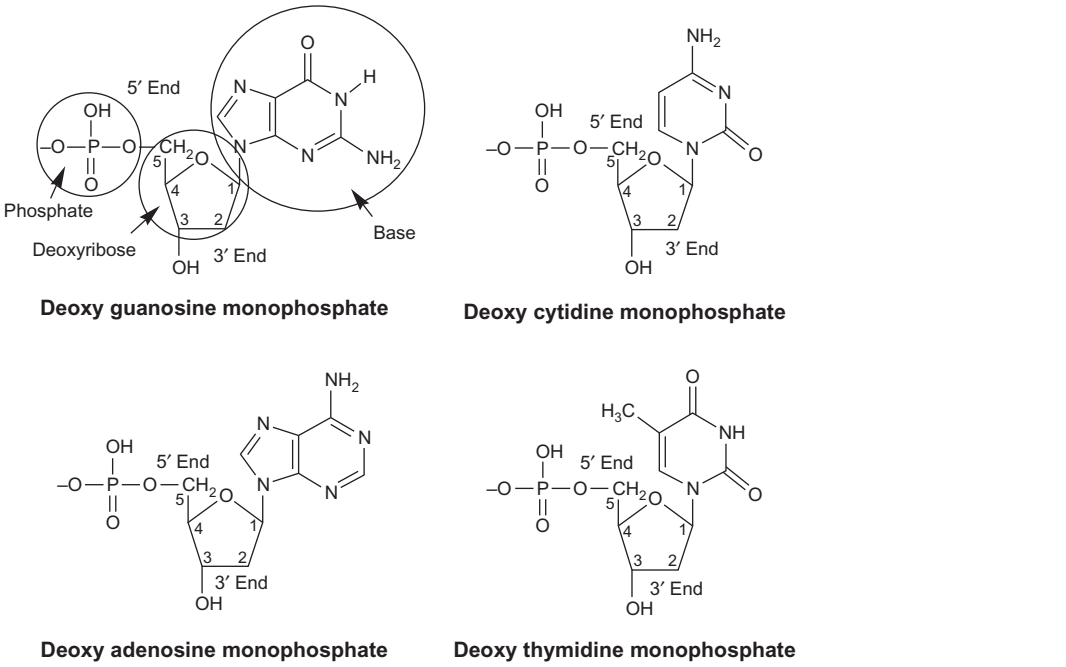

FIGURE 2.2.1 Structures of the nucleoside monophosphates.

one, using the original DNA as a template. The structure of single-stranded DNA is shown in Figure 2.2.2.

The single-stranded DNA described in Figure 2.2.2 is just half of the story. In humans, DNA is normally present as double strands that are held together by **hydrogen bonds**, as shown in Figure 2.2.3. In double-stranded DNA, adenine on one strand pairs with thymine on the opposite strand and guanine on one strand pairs with cytosine on the opposite strand. The two strands have opposite polarity: the 5′ end of one strand is opposite to the 3′ end of the other.

HYDROGEN BONDING ALLOWS FOR DNA STABILITY WITH RAPID DISSOCIATION

As discussed in Chapter 1.4, hydrogen bonds involve sharing of the positive H atom between two electronegative centers. It requires the right spatial separation and orientation of these centers, and has low dissociation energy. This allows the H-bond to form or break rapidly. However, many hydrogen bonds can stabilize large structures like DNA and proteins.

THE DNA TEMPLATE SETS THE SEQUENCE OF NUCLEOTIDES

DNA polymerase adds nucleotides to the 3′ end, using a nucleotide triphosphate as a substrate. The base on the opposite strand determines which nucleotide is incorporated. Thus DNA polymerase replicates DNA on the basis of the DNA already present. The DNA strand unwinds to form two single strands. The DNA polymerase adds nucleotides on both strands to form two complete DNA double strands. The hydrogen bonding between nucleotides is crucial to the ability of DNA to

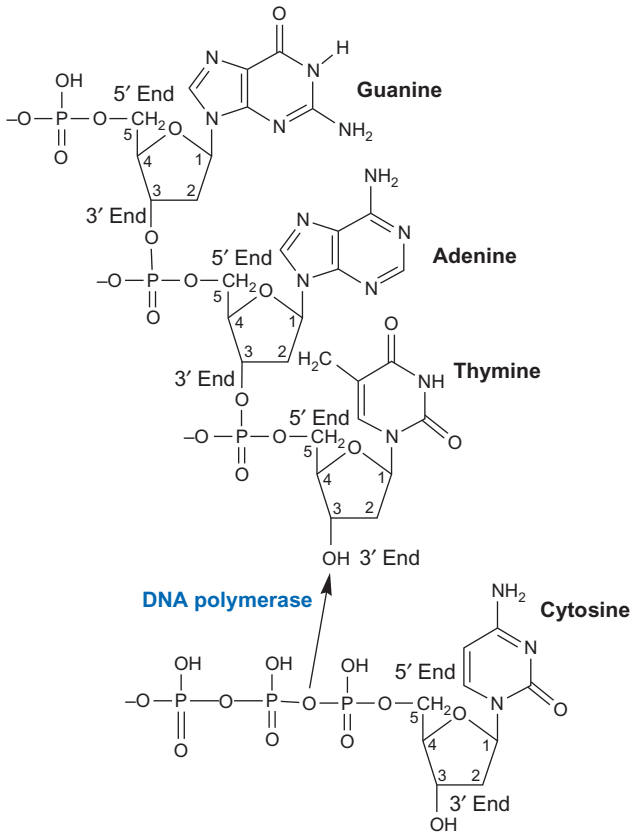

FIGURE 2.2.2 Arrangement of bases in single-stranded DNA. The phosphate–deoxyribose part of the nucleotide triphosphates forms a backbone of alternating phosphate and deoxyribose molecules. Attached to this backbone are the four bases: guanine, adenine, thymine, and cytosine. The 3′ and 5′ ends of the strand derive from the numbering of the ribose carbons.

