28 Biomolecules: Nucleic Acids

Organic KNOWLEDGE TOOLS

ThomsonNOW™ Throughout this chapter, sign in at **www.thomsonedu.com** for online self-study and interactive tutorials based on your level of understanding.

Online homework for this chapter may be assigned in Organic OWL.

The nucleic acids, **deoxyribonucleic acid** (DNA) and **ribonucleic acid** (RNA), are the chemical carriers of a cell's genetic information. Coded in a cell's DNA is the information that determines the nature of the cell, controls the cell's growth and division, and directs biosynthesis of the enzymes and other proteins required for cellular functions.

In addition to the nucleic acids themselves, nucleic acid derivatives such as ATP are involved as phosphorylating agents in many biochemical pathways, and several important coenzymes, including NAD^+, FAD, and coenzyme A, have nucleic acid components.

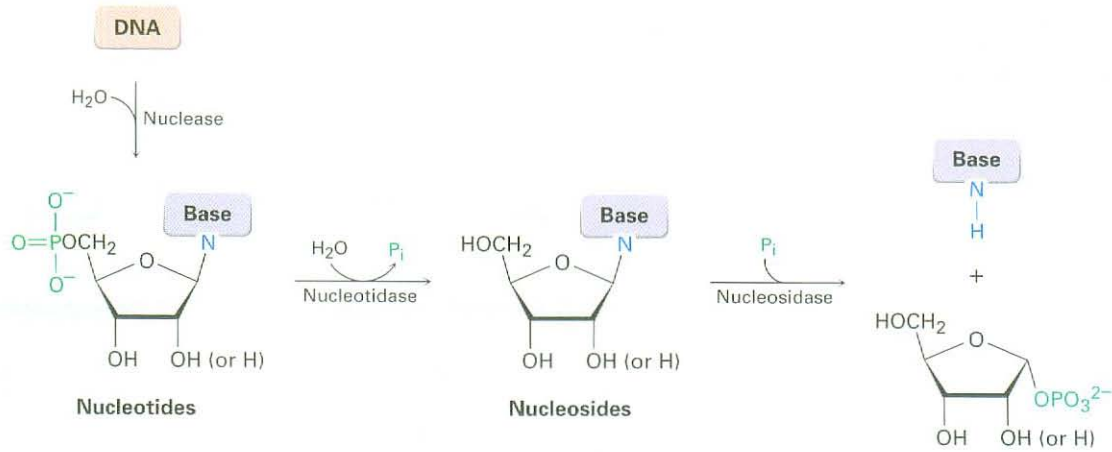
WHY THIS CHAPTER?

Nucleic acids are the last of the four major classes of biomolecules we'll consider. So much has been written and spoken about DNA in the media that the basics of DNA replication and transcription are probably known to you. Thus, we'll move fairly quickly through the fundamentals and then focus more closely on the chemical details of DNA sequencing and synthesis.

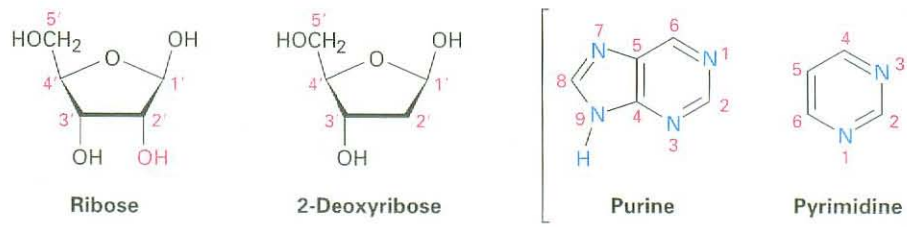
28.1 Nucleotides and Nucleic Acids

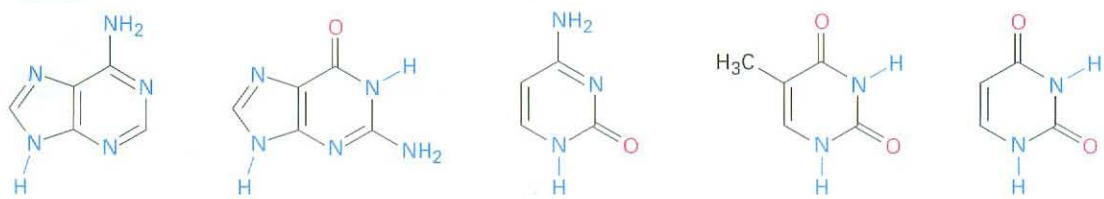
Just as proteins are biopolymers made of amino acids, nucleic acids are biopolymers made of **nucleotides** joined together to form a long chain. Each nucleotide is composed of a **nucleoside** bonded to a phosphate group, and each nucleoside is composed of an aldopentose sugar linked through its anomeric carbon to the nitrogen atom of a heterocyclic purine or pyrimidine base.

Sean Duggan



The sugar component in RNA is ribose, and the sugar in DNA is 2′-deoxyribose. (The prefix *2′-deoxy* indicates that oxygen is missing from the 2′ position of ribose.) DNA contains four different amine bases, two substituted purines (adenine and guanine) and two substituted pyrimidines (cytosine and thymine). Adenine, guanine, and cytosine also occur in RNA, but thymine is replaced in RNA by a closely related pyrimidine base called uracil.





