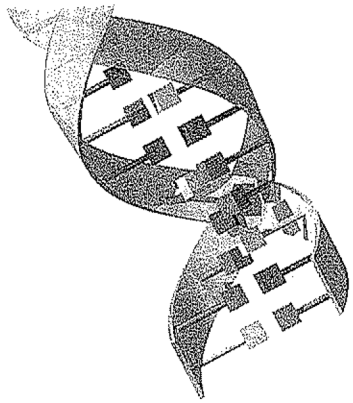


CHAPTER 17
Nucleic Acids

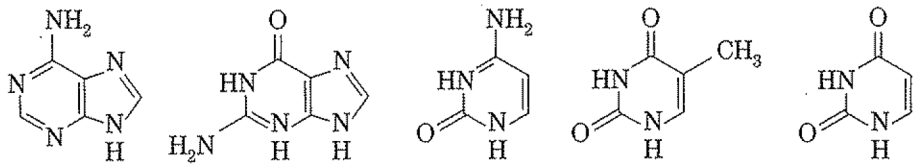
The virtual explosion of biotechnology in the last two decades is ample evidence of the central importance of nucleic acids to chemistry as well as biology. These biopolymers are in the public eye because of their biological and medical promise and have generated serious discussion in the areas of economics, politics, sociology, ethics, and theology.

Nucleic acids are the constituents of our genes. Although their fundamental structures are relatively simple, the process of nucleic acid or gene replication and the translation of the genetic message into tens of thousands of proteins for which it codes is a complex process. This chapter will only touch the surface of a complicated and growing field. Individual monomer units, or nucleotides, as well as dinucleotides also serve as energy carriers and oxidation/reduction agents in metabolism and as chemical messengers relaying life-regulating information within and between cells.

17.1 The Chemical Structure of Nucleic Acids

As the term **nucleic acid** suggests, these biopolymers are acidic in nature and are found in the nucleus of the cell as well as in chloroplasts and mitochondria. The fundamental unit of the polymer is the **nucleotide,** which consists of a heterocyclic base, a sugar, and inorganic phosphate.

Five common heterocyclic bases are found in **DNA** (deoxyribonucleic acid) and **RNA** (ribonucleic acid): two are related to the bicyclic base purine and three to the monocyclic base pyrimidine. Three of the five bases are common to both DNA and RNA, while the two remaining pyrimidines help to distinguish DNA from RNA.



Found in both DNA and RNA

Found in DNA

Found in RNA

nucleic acid
biopolymer whose monomer unit, a nucleotide, consists of a heterocyclic base, a sugar, and a phosphate group

nucleotide
the monomer unit of a nucleic acid consisting of a purine or pyrimidine base covalently bonded to a ribose or deoxyribose unit, which in turn is bonded to a phosphate group

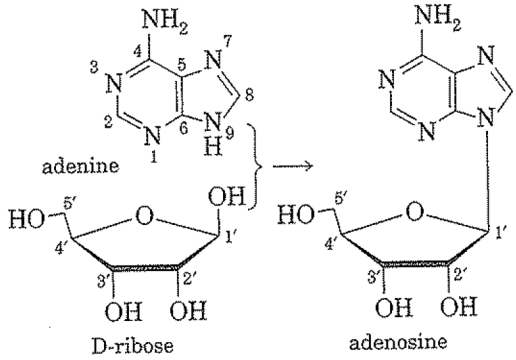
DNA
deoxyribonucleic acid

RNA
ribonucleic acid

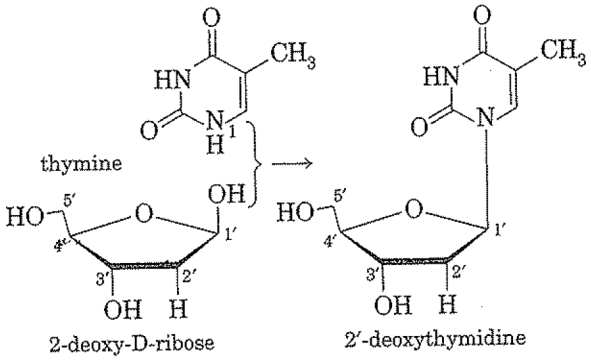
There are variations found in the structures of RNA bases, and DNA can undergo a natural process of methylation. However, the bases mentioned above are the ones found in greatest abundance and are those upon which the genetic code was established.

nucleoside
heterocyclic base bonded to a ribose or deoxyribose unit

These bases are bonded to the monosaccharide, forming a **nucleoside.** Adenine and guanine are attached through the N-9 position of the purine ring system to the hemiacetal group, C-1, or more properly C-1′, of deoxyribose (for DNA) and ribose (for RNA). Note that the glycosidic linkage from the sugar to the base has the β-configuration. The pyrimidines are linked through position N-1 in the ring.



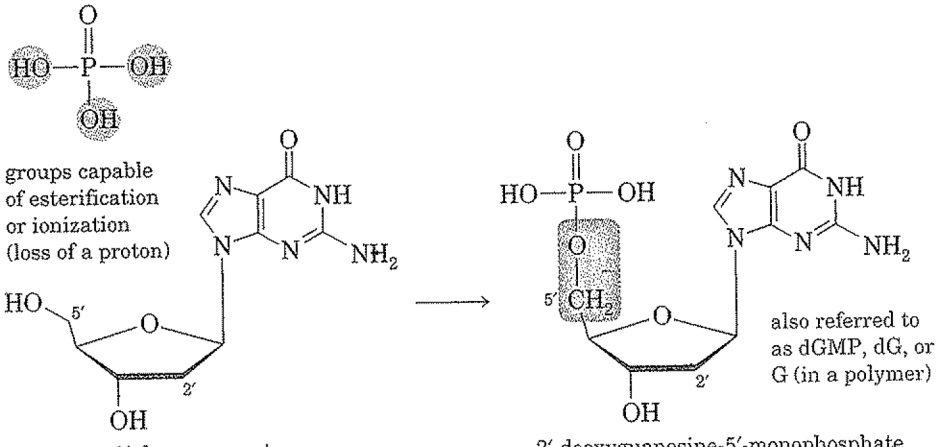
Formation of a Ribonucleoside



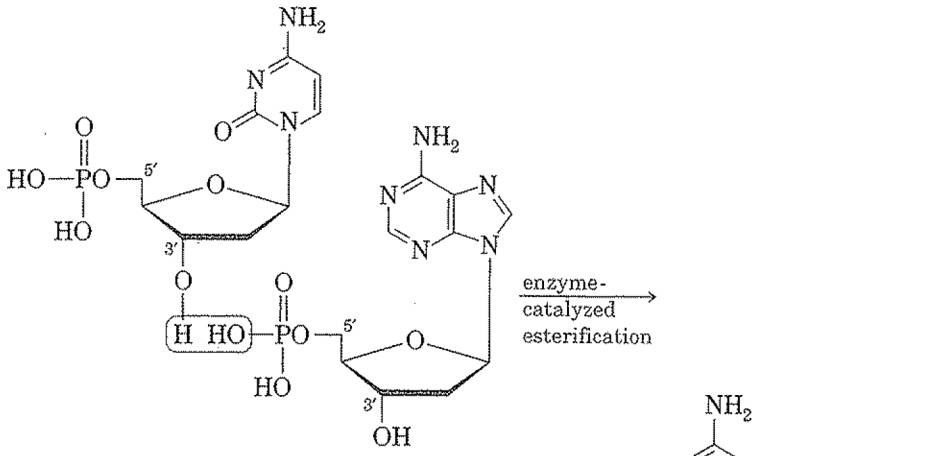
Formation of a 2′-Deoxyribonucleoside

The other nucleosides are named 2′-deoxyadenosine, 2′-deoxyguanosine, and 2′-deoxycytidine for the DNA combinations and adenosine, guanosine, and uridine for the RNA components.

The nucleoside is then esterified through the sugar to a phosphate group to make a nucleotide. Phosphoric acid is a triprotic acid and can react as an acid with each ionizable hydrogen. It can also form one or more ester bonds with available alcohol groups. A nucleotide has a phosphate esterified to position 5′ of the ribose or deoxyribose.



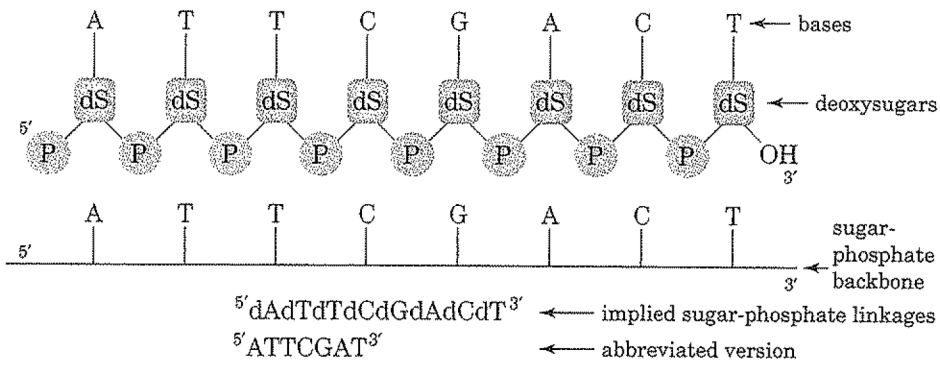
Because of the complicated structure of a nucleotide, shorthand notations are used to designate the bases, nucleosides, and nucleotides. The polymerization of nucleotides into nucleic acids involves the formation of a phosphodiester bridge from the 3′ hydroxy group of one nucleotide to the 5′ phosphate of another.



a dinucleotide

As this enzyme-catalyzed polymerization proceeds, a regular array develops consisting of a phosphate-sugar "backbone" from which the heterocyclic bases protrude. The backbone can be shown or simply assumed to be the way in which the bases are connected.