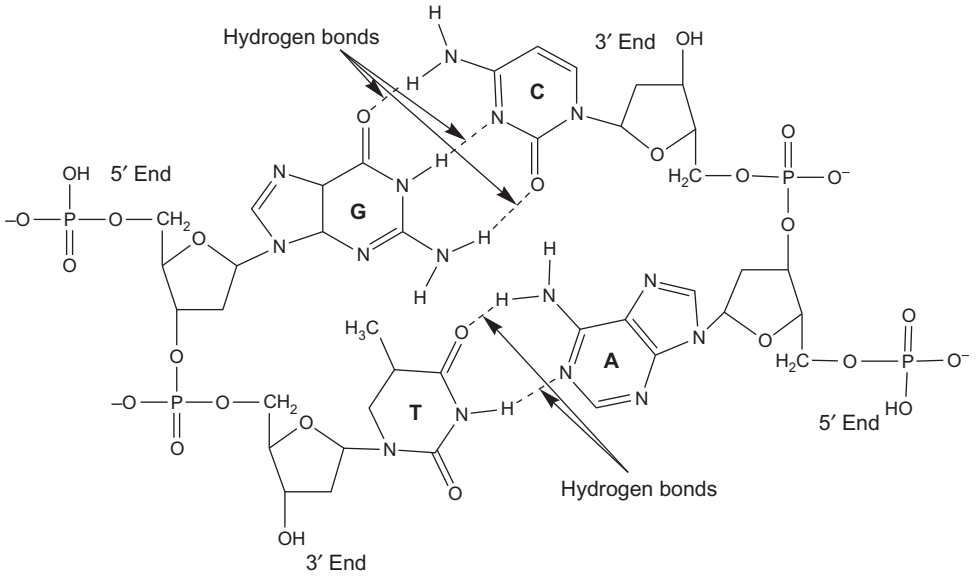

FIGURE 2.2.3 Pairing of bases in double-stranded DNA.

serve as a template for its own replication and for the synthesis of RNA.

THE DOUBLE HELIX IS SUPERCOILED IN CHROMOSOMES

The two DNA strands intertwine around each other to form a double helix. This helix can be further wrapped around proteins called **histones**. The complex of DNA and its associated proteins is called **chromatin**. The complex resembles "beads on a string" that are visible in electron micrographs (see Figure 2.2.9). The entire structure can be further coiled and coiled again to form the chromosomes. All of the DNA becomes condensed like this when the cell divides. Between divisions, the chromosomes partially "unravel" to form a less dense form of chromatin that is the working state of DNA.

THE DOUBLE HELIX POSES SPECIAL CHALLENGES FOR DNA REPLICATION

The replication of DNA by adding nucleotides only to the 3′ end of the growing DNA strand poses a problem for the duplication of double-stranded DNA because the strands are of opposite polarity. This leads to the replication of DNA in spurts, as shown in Figure 2.2.4 and described in the legend.

RNA IS CLOSELY RELATED TO DNA

RNA is **ribonucleic acid**. Structurally, it is very similar to DNA but it differs in several ways. First, the sugar in the backbone in DNA is **deoxyribose**, whereas in RNA it is **ribose**. Second, the nucleotide base thymine in DNA is replaced by **uracil** in RNA. Uracil hydrogen bonds with adenine, taking the place of thymine. Third, RNA in eukaryotic cells is single stranded. Fourth, RNA is not replicated. All of the RNA is produced from DNA using DNA as a template.

There are different kinds of RNA:

- mRNA: "messenger" RNA
- tRNA: "transfer" RNA
- rRNA: "ribosomal" RNA
- snRNA, scRNA: "small nuclear" and "small cytoplasmic" RNA
- Mitochondrial RNA.

MESSENGER RNA CARRIES THE INSTRUCTIONS FOR MAKING PROTEINS

mRNA is "messenger" RNA. mRNA is synthesized in the nucleus using the nucleotide sequence of DNA as a template. This process requires nucleotide triphosphates as substrates and is catalyzed by the enzyme **RNA polymerase II**. The process of making mRNA from DNA is called **transcription**, and it occurs in the nucleus. The mRNA directs the synthesis of proteins, which occurs in the cytoplasm. mRNA formed in the nucleus is transported out of the nucleus and into the cytoplasm where it attaches to the **ribosomes**. Proteins are assembled on the ribosomes using the mRNA nucleotide sequence as a guide. Thus mRNA carries a "message" from the nucleus to the cytoplasm. The message is encoded in the nucleotide sequence of the mRNA, which is complementary to the nucleotide sequence of the DNA that served as a template for synthesizing the mRNA. Making proteins from mRNA is called **translation**.

RIBOSOMAL RNA IS ASSEMBLED IN THE NUCLEOLUS FROM A DNA TEMPLATE

As discussed in Chapter 2.1, ribosomes are complex structures comprised of ribosomal RNA (rRNA) and a number of proteins. **RNA polymerase I** makes rRNA form a large loop of DNA called the **nucleolar organizer** region. The rRNAs then combine with proteins that migrate into the nucleolus from the cytoplasm to form the small and large ribosomal subunits. These ribosomal

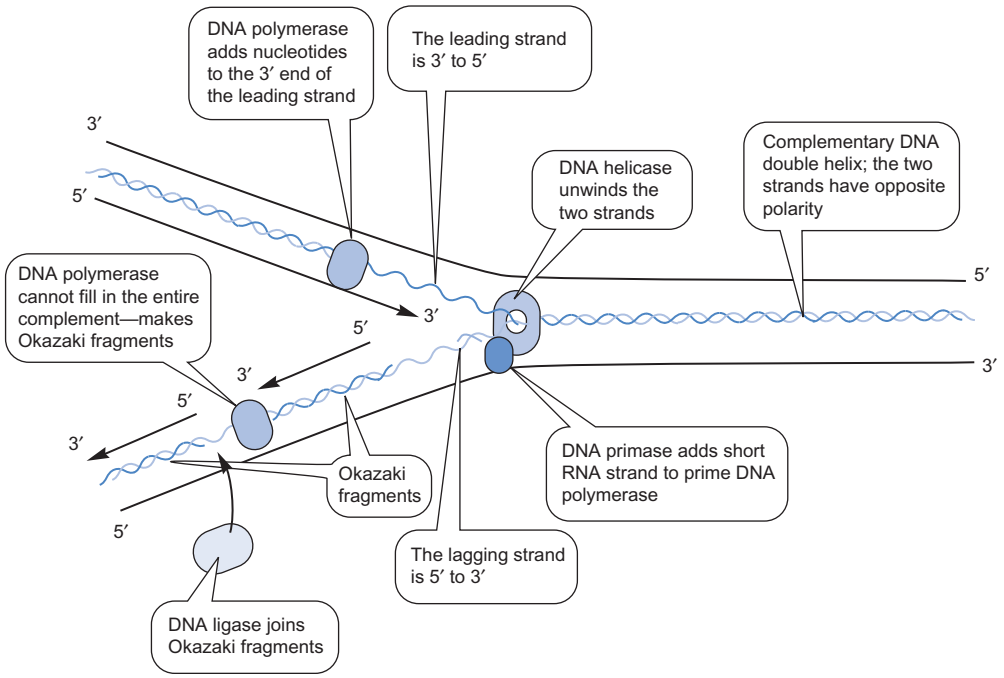

FIGURE 2.2.4 Replication fork of DNA. The double-stranded DNA consists of two strands of opposite polarity. **DNA helicase** unwinds the strands, forming two single-stranded DNAs. The **leading strand** is 3′ to 5′ so its complementary strand being newly synthesized is 5′ to 3′ and DNA polymerase adds nucleotides to the 3′ end as they become available from the helicase. The **lagging strand** is 5′ to 3′ and its complementary strand is 3′ to 5′. DNA primase adds a short RNA strand that primes the DNA polymerase. DNA polymerase then makes the complement by progressing away from the **replication fork**. While it is making DNA, the helicase unwinds more DNA so another DNA polymerase starts replication nearer the replication fork. In this way, the lagging strand is filled in with **Okazaki fragments** that bind to the lagging strand but are not connected. **DNA ligase** connects the Okazaki fragments to complete replication of the lagging strand.

subunits are then transferred to the cytoplasm where they are fully assembled to form an 80S functional ribosome and become protein factories.

TRANSFER RNA *COVALENTLY BINDS AMINO ACIDS AND RECOGNIZES SPECIFIC REGIONS OF MRNA*

How does mRNA specify the sequence of amino acids in a protein? Which amino acid is to be incorporated into the protein is specified by a sequence of three nucleotides called a **codon**. The mRNA triplets do not directly recognize and specify the amino acids; they do so through the use of another kind of RNA called **transfer RNA** or **tRNA**. These remarkable molecules are adapters that can link with an amino acid and recognize the triplets of nucleotides on the mRNA, the codons. They do this by containing a sequence complementary to the codon: the **anticodon**. The function that maps triplets of nucleotides on the mRNA to specific amino acids is called the **genetic code**. Figure 2.2.5 shows the genetic code in look-up table format.

The tRNA consists of a single strand of RNA from 70 to 90 nucleotides long that is held together by hydrogen bonding within nucleotides on the same chain. One end of the tRNA allows for covalent attachment of an amino acid. Another section of the tRNA contains a sequence of three nucleotides that forms the anticodon. Precursors to the tRNA are transcribed from DNA by **RNA polymerase III**.

Another key in the formation of proteins is the attachment of amino acids to the specific tRNA. Specific enzymes called **aminoacyl-tRNA synthetases** couple the amino acid to the appropriate tRNA. There is a different synthetase for each amino acid. One attaches glycine to $tRNA^{Gly}$, another attaches alanine to $tRNA^{Ala}$, and so on. These synthetases must recognize both the amino acid and the tRNA that contain the right anticodon. The overall processing of RNA and protein synthesis is shown in Figure 2.2.6. Translation is shown in Figure 2.2.7.

THE GENETIC CODE IS A SYSTEM PROPERTY

The genetic code shown in Figure 2.2.5 lists an amino acid or other signal (such as STOP) for every triplet nucleotide in mRNA. The code is the function that maps the nucleotide sequence onto instructions for protein synthesis. That is,

$$[2.2.1] \qquad f\{N_i : \{\mathrm{A, U, C, G}\}\} = \{A_j : \{\text{control steps, amino acids}\}\}$$

This describes the function that turns a set of N_i nucleotides on the mRNA (selected from adenosine, uracil, cytosine, and guanosine nucleotides) into a set of A_j amino acids. Where does this code reside? The code should not be confused with the message or the means of writing the message. Therefore, the code itself is not a

First position (5′ end)	Second position U	Second position C	Second position A	Second position G	Third position (3′ end)
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Amino acid	Three-letter code	One-letter code	Codons
Alanine	Ala	A	GCA GCC GCG GCU
Arginine	Arg	R	AGA AGG CGA CGC CGG CGU
Asparagine	Asn	N	AAC AAU
Aspartic acid	Asp	D	GAC GAG
Cysteine	Cys	C	UGC UGU
Glutamic acid	Glu	E	GAA GAG
Glutamine	Gln	Q	CAA CAG
Glycine	Gly	G	GGA GGC GGG GGU
Histidine	His	H	CAC CAU
Isoleucine	Ile	I	AUA AUC AUU
Leucine	Leu	L	UUA UUG CUA CUC CUG CUU
Lysine	Lys	K	AAA AAG
Methionine	Met	M	AUG
Phenylalanine	Phe	F	UUC UUU
Proline	Pro	P	CCA CCC CCG CCU
Serine	Ser	S	AGU AGC UCU UCC UCA UCG
Threonine	Thr	T	ACA ACC ACG ACU
Tryptophan	Trp	W	UGG
Tyrosine	Tyr	Y	UAC UAU
Valine	Val	V	GUA GUC GUG GUU

FIGURE 2.2.5 The genetic code. Each amino acid that is incorporated into a protein is specified by a triplet sequence of nucleotides on the mRNA. These triplets are called codons. Which codons specify which amino acids is shown here in two formats. The left format shows which amino acids correspond to which codons given as a first, second, and third position. Here we use the Biochemists' shorthand for RNA bases and amino acids, where U = uracil, C = cytosine, A = adenine, and G = guanosine, and each amino acid is given by its three-letter shorthand designation, where Phe = phenylalanine, Ser = serine, Tyr = tyrosine, Cys = cysteine, Leu = leucine, Pro = proline, His = histidine, Arg = arginine, Gln = glutamine, Ile = isoleucine, Thr = threonine, Asn = asparagine, Lys = lysine, Met = methionine, Val = valine, Ala = alanine, Asp = aspartic acid, Gly = glycine, and Glu = glutamic acid. The right format lists the amino acids together with their three-letter designation and single-letter designation, with a list of the codons that specify them.