Representations of a polynucleotide chain



The result is an **oligonucleotide** (just a few units), a **polynucleotide,** or a nucleic acid. You should become familiar with each way to represent polynucleotides. Notice that in the bottom representation above that deoxyribose (dS) is not indicated. The presence of thymine (T) is enough to identify the sequence as DNA.

oligonucleotide
polymer containing a few nucleotide units

polynucleotide
polymer containing more than a few nucleotide units

In all of the phosphates linking the nucleotides, one -OH group remains *underivatized.* The high K_a for this group allows it to deprotonate at physiological pH, producing an anion (−). This means that the phosphate-sugar backbone is highly negatively charged and hydrophilic.

GETTING INVOLVED

- ✓ Why are the purines and pyrimidines categorized as bases?
- ✓ What are two major differences between the structures of a ribonucleotide and a deoxyribonucleotide?
- ✓ Be sure that you know how to link a purine or a pyrimidine base to a ribose unit and then to a phosphate. Draw the structure of mononucleotide with uracil.
- ✓ Draw the structure of a dinucleotide containing a uracil mononucleotide and a cytidine mononucleotide with a 5′, 2′-phosphodiester link.

Problem 17.1

Compare the functional groups in carbohydrates, proteins, lipids, and nucleic acids. How are they alike and how do they differ?

Problem 17.2

Draw three other representations for the structure of the polydeoxynucleotide **GTCC**.

Problem 17.3

Write the complete structure for the polyribonucleotide **UCAG**.

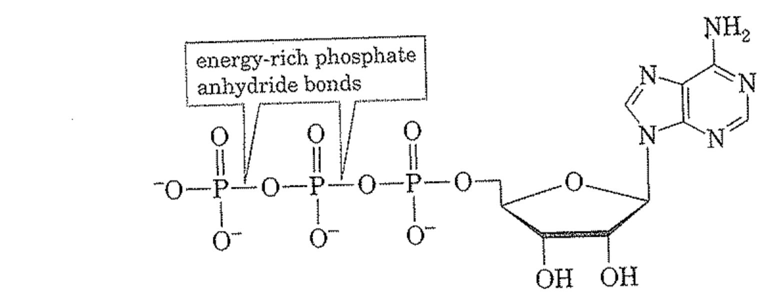
See related problems 17.10, 17.13, 17.14.

17.2 Other Structures Involving Nucleotides

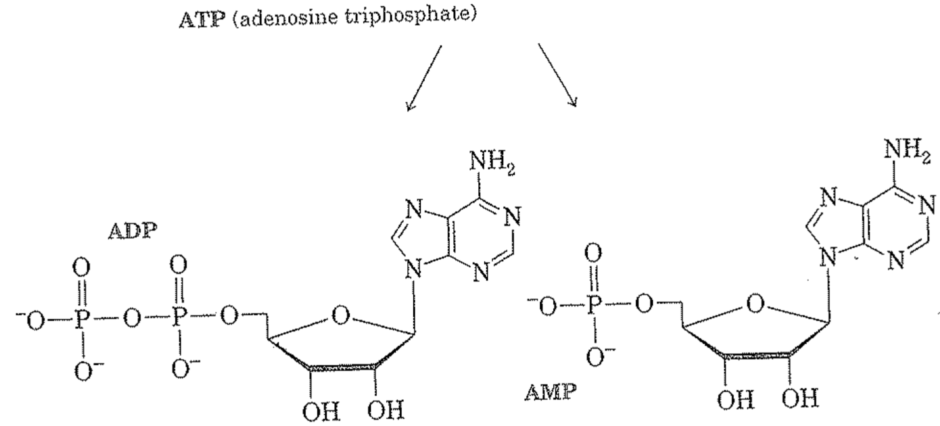
A. Energy Intermediates

ATP, ADP, AMP adenosine tri-, di-, and monophosphate; energy carriers in metabolism

Nucleotide di- and triphosphates contain high-energy phosphate anhydride bonds that are made during metabolic catabolism (nutrient breakdown) and used during the process of biosynthesis. **Adenosine triphosphate (ATP)** is the best known and most ubiquitous of these molecules, although guanosine and cytidine triphosphates are also important in metabolic processing.



ATP (adenosine triphosphate)



GETTING INVOLVED

✓ Draw the structures for UDP and CMP.

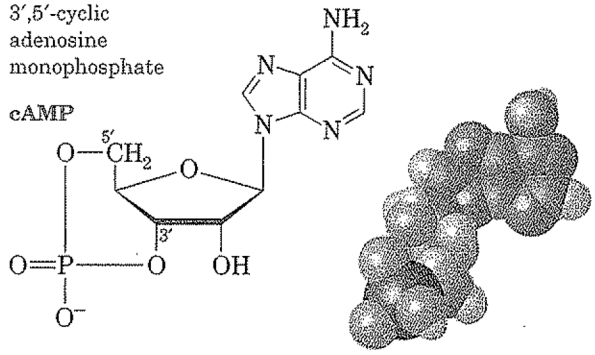
Problem 17.4

A product of the hydrolysis of ATP to AMP is inorganic pyrophosphate (PP_i) or $HP_2O_7^{3-}$, wherein the anhydride bond between the two phosphates has not been broken. Draw the structure of PP_i in its ionized state.

B. Chemical Messengers

The communication of hormone- and nerve-mediated signals can also involve the formation of intracellular messengers known as cyclic nucleotides. 3′,5′-cyclic AMP or **cAMP**, and cGMP are such biomessengers.

cAMP
cyclic adenosine monophosphate; chemical messenger



GETTING INVOLVED

Problem 17.5

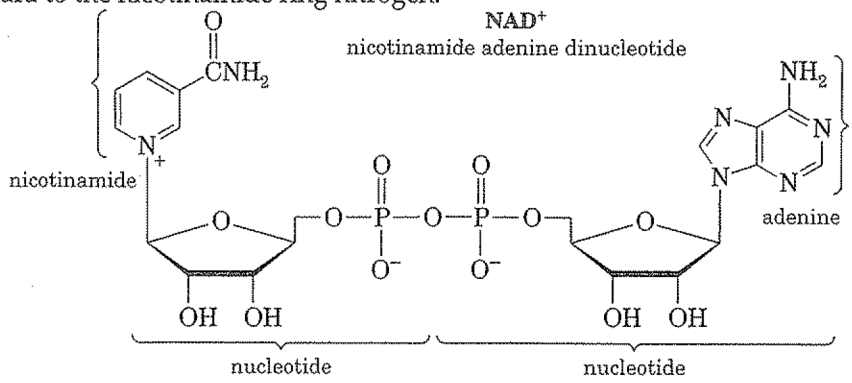
Draw the structure of 3′,5′-cyclic guanosine monophosphate, cGMP. Should this molecule be acidic, basic, or neutral? Explain your answer.

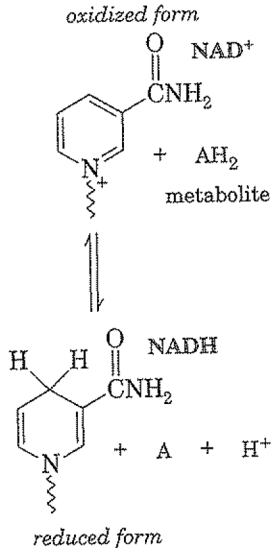
See related problem 17.15.

C. Redox Factors—Nucleotide Vitamins

Several variations of nucleotides participate in enzyme-catalyzed reactions as cofactors. They contain water-soluble vitamins; that is, they are organic compounds that are essential to life (vitamin), water-soluble, not synthesized within the body, and obtained through the diet. One of these is nicotinamide, not related to nicotine, which is found joined to AMP as nicotinamide adenine dinucleotide or NAD. NAD has two redox forms: **NAD^+** (shown below) and **NADH** in which the oxidized NAD^+ has undergone a hydride reduction at a position para to the nicotinamide ring nitrogen.