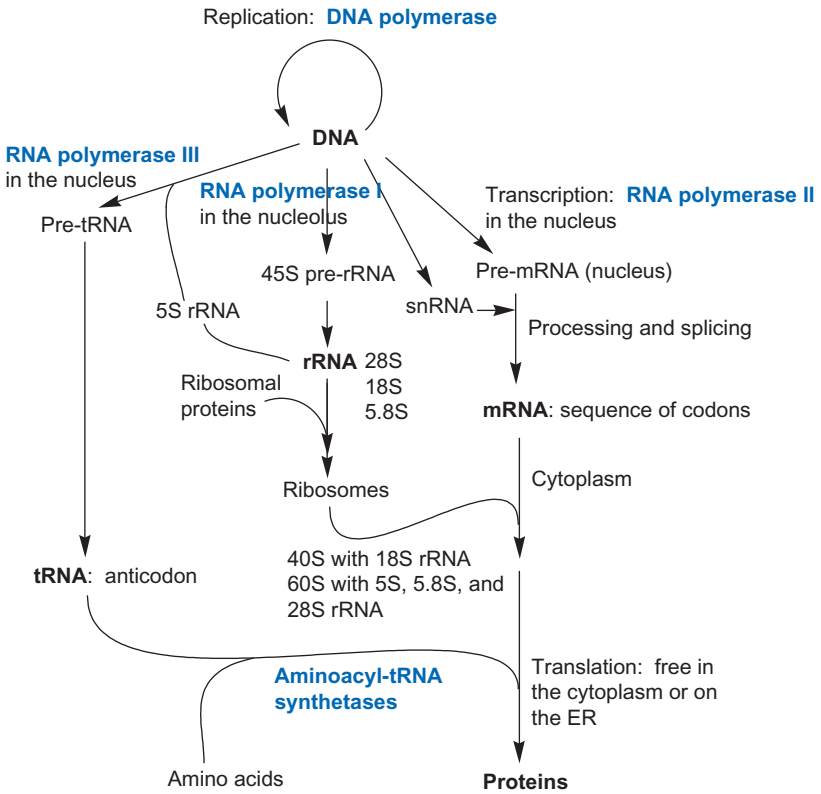

FIGURE 2.2.6 Processing of RNA and DNA. Replication of DNA is accomplished by DNA polymerase using the original DNA as a template. Messenger RNA is synthesized in the nucleus as a precursor that is processed to form the final mRNA. The synthesis of mRNA is called transcription and is accomplished by RNA polymerase II. The final mRNA travels to the cytoplasm where it binds to ribosomes. The ribosomal subunits are formed in the nucleolus from proteins and ribosomal RNA that is made as a precursor and cut into the final rRNA strands. rRNA is made from DNA by RNA polymerase I. The mRNA directs the sequential addition of amino acids to form proteins in a process called translation. This requires tRNA, made from DNA by RNA polymerase III.

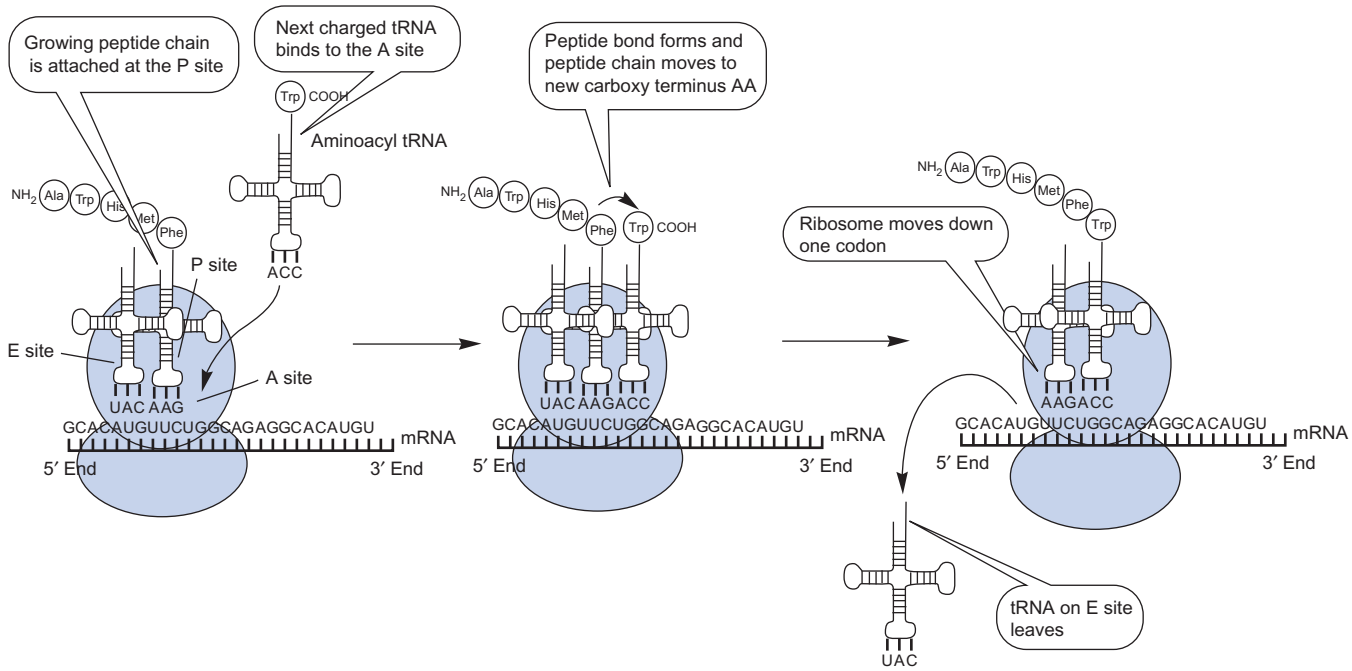

FIGURE 2.2.7 Elongation of a growing polypeptide chain. The mRNA binds to a ribosome consisting of some 82 proteins and 4 separate rRNA strands. There are 3 tRNA binding sites: an **aminoacyl** or A site, a **peptidyl** or P site, and an **exit** or E site. The aminoacyl tRNA is escorted to the ribosome by an **elongation factor** that hydrolyzes GTP. The scheme begins with a short polypeptide bound to a tRNA at the peptidyl site along with the $tRNA^{Met}$ that remains at the exit site. The next aminoacyl tRNA binds to the aminoacyl site; in this case it is $tRNA^{Trp}$ that is charged with its amino acid. The peptide bond is formed between the peptide and its next amino acid on the carboxy terminus. The ribosome shifts over one codon; the tRNA at the exit site leaves and the former occupant of the peptidyl site now occupies the exit site. The peptide now occupies the peptidyl site one codon further along the mRNA.