NAD^+/NADH
nicotinamide-adenine dinucleotide (oxidized/reduced forms); oxidation-reduction cofactor in metabolism



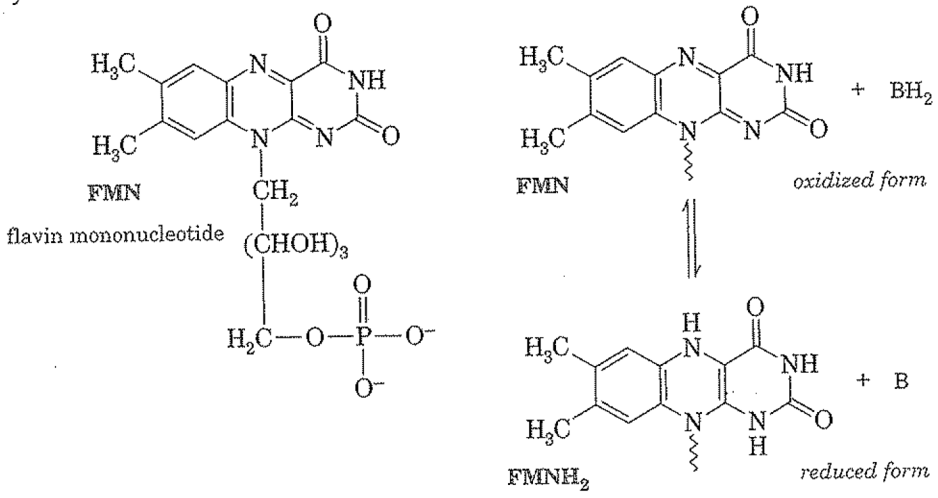


Nicotinamide, in the form of the pyridine carboxylic acid niacin, is found in yeast, meats, and wheat germ. Its absence from the diet results in pellagra. Pellagra's symptoms include diarrhea, indigestion, and dermatitis. It can be fatal if left untreated. Excessive intake of niacin causes flushing of the skin and may lead to liver damage.

Riboflavin, vitamin B_2, consists of a heterocyclic base called flavin and the reduced form of ribose, ribitol. In the body riboflavin can be joined with a phosphate group to form flavin mononucleotide (FMN). FMN can also be linked with AMP to produce flavine adenine dinucleotide **(FAD).** Both FMN and FAD can undergo reversible oxidation-reduction with the elements of molecular hydrogen adding in a 1,4 manner to the part of the system shown here. The products are abbreviated as $FMNH_2$ and **$FADH_2$.**

FAD/$FADH_2$
flavin adenine dinucleotide (oxidized/reduced forms); oxidizing reducing agents

A deficiency of B_2 produces dermatitis of the face, an inflammed tongue, and eye disorders.



GETTING INVOLVED

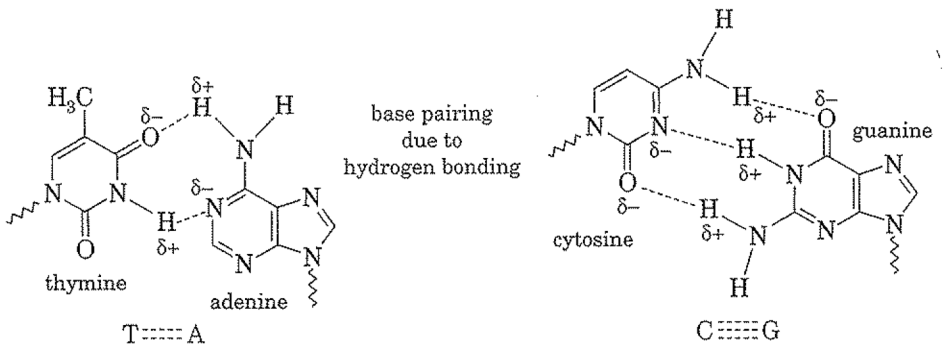
✓ Flavin adenine dinucleotide (FAD) is also an oxidation–reduction cofactor. Draw the structure of FAD.

17.3 The Hierarchy of Nucleic Acid Structure

In Chapter 16 we saw that proteins have several levels of superstructure that depend on the ability of various functional groups to participate in covalent and noncovalent interactions. The most important noncovalent intermolecular interaction is hydrogen-bonding. The same types of interactions are exhibited by nucleic acids. The bases establish hydrogen-bonding patterns that result in the well-known double-stranded helix of DNA. Hydrogen-bonding between bases is also used to direct the replication of genetic material and the transcription and translation of the DNA coded message for the production of proteins through RNA.

A. DNA Structure: The Double Helix

It had been known prior to 1953 that the mole ratio of adenine to thymine and guanine to cytosine was usually 1, no matter the source of the DNA. The numbers of the individual bases varied, but that ratio essentially remained the same. The reason for this depends on a recurring phenomenon in chemistry, hydrogen-bonding. Looking at the structures for the bases, we can see that the opportunity for hydrogen-bonding exists since there are electronegative oxygens on the carbonyl groups, electronegative ring nitrogens, and electropositive hydrogens on the amine or imine groups.



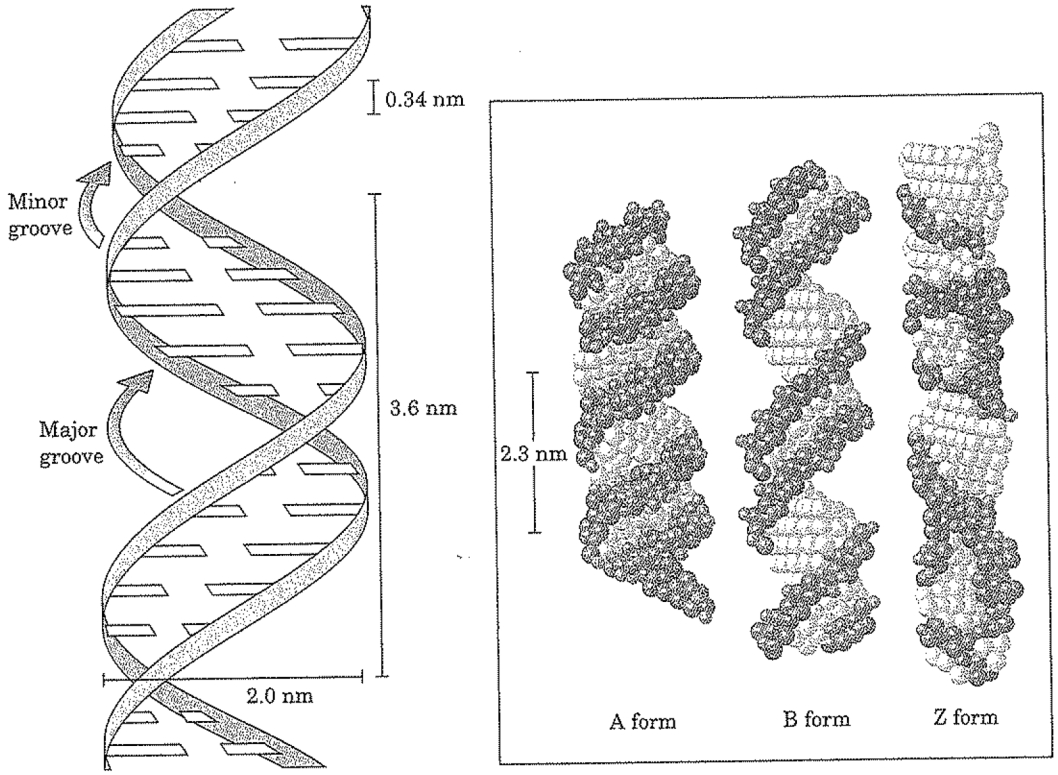
The hydrogen-bonding between adenine and thymine, guanine and cytosine, is called complementary **base-pairing.** Maximum base-pairing occurs when A and T are joined by two hydrogen bonds and G and C by three.

base-pairs
complementary bases that can hydrogen-bond to each other; A ≡ T, G ≡ C, A ≡ U

The actual physical orientation of the entire DNA polymer was not known until 1953, when James Watson, Francis Crick, and Maurice Wilkins interpreted X-ray data (produced by Rosalind Franklin) to indicate a **double-stranded helix.** The Watson–Crick hypothesis, for which the three won the Nobel Prize in 1962, shows two complementary hydrogen-bonded strands aligned in an antiparallel manner.

double helix
Watson–Crick model of DNA in which the heterocyclic bases are oriented toward the interior axis and the sugar-phosphate backbone on the outside of the helix

The two polynucleotide chains are twisted around each other with the bases oriented toward the center axis of the helix and the sugar-phosphate backbone on the outside of the helix, exposed to the aqueous environment of the cell. There are three commonly known forms of the helix called A, B, and Z. These differ in the number of water molecules interacting with the helix as well as in the orientation of the bases to the center of the helix, rise of each turn, and overall handedness of the helix (Figure 17.1). The B helix is the one assumed to exist in water solution and was proposed by Watson and Crick. It is right-handed and rises 34 angstroms per turn (1 Å = 10^{-8} cm) or 3.4 nanometers, containing ten nucleotide bases per complete turn. Two grooves appear in the overall structure, one wider than the other. The wider is known as the *major groove*, whereas the smaller is the



(a) General architecture

(b) A, B, and Z forms

Figure 17.1 The double helix of DNA. (Adapted from Lehninger, Nelson, and Cox, *Principles of Biochemistry*, 2nd ed. Used with permission.)

minor groove. Other biomolecules interact with DNA in these grooves, helping in its function. For example, basic proteins known as histones stabilize DNA by forming charged (+/−) complexes with it. Single-stranded DNA (ssDNA) does exist in certain organisms but is not common.

genome
the entire genetic makeup of an organism

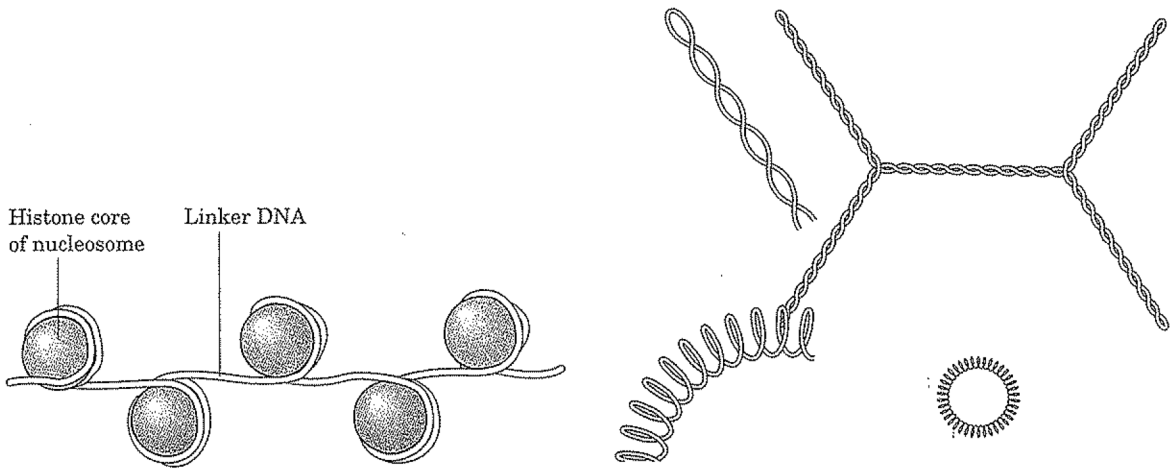
DNA is the material of the **genome**, or hereditary material, of all living organisms from bacteria to human beings. Since the entire human genome, all estimated 100,000 genes, must fit into each cell, there must be much more efficient packing of a double helix. DNA can form into circular rings in lower organisms and is linear in higher organisms. The helices, whether in rings or linear form, can also intertwine, forming **supercoils** (Figure 17.2). In fact, the strain that is induced in superwound coils can aid in transferring genetic information. Supercoiling as well as wrapping around **histone** proteins allow for the compaction necessary to fit the total length of DNA into a single cell.

supercoil
form of compacted DNA

histone
basic protein associated with nucleic acids in eukaryotic organisms

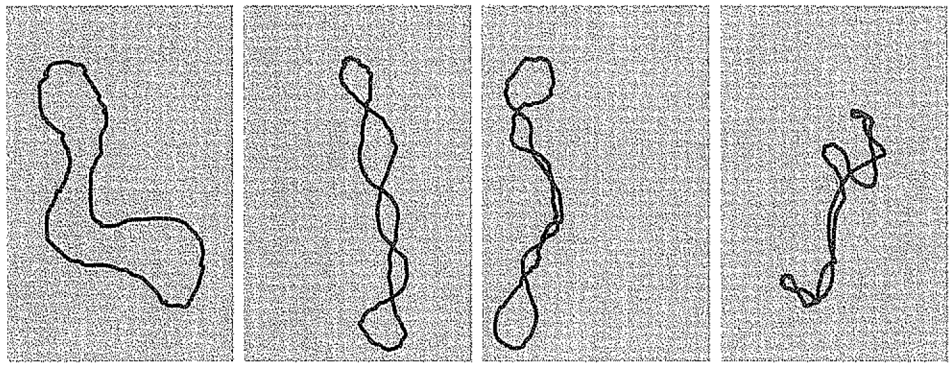
GETTING INVOLVED

- ✓ RNA contains uracil, which forms base pairs with adenine. Draw a uracil–adenine base pair, indicating the possible sites of hydrogen-bonding.
- ✓ What types of macro- (large) structures are formed by DNA?
- ✓ What is a genome?
- ✓ Briefly mention a few ways in which DNA differs from RNA in its fundamental chemical structure and in its superstructure.
- ✓ Describe the process of base-pairing in DNA
- ✓ How does base-pairing support the theory of Watson and Crick?
- ✓ Is the double helix the largest assembly possible for DNA?



(*a*) Histone-DNA complexes

(*b*) Circular DNA and coiled coils



(*c*) Relaxed and supercoiled DNA

Figure 17.2 Compaction of nucleic acids through coils and histone interactions. (Adapted from Lehninger, Nelson, and Cox, *Principles of Biochemistry*, 2nd ed. Used with permission.)

Problem 17.6

If the histones are basic proteins, which amino acids should occur in large proportion in them? With which negative groups of the polynucleotide will the histones interact?

B. RNA Structure

Although double-stranded RNA (dsRNA) does exist, it is not common. Most RNA is single-stranded (ssRNA), forming a greater variety of superstructures than DNA, which suits its different roles. There are three general types of RNA: messenger RNA **(mRNA)**, ribosomal RNA **(rRNA)**, and transfer RNA **(tRNA)**. We will discuss the overall structures of RNA as we explain their functions.

mRNA
messenger RNA; contains codons for the construction of protein

rRNA
ribosomal RNA; RNA associated with proteins to form the ribosome

tRNA
transfer RNA; brings amino acids to the ribosome for protein synthesis; contains anticodons

17.4 The Genetic Code

DNA, found primarily in the nucleus but to a small extent in mitochondria and chloroplasts, is the ultimate carrier of the genetic code encrypted in the sequence of its bases. RNA acts as a transcribing agent, copying the nuclear DNA message and carrying it to the cytoplasm, where amino acids are assembled into the

correct sequence. The proteins that resul
in all the other functions mentioned in C
a brief overview of this process.

A. DNA Replication

replication
process of duplicating DNA

In order for a single cell to grow into a complete organism that passes on genetic information to ensuing generations, it is necessary that DNA be able to reproduce itself. This process, known as **replication,** first involves the uncoiling of the double helix, which occurs in sections. Once uncoiled, a DNA strand is base-paired to the corresponding dNTPs (deoxyribonucleotide triphosphates). Since nucleic acid biosynthesis occurs continuously from the 5′ to the 3′ end, only one strand of the DNA helix is made in one continuous polymer (this will complement the 3′ ---------→ 5′ original DNA strand). The second strand of the opened DNA helix is replicated from shorter 3′ ---------→ 5′ fragments produced from the base-pairing to the second original DNA strand. Lengths of RNA known as **primers** serve as starting points for DNA formation. After the DNA polymer has been added to the primers, the primers are eventually removed.

primer
lengths of RNA that serve as starting points for DNA formation

semiconservative replication
DNA that is composed of one parent strand (template) and one daughter strand (formed from base-pairing) with a new set of bases

DNA replication is said to be **semiconservative** in that the second generation double helix is composed of one "parent" or original DNA strand and one "daughter" strand (see Figure 17.3). In conservative replication the parent strands would have recombined, and the daughter strands would have formed an entirely new DNA double helix.

Replication proceeds through a complex interplay of DNA, RNA, deoxyribonucleotide triphosphates, enzymes, and other biochemical factors. Enzymes include DNA polymerases, helicases, primases, and **ligases,** which not only sew together the genetic fragment but also ensure an extraordinary fidelity in replication. Sequences are proofread and can be corrected as the DNA is incorporated into a new copy of the genome. Binding proteins help to keep the helix open.

ligase
enzyme that connects pieces of polynucleotides

GETTING INVOLVED

Problem 17.7

Attempt to solve a DNA sequence given the following information from the electrophoresis results of a Sanger-type experiment (See Connections 17.1).

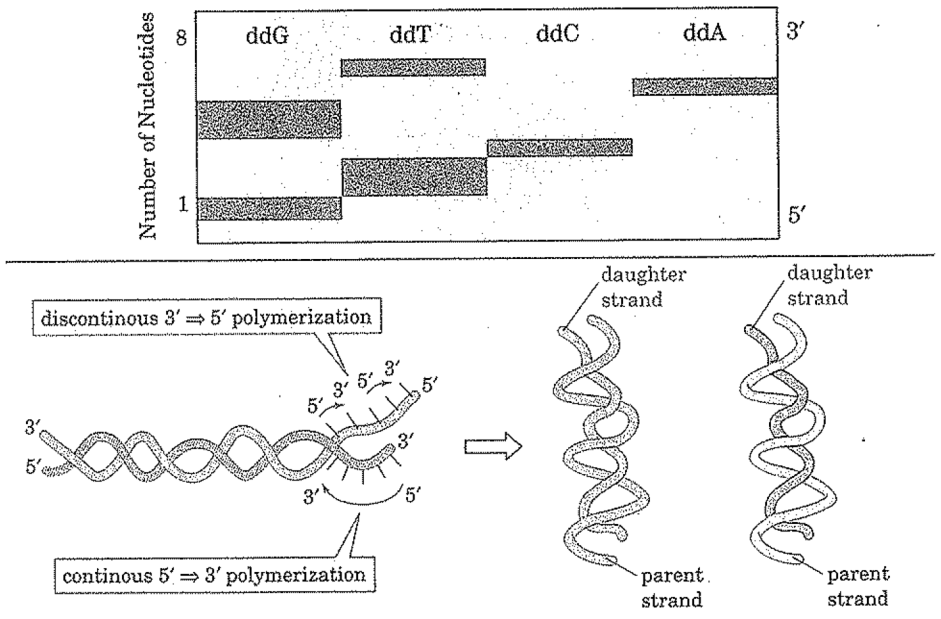


Figure 17.3 Semiconservative mode of replication of DNA.