property of the mRNA. The DNA itself also does not contain the code, as its message has no meaning without the tRNA. The tRNA is synthesized from other parts of the DNA that are not directly transcribed for protein synthesis. Is the code in the tRNA that links the triplets of mRNA to a specific amino acid? Or is the code in the aminoacyl-tRNA synthetases, the enzymes that couple the amino acid to the specific tRNA? If the code is in the aminoacyl-tRNA synthetases, then the code is in some sense also in the DNA because the DNA directs the synthesis of the aminoacyl-tRNA synthetases! But the DNA does not "make" proteins; mRNA does not "make" proteins. The proper proteins are synthesized with DNA as the store of information, mRNA carrying the message, tRNA converting, in a single step, the nucleotide information into the protein information, and preexisting proteins catalyzing the entire series of events. The code itself is an emergent property. The genetic code does not exist in any single component of the mechanism for making proteins. It is not "in" the DNA, or the mRNA or the tRNA or the tRNA synthetases. Rather, it is a system property that emerges from the interactions of all of these parts.

REGULATION OF DNA TRANSCRIPTION DEFINES THE CELL TYPE

The differentiation of cell types produces the wide spectrum of cell types in the body, but our understanding of the process is still rudimentary: we do not know how to change one cell type into another. We do know, however, that the hallmark of differentiation is selective expression of the cell's DNA. Selective expression of DNA as proteins requires selective transcription. Initiation of transcription by RNA polymerase II requires a number of specific proteins called **transcription factors**. These come in two flavors: some are required for activity at all genes and are therefore called general transcription factors. Other transcription factors bind to DNA sequences that control the expression of specific genes. The overall process is shown in Figure 2.2.8.

Most genes contain both transcribed and untranscribed regions. On the 3′ end of the DNA gene is a specialized sequence called the **promoter** that helps regulate gene expression. This region contains a TATAA sequence (called a TATA box) some 25–30 nucleotides upstream from the initiation site. Initiation begins when a transcription factor **TFIID** (transcription factor polymerase II) binds to the DNA TATAA sequence. Some promoters do not contain a TATA box, but instead contain an initiator sequence. Nevertheless, TFIID is involved in initiation of transcription even on promoters that lack the TATA box.

In addition to these general transcription factors, there are a number of transcription factors that may act as **enhancers** or **repressors** of gene expression. These factors interact with specific sites on the DNA that are generally further away from the gene than the promoter region and act as regulatory elements for gene transcription. The signals that turn on or turn off the production of transcription factors ultimately determine the phenotypic fate of cells. How these transcription factors work is still being investigated. An example of this is the

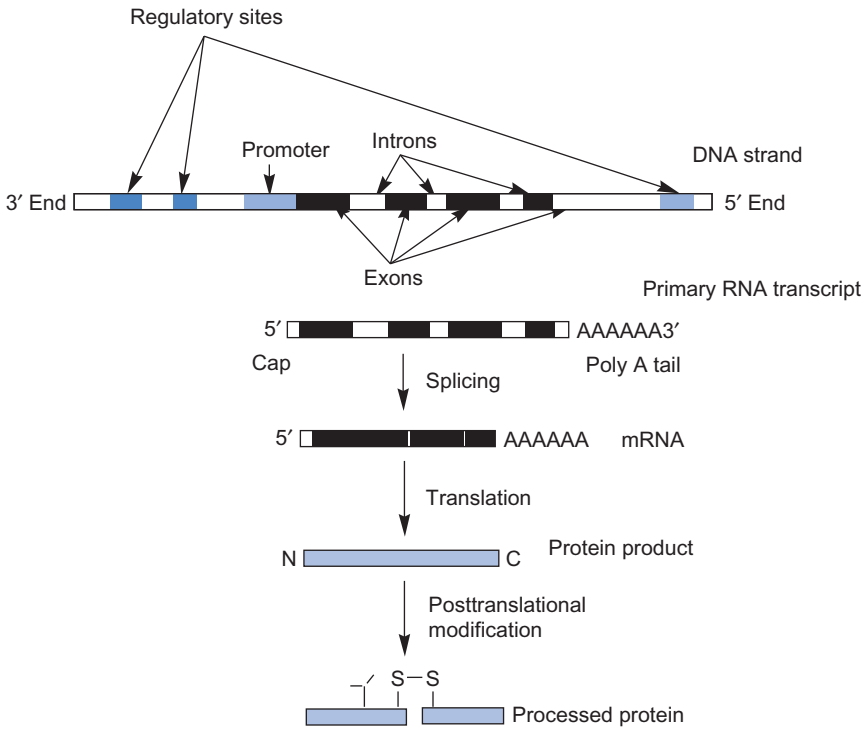

FIGURE 2.2.8 Overall processes of transcription and translation. The gene has a promoter that often contains a TATAA sequence for initiation of transcription by RNA polymerase II, several **exons** (expressed or coding regions of the DNA sequence) and **introns** (intervening or noncoding regions), and regions that bind transcription factors to either enhance or repress expression. The primary gene transcript contains the sequence of bases complementary to the introns and exons. This primary RNA transcript is then further processed on large complexes called **spliceosomes**, composed of protein and **small nuclear RNAs** or **snRNA**. The spliceosomes remove the introns by clipping them out and then splice together the remaining exons to form a sequential mRNA. This mRNA is then translated in the cytosol to a primary protein product which may then require further processing.

steroid hormone receptors. Steroid hormones and similar materials like vitamin D bind to receptor proteins that in turn bind to specific sequences on the DNA called **response elements**. Binding of the receptor proteins then activates gene expression. A number of accessory proteins are required for this process.