Connections 17.1

www.prenhall.com/bailey

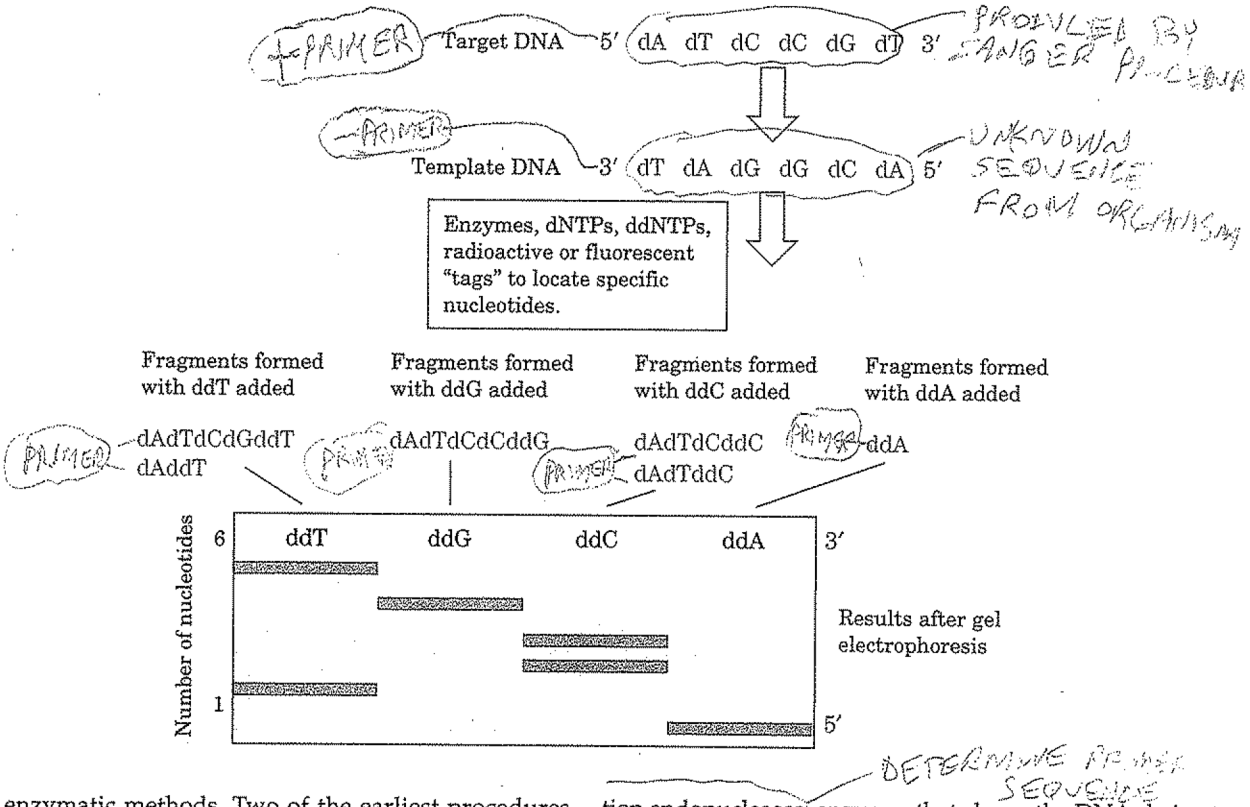
The Human Genome Project

The scientific community embarked on one of its greatest projects during the last decades of the twentieth century. The quest was to decipher the DNA sequence for the entire human genome—all 3 billion base pairs located in the genes of the 46 chromosomes. It will be left to the scientists of the twenty-first century to determine the functions of the protein products coded in the DNA. As might be expected, the DNA sequences of simpler organisms have already been under investigation and their results will lead the way to the ultimate goal.

The sequencing process is a combination of chemical and enzymatic methods. Two of the earliest procedures were that of Sanger and that of Gilbert and Maxam.

In the Sanger process a template is synthesized that is complementary to a single strand of DNA (see diagram in Problem 17.7). DNA polymerizing enzymes facilitate making the primer copy. This template is then used in further DNA replication; however, at this point small amounts of specific 2′,3′-dideoxyribonucleotides are added (ddATP, ddGTP, ddCTP, ddTTP) along with the normal 2′-deoxynucleotide triphosphates (dNTPs). Without the 3′ hydroxy group the ddNTPs will stop the lengthening chain as they are incorporated. The result is a mix of polynucleotides of varying lengths. These polymers can be separated using electrophoresis on the basis of their molecular weights. The shorter fragments represent the 5′ end of the primer while the longer segments are complements to the 3′ end. In the following diagram, try to coordinate the electrophoretic pattern with the fragments and then the original sequence.

The Gilbert–Maxam method uses specific radioactive labels to "tag" areas of the DNA and then applies restriction endonucleases, enzymes that cleave the DNA chain at specific nucleotides, to break the polymer into smaller fragments. The radioactive markers help in the elucidation of the sequence.



The ability to locate specific loci in chromosomes has produced many diagnostic tests for genetic conditions such as sickle-cell anemia, cystic fibrosis, Huntington's disease, and various types of cancer. Gene therapy techniques are in their infancy and hold great promise for the twenty-first century.

B. Transcription and Translation

transcription
process of making mRNA by complementary base-pairing of ribonucleotides with a piece of DNA chain

translation
process that involves mRNA binding to ribosomes and base-pairing with specific tRNAs holding amino acids; the end product is a protein

sense (+) DNA strand
strand of DNA double helix that is transcribed to mRNA

antisense (−) DNA strand
strand of DNA double helix that is not transcribed to mRNA

exon
expressed sequence; portion of DNA (and mRNA) that is transcribed and translated into protein

intron
an intervening sequence; portion of DNA (and mRNA) that is not transcribed and translated into protein

The processes whereby the genetic code is interpreted to form protein are called **transcription** and **translation.** First the DNA sequence is transcribed into messenger RNA (mRNA) in the nucleus. This process involves the base-pairing of ribonucleotide triphosphates (NTPs) with an unwound portion of the template or (−) or **"sense" strand** of the DNA helix and then enzyme-catalyzed polymerization. Only one DNA strand is transcribed at a time. The untranscribed strand is called the coding, (+), or **antisense strand,** because its sequence will be the same as that for the mRNA produced, with the substitution of a U for a T. The DNA sequence on the template strand is read 3′ to 5′ while the mRNA is synthesized 5′ to 3′. However, when correlations are made between the mRNA and its parent DNA, it is the antisense (−) strand that is usually referred to (see Figure 17.4).

The DNA code for a protein is usually found in several locations, either along one chromosome or on separate chromosomes. Not all of the DNA sequence codes for protein. Some segments are in between coding sequences. The coding sequences are known as **exons** and the intervening sequences are called **introns.** There are usually fewer than 10^3 nucleotides per exon with most in the range of 100 to 200 hundred base pairs. Intron lengths have a much wider variation of anywhere from 50 to 20,000 nucleotides. When an mRNA (the primary transcript) is first made it contains the complementary sequence of both exons and introns. Richard Roberts and Phillip Sharp won the 1993 Nobel Prize in medicine for their 1977 discovery of "split genes," that is, introns and exons. The primary transcript mRNA is edited to remove intron pieces (Figure 17.5). The cutting and splicing of exons has provided an interesting insight to the evolution of proteins and allows us to understand how variations of one type of protein can be found. The processes of gene duplication, mutation, and gene fusion lead to the production of large families of proteins related in structure and/or function either as a whole or in their domains.

The heavy and light chains of antibodies, for example, are made from several exons that are mixed and matched, resulting in many proteins that can respond specifically to the large number of nonself entities encountered by a human. It should be noted that, almost without exception, bacterial cells contain only exons and no introns.

mRNA carries a complementary sequence to the DNA exons onto the ribosomes that are associated with the endoplasmic reticulum in the cytoplasm of the cell. Specific sequences of three bases relate to the amino acid. Since there are only four DNA and four RNA bases, the correlation cannot be one-to-one. Even a two-to-one relationship would produce a code for only 16 of the 20 common

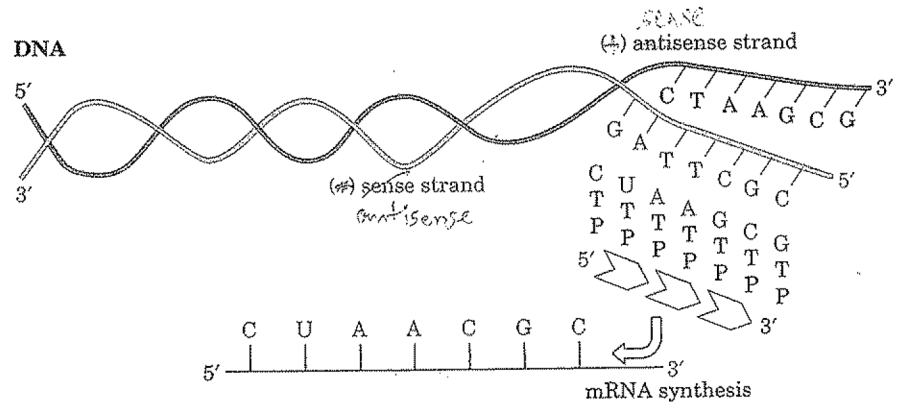


Figure 17.4 The transcription of DNA to RNA.

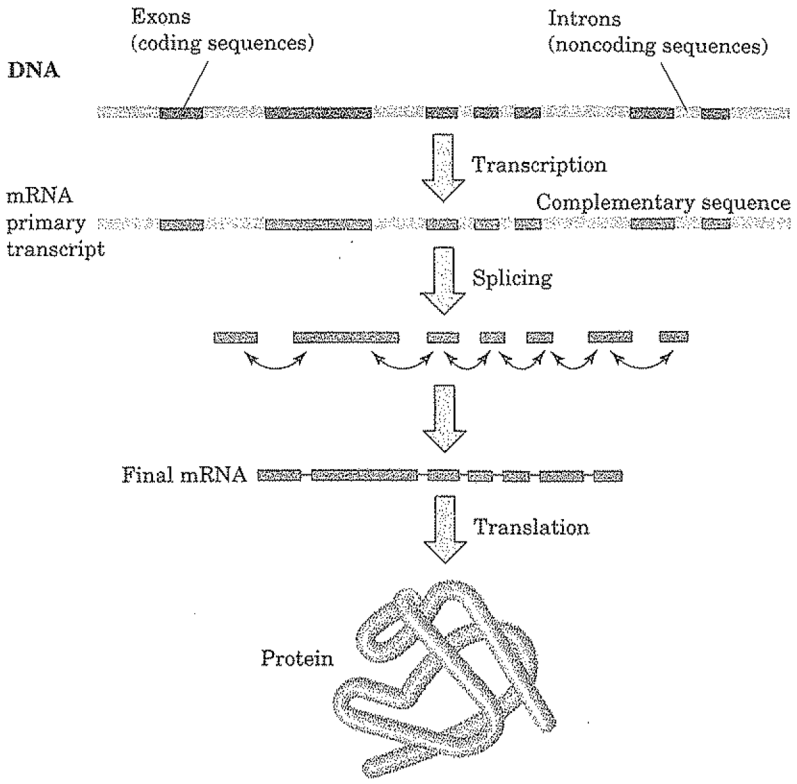


Figure 17.5 The removal of intron sequences from mRNA.

amino acids. However, at three-to-one, 64 different combinations are possible. This means that some amino acids may have more than one "code word." In actuality there are 61 amino acid **codons** (in mRNA) and 3 codons for initiation and termination of the protein chain (sometimes called *nonsense codons*). See Table 17.1 for a list of mRNA codons and their corresponding amino acids or start/stop directions. The availability of more than one codon for most amino acids is called the **degeneracy** of the code. Looking at Table 17.1, one can see that the first base in codons for the same amino acid is usually the same or almost so. The third can be highly variable and is referred to as the "wobble" base.

codon
three-base polynucleotide sequence of mRNA corresponding to an amino acid or protein synthesis directive (start or stop)

degeneracy
characteristic of the DNA code; there is more than one three-base code for most amino acids

The codons found in bacteria are the same as those seen in higher organisms. It seems then that the genetic code is nearly universal for all organisms, whether prokaryotic or eukaryotic.

Ribosomes are large complexes consisting of ribosomal RNA (rRNA), protein, and cofactors. Each ribosome has two major subunits, 50s and 30s in prokaryotes (lower organisms) and 60s and 40s in eukaryotes (higher organisms). Ribosomes provide the environment for translation of the nucleic acid code to a protein amino acid sequence. Ribosomal RNA (rRNA) provides a scaffolding upon which enzymes can interact with the key factors in the manufacture of proteins. mRNA locates itself in a cleft formed by two major portions of the ribosome. From 10 to 100 ribosomes can be associated along one strand of mRNA giving rise to a polysome. More than one protein molecule can thereby be synthesized simultaneously.

ribosome
RNA-protein complexes that serve as the environment for protein synthesis

The third major type of RNA, transfer RNA (tRNA), joins the ribosome-mRNA super complex, carrying with it amino acids for the protein biosynthetic process. tRNA is single-stranded and has a three-dimensional structure that appears elongated and L-shaped, stabilized by hydrogen bonding between base pairs within the molecule. At the 3′ arm of the molecule an amino acid is attached.

Table 17.1 Messenger RNA Codons

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

There are more than 30 different tRNAs for the 20 common amino acids. Some tRNAs are more amino acid-specific than others. Located at a polynucleotide loop on the other end of the tRNA is a three-base **anticodon** sequence complementary to the three-base codon on the mRNA (Figure 17.6).

anticodon
three-base polynucleotide sequence of tRNA that base-pairs with a specific codon

The amino acid-bearing tRNA becomes noncovalently attached to a ribosomal site called the "P" site. For bacteria the "start" codon is for an N-formyl methionine (methionine with a formyl group at the amino N), whereas for eukaryotic organisms the sequence begins with the codon for a methionine. A second tRNA then binds to an adjacent "A" site on the ribosome-mRNA complex.

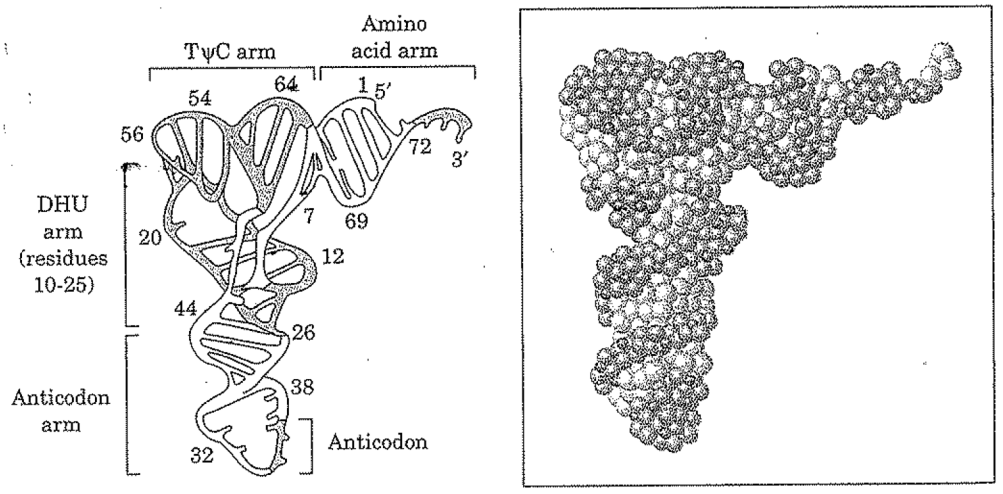


Figure 17.6 Models of transfer RNA (tRNA). (Adapted from Lehninger, Nelson, and Cox, *Principles of Biochemistry*, 2nd ed. Used with permission.)

The first amino acid, which will be the N-terminus of the protein chain, is linked to the second. This results in the liberation of the first tRNA from the P site and the movement of the A site tRNA to the P site, now with a dipeptide attached to the P site. A third amino acid-carrying tRNA binds to the vacant A site and the procedure is repeated. In this manner the protein grows until a "stop" codon is encountered. Then the protein is released from the complex and is transported to cellular areas for modification and/or incorporation into the cell matrix (see Figure 17.7).