THE *HISTONE CODE* PROVIDES ANOTHER LEVEL OF REGULATION OF GENE TRANSCRIPTION

As described earlier, the double helix of DNA winds around specific proteins called histones. These histones form the basic unit of chromatin called the **nucleosome**. A schematic of this structure is shown in Figure 2.2.9. When wrapped up in this way, DNA is inaccessible to RNA polymerase II and so cannot be transcribed to form mRNA. Most cells sequester away large portions of their DNA in this way. In order to express DNA, cells must unwrap it from the chromatin. Determining which sections of DNA should be unwrapped is the first step in regulating gene transcription. This is accomplished through covalent modifications of the histones. Covalent modification of histones after synthesis is called **posttranslational modification**. Histones undergo

- acetylation of lysine and arginine amino acids in the histones;
- methylation of lysine and arginine amino acids;
- phosphorylation of serine and threonine amino acids in the histones;
- ubiquitinylation of lysines.

All of these modifications are accomplished by enzymes that must themselves be regulated. These enzymes possess **histone acetyltransferase** (HAT) activity, **histone deacetylation** (HDAC) activity, **histone methyltransferase** (HMT) activity, and **histone kinase** activity. The function of these enzymes is evident from their names: HAT adds acetyl groups to the histones; HDAC removes them; HMT adds methyl groups; and histone kinase phosphorylates the histones. Figure 2.2.10 summarizes the various known modifications of the core histones.

These posttranslational modifications of the histones have functional consequences. For example, the combination of acetylation at lysine at position 8 on H4 and lysine at position 14 on H3 with phosphorylation of serine at position 10 on H3 is associated with transcription. Trimethylation of lysine at position 9 on H3 with the lack of acetylation of H3 and H4 correlates with transcriptional repression. These observations have led to the hypothesis that gene expression is regulated in part by a "histone code." This hypothesis requires two components:

1. Specific enzymes write the code by adding or removing modifications at specific sites in the histones.
2. Other proteins recognize the histone markers and interact with histones and other factors to mediate functional effects.

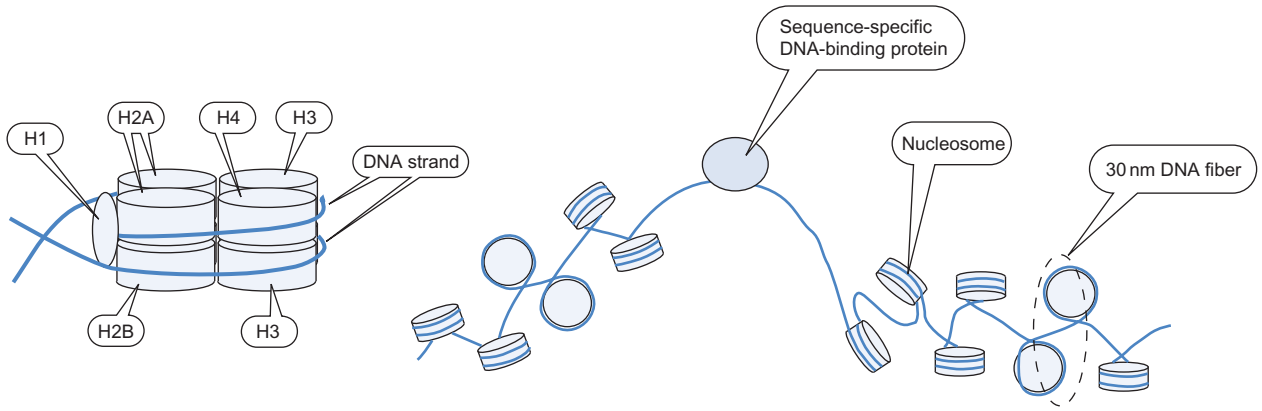

FIGURE 2.2.9 Schematic diagram of nucleosome structure and 30-nm DNA fiber. DNA is wrapped around a complex of eight histone proteins: a tetramer of H3 and H4 and two dimers of H2A and H2B, to form the nucleosome. Strands of DNA bound to nucleosomes resemble beads on a string. The interaction of DNA with the nucleosomes is altered by H1, a linker histone, that enables the nucleosomes to condense to form a 30-nm DNA strand. H1 interacts with both the histones in the core and DNA. The nucleosome contains 147 base pairs of DNA wrapped nearly twice around the core histones. A short "linker" of 10–60 base pairs separates each nucleosome from its neighbor. The linear sequence of nucleosomes forms "beads on a string" in electron micrographs. This becomes more highly condensed to form 30-nm-thick fibers that are stabilized by the H1 histones.

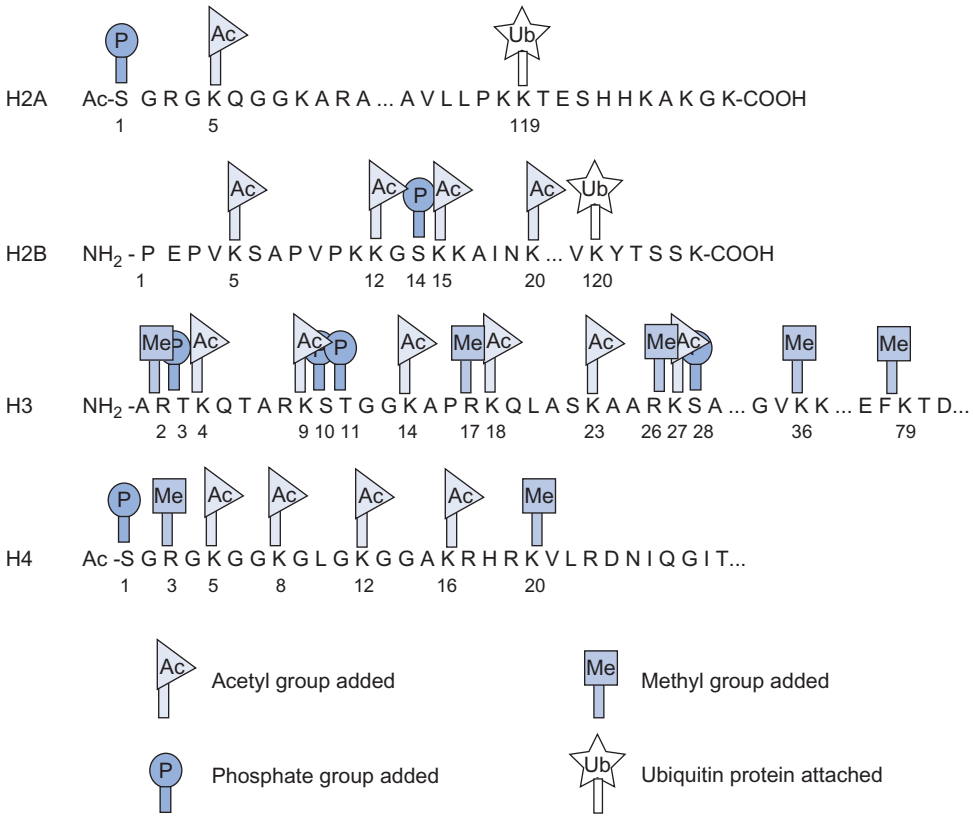

FIGURE 2.2.10 Posttranslational modifications of the core histones. The one-letter amino acid abbreviation follows that in Figure 1.3.5. The numbers below the amino acids refer to the number of amino acids in the sequence, starting from the amino terminus. Source: *From C.L. Peterson and M.-A. Lanie, Histones and histone modifications,* Current Biology ***14****:R546–R551, 2004.*

The histone code may be a misnomer if it is viewed as being like the genetic code. In the genetic code, given triplets of nucleotides on mRNA *always* produce the same result, independent of which cell or tissue is being analyzed. The term "histone code" implies that a particular combination of histone modifications will *always* produce the same biological result. Evidence suggests that the same pattern of histone modifications can be interpreted differently by different cells depending on the gene and its cellular context.

DNA METHYLATION REPRESSES TRANSCRIPTION

A methyl group can be added to the 5 position of cytosine to form 5 methyl cytosine by the action of DNMT, DNA methyltransferase, as shown in Figure 2.2.11. These enzymes come in two classes. DNMT1 is responsible for maintenance methylation, which methylates the new strand of recently replicated DNA, so that the methylation pattern is passed down from stem cells to daughter cells (see Figure 2.2.12). This explains the unidirectionality of most developmental processes. Stem cells become differentiated cells and the differentiated cells maintain their differentiation, partly through a pattern of DNA methylation. The second class of DNA methylation transferases, represented by DNMT3a and DNMT3b, is responsible for de novo methylation. Both the maintenance methylation enzymes and de novo methylation enzymes methylate cytosine in a CpG sequence in the DNA (see Figure 2.2.11).

The consequence of DNA methylation is typically repression of transcription for the genes that are methylated. The methyl group does two things: it interferes with the binding of transcription factors that eventually recruit RNA polymerase II, and it also allows the binding of a set of proteins that specifically recognize the methylated DNA, and these proteins recruit histone deacetylases that modify the histones associated with the DNA. These actions result in a repressed transcription of the methylated parts of the DNA.

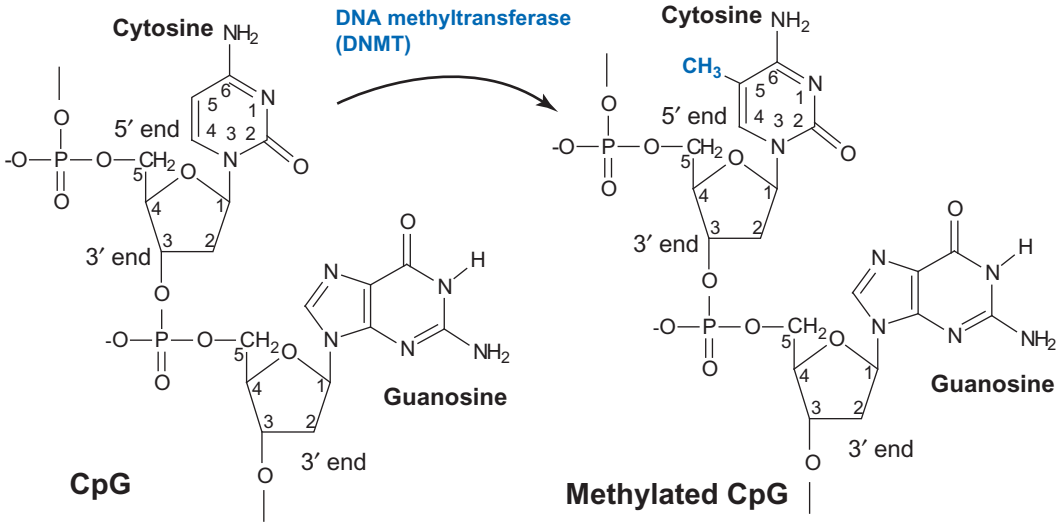

FIGURE 2.2.11 DNA methylation. DNA is methylated at sites with the sequence CpG by DNA methyltransferase. It can also be demethylated, but the process is not simply a reversal of the methylation reaction. Demethylases form hydroxymethyl cytosine, which is then cut out and replaced with cytosine by DNA repair mechanisms.