Most proteins are biosynthesized with an N-terminal hydrophobic sequence of 15 to 30 amino acids, called the **signal sequence,** which facilitates the targeting of a protein as well as its passage through the membranes of organelles, where storage or posttranslational modification takes place. As the protein is threaded through the membrane, the signal sequence is enzymatically removed.

signal sequence
N-terminal protein sequence

Alterations in the nascent protein are called **posttranslational modifications** and may include removal of the N-formyl methionine or methionine that started the chain, addition of carbohydrate (glycosylation), methylation, esterification, phosphorylation, isoprenylation, or cleavage of the single chain into multiple chains to produce a fully functional protein.

posttranslational modifications
chemical changes made on a completed protein such as the addition of lipid or carbohydrate or the cleavage of the polypeptide chain

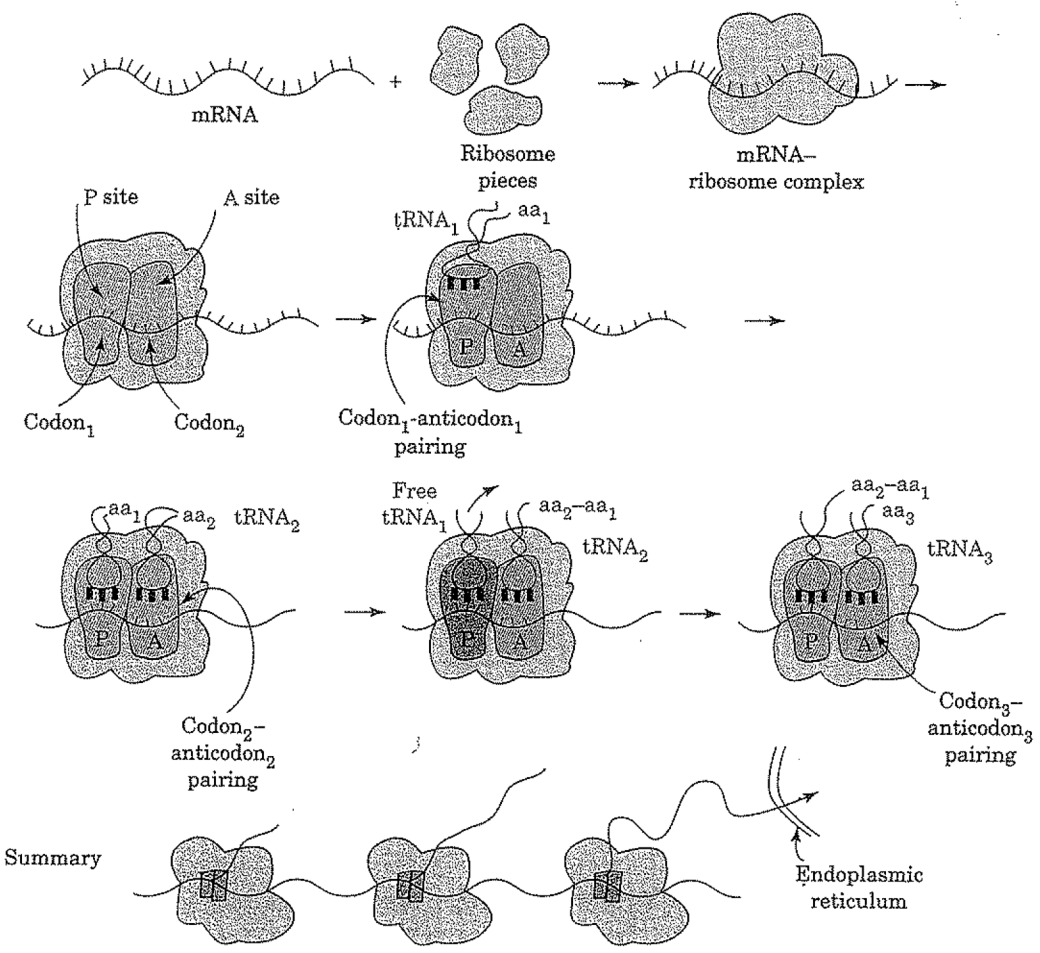


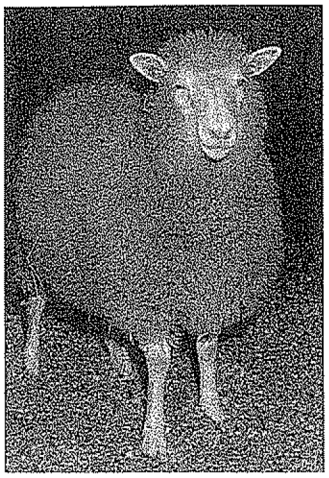
Figure 17.7
Translation of mRNA into protein.

GETTING INVOLVED

- ✓ What is meant by a codon? anticodon?
- ✓ What makes the genetic code "degenerate"?
- ✓ Which amino acid codons are not degenerate?

Problem 17.8

What would be the tRNA anticodon sequences for the following mRNA codon: 5'-GGU ACU CCC UGA-3'? Write the tetrapeptide that is being coded. What is the original DNA sequence for both the sense and antisense strands?



Dolly, a sheep cloned from an existing animal

Problem 17.9

Mistakes in making mRNA occur by inserting or deleting bases.

(a) What would happen to the polypeptide sequence coded for in Problem 17.8 if an A were inserted in between the A and C of the second codon or the A were deleted from the second codon?

(b) What would happen if the codon GUG was inserted between the third and fourth codons?

See related problems 17.10, 17.16, 17.17.

17.5 Characteristics of Transcription and Translation

There are several key points that should be remembered about the genetic code and its direction of protein synthesis.

1. The genetic code is nearly universal. The three-base mRNA codons and their anticodons can be found in prokaryotic as well as eukaryotic organisms.
2. The code is degenerate. Most amino acids have more than one codon.
3. One RNA base sequence is usually read in the same way to produce the same protein in a repeatable manner, that is, there are no overlapping codes. There are a few exceptions to this feature, but it generally holds true.
4. There is a great deal of reliability in the process, but mutations can occur that may or may not lead to viable proteins.
5. The biosynthesis of proteins is energy-consuming.

mutation
a change in the nucleotide sequence of a DNA molecule which may or may not lead to an alteration in protein structure and/or function

17.6 Mutation of DNA

The structure of nucleic acids is sensitive to chemical and physical factors that are present naturally or may be introduced into cells through the environment. During the process of metabolism, free radicals are generated that may affect the reactivity of the nucleotide bases. Dimers of adjacent bases such as thymine occur. Hydrogen-bonding patterns can be altered by tautomerism induced by exposure to radiation or chemical agents. In addition, incorrect purine or pyrimidine bases may be inserted or bases may be deleted during the replication or transcription process, giving rise to stop codons, shifts in the reading frame, or substitutions that change the nature of the protein to be biosynthesized.

An organism has natural mechanisms to deal with many such mutations, for example, excision and replacement of dimers and double-checking the integrity of the reading frame. However, these mechanisms cannot cover all changes and can be overwhelmed when faced with a "flood" of mutational

Connections 17.2 www.prenhall.com/bailey

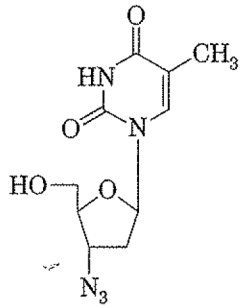
Acquired Immune Deficiency Syndrome (AIDS)

We are all aware of the challenge that AIDS presents to the entire world community. Caused by the human immunodeficiency retrovirus (HIV), this condition results in the collapse of the immune system. The mode of transmission is usually sexual or involves a mingling of blood between a host and a subject such as occurred in blood transfusions during the 1970s or can still occur in intravenous drug addicts through shared needles. Newborns can be infected with HIV from their mothers. HIV has two major forms, called I and II, in which the former was originally prevalent in the homosexual community in the United States and the latter was spread heterosexually in Third World countries. Today both forms are ravaging many continents.

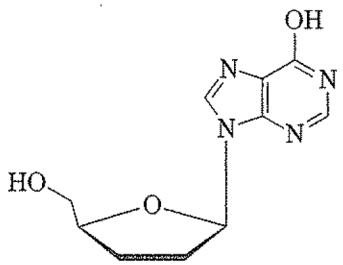
HIV invades a host organism and, as a retrovirus, incorporates its genome into the host DNA. After a period of time the virus destroys the T_4 cells of the immune system, leaving the body open to opportunistic infections such as pneumonia caused by *Pneumocystis carinii* and/or tuberculosis. Many of those afflicted eventually succumb to a form of cancer known as Kaposi's sarcoma.

The lives of HIV victims have been extended by drug treatment, but an AIDS vaccine is still in the future. As more is learned about HIV, its structures and modes of reproduction and action, research has produced veritable "cocktails" of drug mixtures whose targets attempt to stop the invasion of the virus as well as its incorporation into the host genome. There are three general classes of medications: reverse transcriptase inhibitors (RTIs), non-nucleoside reverse transcriptase inhibitors (NNRTIs), and protease inhibitors. The first two types slow the incorporation of viral RNA into the host DNA. The third type targets a protease enzyme used in the reverse transcription process. Further research into the protein coat of HIV is leading to new methods of approaching drug treatment. The structures of some RTIs and NNRTIs are shown below.

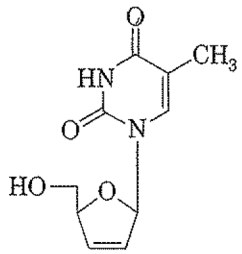
Reverse transcriptase inhibitors



AZT
3′ α-azido-2′3′-dideoxythymidine
(Retrovir®)

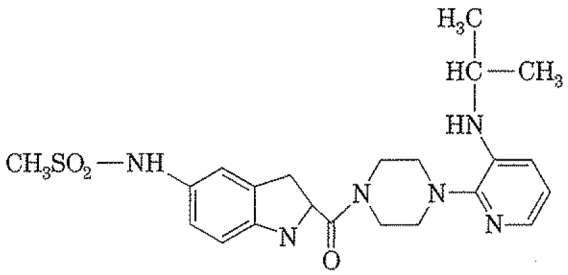


ddI
dideoxyinosine
(Videx®)

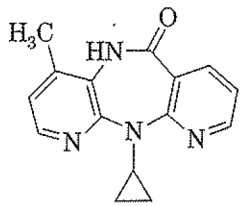


3TC
(Epivir®)

Non nucleotide reverse transcriptase inhibitors



Delaviridine mesylate



Nevirapine

events. The end result may be positive or negative: positive in that this is a natural way for the evolution and adaptation of an organism; negative in that it can result in the inability of the organism to survive. Site-directed mutagenesis has helped us to understand the roles of various amino acids in the structure of a pro-

tein; it is possible to design and produce DNA that will change a single specific amino acid or substitute or delete entire sections of a protein molecule. Protein domains and subunits have been shuffled and recombined into chimeras that retain the properties of the component parts. Animals have had specific genes "knocked out" in order to ascertain the importance of a specific protein in the overall viability of that organism. The possibilities are virtually limitless.