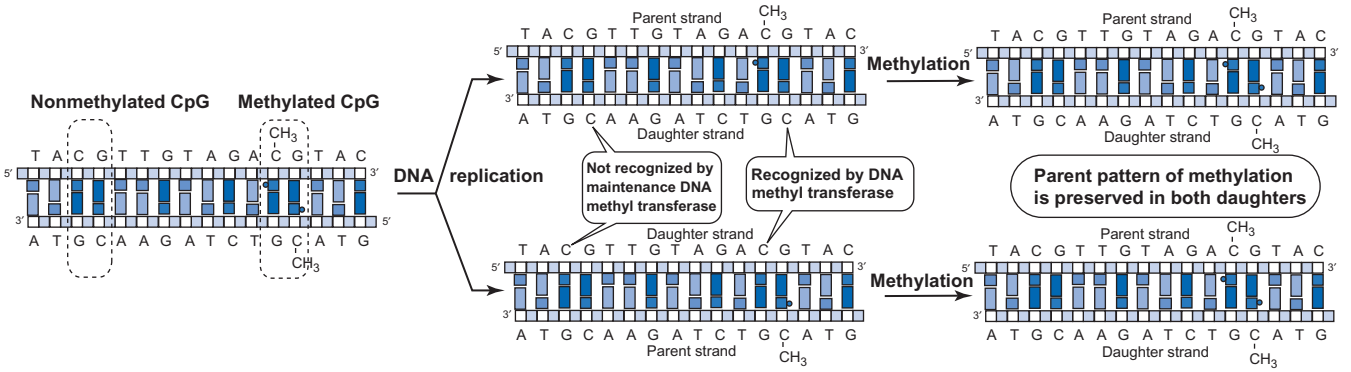

FIGURE 2.2.12 Maintenance methylation. The parent DNA has some CpG sequences that are methylated, and some that are not. These are determined by de novo methylation reactions catalyzed by DNA methyl transferase 3A and 3B (DNMT3a and DNMT3b). When the DNA is replicated, the daughter strands are not methylated. The maintenance DNA methyl transferase (DNMT1) recognizes the CpG sequence on the daughter strand that corresponds to the methylated CpG sequence on the parent strand, and then methylates the daughter strand cytosine. DNMT1 does not bind to the unmethylated CpG sequences and therefore does not methylated these. The result is that the methylated pattern of the parent strand is replicated in the methylated pattern of the two daughter strands. In this way, the pattern of gene repression in a differentiated cell line is passed onto its descendants.

SUMMARY

Genetic information is stored in DNA in the nucleus and mitochondria of cells. DNA consists of two strands of nucleotides on a phosphodeoxyribose backbone. The two strands form a double helix that is stabilized by the formation of hydrogen bonds between nucleotide bases on the two strands. Replication of DNA is based on making strands complementary to the two strands produced when the double-stranded DNA is unwound. The four bases include two purines (adenine and guanine) and two pyrimidines (cytosine and thymine). They pair up as A:T and G:C, with two hydrogen bonds between A and T and three between G and C.

DNA serves as a template for making a variety of RNA types: RNA polymerase I makes ribosomal RNA (rRNA) from nuclear DNA; RNA polymerase II makes messenger RNA (mRNA); RNA polymerase III makes transfer RNA (tRNA). All of these contribute to the synthesis of proteins and control of gene expression.

Protein synthesis occurs on the ribosomes, using mRNA as a template for tRNA. The ribosomes themselves are large and complex structures consisting of rRNA and a number of proteins. The ribosomes consist of 60S and 40S subunits. The 60S subunit has three rRNAs (28S, 5S, and 5.8S) and 49 other proteins; the 40S subunit has an 18S rRNA and another 33 proteins. The ribosome binds to mRNA and provides a reaction site for the peptide bound to tRNA (P site), a second reaction site for the next amino acid covalently attached to its tRNA (A site), and a third site for the tRNA about to leave the ribosome (E site). The ribosome brings the peptide carboxyl terminal close to the amino terminal of the next amino acid, and then forms the peptide bond, simultaneously shifting the peptide from the P to the A site. The mRNA provides a sequence of nucleotides. Groups of three nucleotides form codons that are recognized by complementary nucleotide sequences on the tRNA—the anticodons. The specific binding of anticodon to codon allows the mRNA to determine the sequence of amino acids in the proteins being made. The attachment of specific amino acids to specific tRNAs, however, is assured by the tRNA synthetases that hook the amino acids onto tRNA.

mRNA directs the synthesis of specific proteins by virtue of its sequence of codons. The set of proteins that are made determines the type of cell, because cell structure and activity derives from the kinds and amounts of proteins expressed by the cell. Cells have developed elaborate methods for determining what parts of the DNA are transcribed into mRNA. These methods involve transcription factors, enhancers and repressors, and the histone code. The DNA in cells is wrapped around a complex of histone proteins, forming a nucleosome. Modification of the histones allows specific sections of the DNA to be either used to make proteins or silenced.

A second method for repression of gene transcription is methylation of cytosines in the DNA in sequences of CpG. These are originally methylated by de novo methylation, but the pattern of methylation can survive DNA replication through a maintenance methyltransferase. The result is a stable pattern of gene repression that survives proliferation. DNA methylation interacts with histone modification to determine which DNA sequences will be silenced.

REVIEW QUESTIONS

1. What is the genotype? What is the phenotype? What is an allele? What is the usefulness in having two copies of each gene?
2. What are the parts of a nucleotide? Name the purine bases. Name the pyrimidine bases.
3. How are nucleotides linked together to form single-stranded DNA or RNA? What distinguishes the 5′ end from the 3′ end? What enzyme makes DNA from nucleotides?
4. What holds double-stranded DNA together? Why are hydrogen bonds useful? How many hydrogen bonds link guanine to cytosine? How many such bonds link adenine to thymine?
5. During DNA replication, what determines the sequence of DNA in the new strands?
6. Does DNA polymerase add nucleotides at the 5′ end or the 3′ end of the strand? What problem does this make for DNA replication? What is an Okazaki fragment?
7. How does RNA differ from DNA? What is mRNA? What is tRNA? What is rRNA? What RNA polymerase makes mRNA? Which makes tRNA? rRNA?
8. What is a codon? Is it on mRNA, DNA, or tRNA? What is an anticodon?
9. What couples tRNA with amino acids? Are these enzymes specific for the anticodon?
10. What is a ribosome? What is the A site? What is the P site? What is the E site?
11. What is the genetic code? Where is it located in the cell?
12. What is meant by "transcription"? What is meant by "translation"?
13. How are inactive portions of DNA locked up by the cell? How do they get unlocked? What is a "response element"? What is an "intron"? What is an "exon"?
14. What are histones? What promotes DNA binding? What promotes DNA unraveling from the histones? What is meant by "the histone code"?
15. What is DNA methylation? Where does it occur? What is the consequence of DNA methylation? How does DNA methylation pattern get passed on to daughter cells?