17.7 Viruses

virus
a nonbacterial infectious agent, that consists of DNA or RNA, a few proteins, and a protein coat

ssRNA
single-stranded RNA

dsDNA
double-stranded DNA

ss DNA
single-stranded DNA

reverse transcriptase
enzyme that can incorporate a virus RNA code into host DNA

retrovirus
RNA virus that can encode its genome into the DNA of a host organism, using the enzyme reverse transcriptase

Viruses are unique in that they are mainly nucleic acid with a few enzymes and a protein capsule or coat. The genetic material can be **ssRNA** as well as **ds** and **ss DNA.** The rabies virus, for example, has an ssRNA genome that codes for five proteins: a reverse transcriptase, a nucleoprotein, a phosphoprotein, a matrix protein that lines the inside of the membrane lipid bilayer, and a glycoprotein that constitutes the outer coat. Up to this point we have discussed the passing of genetic information from DNA to RNA to protein. How can a virus be replicated without DNA? The answer lies with a key viral enzyme known as **reverse transcriptase.** It takes the RNA message and puts its complement into the host cell's DNA. The protein synthesis machinery of the host cell is then used to propagate the viral RNA and its proteins. Once the virus particle or virion is assembled and enough virions are present, the host cell is lysed and the virus particles invade other cells. Such viruses are called **retroviruses.** Being RNA-based, they are subject to, and can survive, more mutational events. Therefore, the virus can change its protein coat frequently. This provides a constant challenge to the host defense systems. It is the reason that the flu virus is still with us, as is the rhinovirus causing the common cold. It is also part of the reason that to date it has not been possible to develop an effective vaccine against HIV, human immunodeficiency virus, which causes AIDS (acquired immune deficiency syndrome). Other viruses do not change as much and so have come under control by immunization. Smallpox has essentially been eradicated and poliomyelitis is following. Measles and mumps are controlled wherever there is an effective immunization program.

17.8 Oncogenes

oncogene
gene associated with cancer

"Onco-" refers to cancer. **Oncogenes** are genes connected with cancer, that is, they are related to the uncontrolled growth of cells. These "untamed" cells rob the adjacent tissue of blood, nutrients, and the space to exist. Eventually the organism cannot survive the invasion and dies. More than 100 cancer-related genes have been discovered to date. It is their mutation that gives rise to the proliferation of immature cells. The major questions concerning oncogenes are what proteins do they code for and what effect do these proteins have on cell growth and maturation? Cell growth is a balanced interplay of stimulation and suppression of growth factors in order to maintain the homeostasis or balance within an organism. Uncontrolled growth, therefore, can be a result of direct growth stimulation or the inhibition of suppression. One of the first oncogenes to be discovered and studied extensively is the p53 tumor suppression gene located on human chromosome 17. It has been associated with most types of human cancers from breast cancer to brain tumors. This is because p53 seems to be responsible for controlling the overall mutability of cells in the human genome. Most of the mutations in this gene involve missense codons, that is, changes in one DNA base, giving rise to single amino acid replacements. The alteration of

Connections 17.3 www.prenhall.com/bailey

DNA Fingerprinting

The intron sequences of DNA, far from being merely "junk DNA," are providing a means for the direct identification of individuals. These intron sequences contain regions of short nucleotide segments repeated many times. The length and number of repeats are the keys to identity. Since all cells of the body contain an individual's DNA, minute samples of blood, saliva, semen, hair follicles, or skin can be used for analysis. Of course, small tissue samples mean very small quantities of DNA. The Nobel Prize-winning discovery of an enzyme isolated from thermal hot springs, which will amplify submicroquantities of DNA, has revolutionized molecular biology and its applications, especially forensic science.

The polymerase chain reaction, or PCR, involves isolating nanogram amounts of DNA from tissue (old or new), cleaving it if necessary into manageable fragments, and subjecting it to multiple treatments of a polymerase enzyme in the presence of known DNA primers, with alternate heating and cooling cycles so that the original DNA is uncoiled, replicated, and renatured. The eventual products are separated by electrophoresis and compared. Individuals have unique sets of repeating sequences called polymorphisms that will identify them as reliably as a fingerprint can. Identical twins are the only organisms that have the same genetic sequences. Of course, a drawback to this ultra-sensitive method is the reality of contamination by investigators or others on the scene. Consequently, heightened security and improved sampling techniques are evolving in this new scenario.

DNA fingerprinting has helped to convict rapists and killers as well as exonerate death-row inmates. It has revealed the indiscretions of presidents and located the descendents of ancient populations. It can also be used to diagnose disease states such as Huntington's chorea, cystic fibrosis, and hereditary Alzheimer's, and may eventually help persons with genetic disorders to overcome unfavorable outcomes.

even one amino acid in a protein can severely affect its conformation and function. This is the case with the p53 gene.

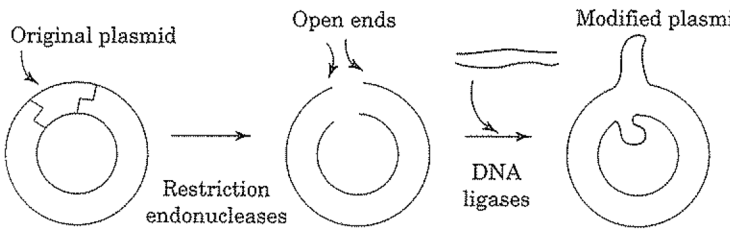
17.9 Recombinant DNA and Biotechnology

recombinant DNA
DNA that has been spliced into a foreign host

With a library of gene sequences as well as the technical expertise to analyze, modify, and produce synthetic DNA and RNA, it is now possible and practical to introduce genes not only into simple bacteria but also into animals and eventually into humans.

At its simplest level, natural, semisynthetic, or totally synthetic genes can be introduced into the genome of lower organisms such as the *E. coli* bacteria that inhabit the human gut. The circular plasmid DNA found in many types of bacteria is an easy vehicle to use. It can be removed from its cell, modified, and then returned in order to generate the proteins encoded. In order to introduce a foreign gene, **endonucleases,** which cleave DNA at specific base sequences, are used to open up a plasmid DNA molecule. A synthetic gene can be made with ends that can be annealed or attached by DNA ligases to the opening in the host DNA.

endonuclease
enzyme that cleaves polynucleotides within the chain



Interpreting DNA

It is in this way that a number of proteins have been mass produced by using the genes from higher animals, including humans. Some of them include human and bovine growth hormones, human insulin, tissue plasminogen activator (dis-

solves blood clots), and components of th
then been spliced into mice to produce t

Final observation: The scope of this ... is full of promise and, for some, dread. The challenge to society will be to support the knowledge while constraining its potential for abuse. Maybe the challenge is not one to science but rather one to our total humanity.

Problems

17.10 **Structure:** Give three important structural differences between DNA and RNA.

17.11 **Structure:** Uracil hydrogen-bonds to adenine in RNA in place of the thymine found in DNA. Draw the structure of the adenine–uracil hydrogen-bonding pairs.

17.12 **Genetic Code:** Given the following "sense," (+), or coding DNA sequence, write the sequences for the corresponding "antisense" DNA, mRNA, and tRNA. Be sure to indicate the 3′ and 5′ ends of the polynucleotides.

5'GTAACGTCGCTT3'

17.13 **Structure:** If one mole of the polynucleotide in problem 17.12 were completely hydrolyzed, what would be the products and how many moles of each would be produced?

17.14 **Structure:** There are ten nucleotide bases per 360° turn of the DNA molecule. This corresponds to a linear distance of about 34 Å (1 Å = 1 angstrom = 10^{-8} centimeters). How long, in meters, would a DNA molecule be if it contained one million nucleotide base pairs?

17.15 **Energy-Related Nucleotides:** What are the hydrolysis products of one mole of each of ATP, FAD, NADH, and FMN?

17.16 **Genetic Code:** Polypeptides, which have physiological activity such as the hormone glucagon, are derived from much larger protein precursors. What is the minimum number of nucleotide base-pairs that would be needed for the exon coding for glucagon, which has 37 amino acids?

17.17 **Genetic Code:** Two naturally occurring variations discovered in the amino acid sequence of adult hemoglobin involve substituting a lysine for a glutamic acid in the β chain (hemoglobin E) and substituting a tyrosine for a histidine also in the β chain (hemoglobin M_{Boston}). Considering the codons for these amino acids, can you offer some explanation for these natural substitutions?