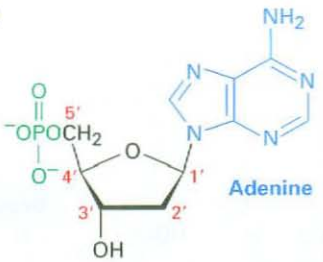
The structures of the four deoxyribonucleotides and the four ribonucleotides are shown in Figure 28.1. Note that in naming and numbering nucleotides, positions on the sugars are given a prime superscript to distinguish them from positions on the amine base. Position 3 would be on the base, for instance, while position 3′ would be on the sugar. Although similar chemically, DNA and RNA differ dramatically in size. Molecules of DNA are enormous, with molecular weights up to several billion. Molecules of RNA, by contrast, are much smaller, containing as few as 60 nucleotides and having molecular weights as low as 22,000.

ThomsonNOW Click *Organic Interactive* to **learn to recognize classes of nucleic acids and their base-pair partners.**

Figure 28.1 Structures of the four deoxyribonucleotides and the four ribonucleotides.

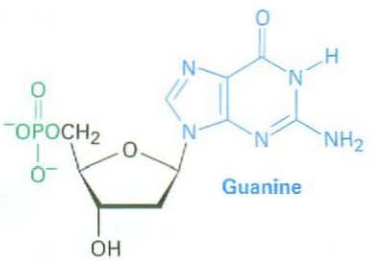
Deoxyribonucleotides

A



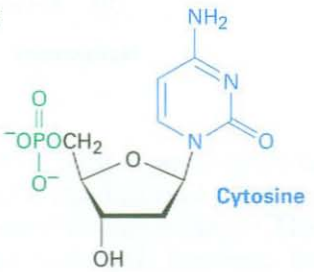
2′-Deoxyadenosine 5′-phosphate

G



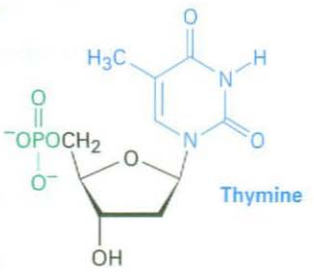
2′-Deoxyguanosine 5′-phosphate

C



2′-Deoxycytidine 5′-phosphate

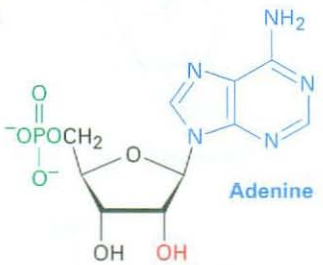
T



Thymidine 5′-phosphate

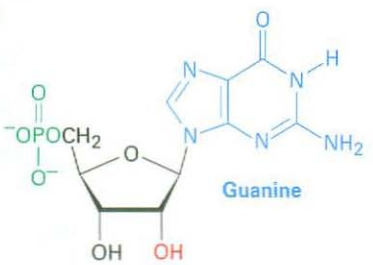
Ribonucleotides

A



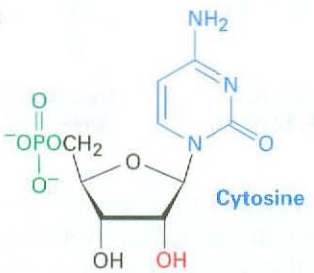
Adenosine 5′-phosphate

G



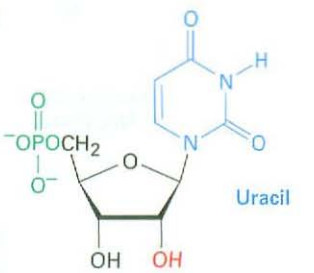
Guanosine 5′-phosphate

C



Cytidine 5′-phosphate

U

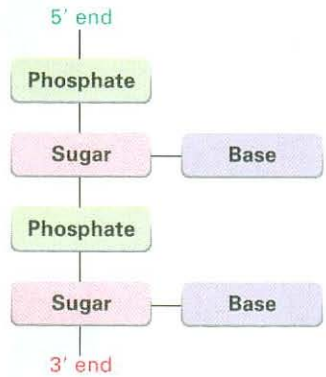


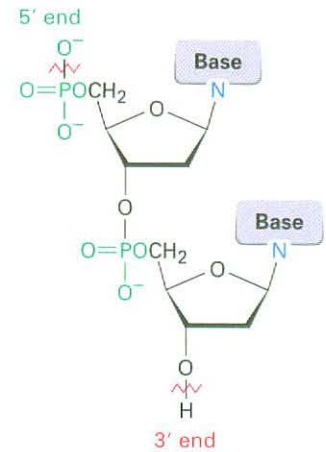
Uridine 5′-phosphate

Nucleotides are linked together in DNA and RNA by phosphodiester bonds [RO—(PO_2^-)—OR'] between phosphate, the 5′ hydroxyl group on one nucleoside, and the 3′-hydroxyl group on another nucleoside. One end of the nucleic acid polymer has a free hydroxyl at C3′ (the **3′ end**), and the other end has a phosphate at C5′ (the **5′ end**). The sequence of nucleotides in a chain is described by starting at the 5′ end and identifying the bases in order of occurrence, using the abbreviations G, C, A, T (or U for RNA). Thus, a typical DNA sequence might be written as TAGGCT.

James Dewey Watson

James Dewey Watson (1928–) was born in Chicago, Illinois, and enrolled in the University of Chicago at age 15. He received his Ph.D. in 1950 at the University of Indiana and then worked at Cambridge University in England from 1951 to 1953, where he and Francis Crick deduced the structure of DNA. After more than 20 years as professor at Harvard University, he moved in 1976 to the Laboratory of Quantitative Biology at Cold Spring Harbor, Long Island, New York. He shared the 1962 Nobel Prize in medicine for his work on nucleic acids.





Problem 28.1 Draw the full structure of the DNA dinucleotide AG.

Problem 28.2 Draw the full structure of the RNA dinucleotide UA.

28.2 Base Pairing in DNA: The Watson–Crick Model

Francis Harry Compton Crick

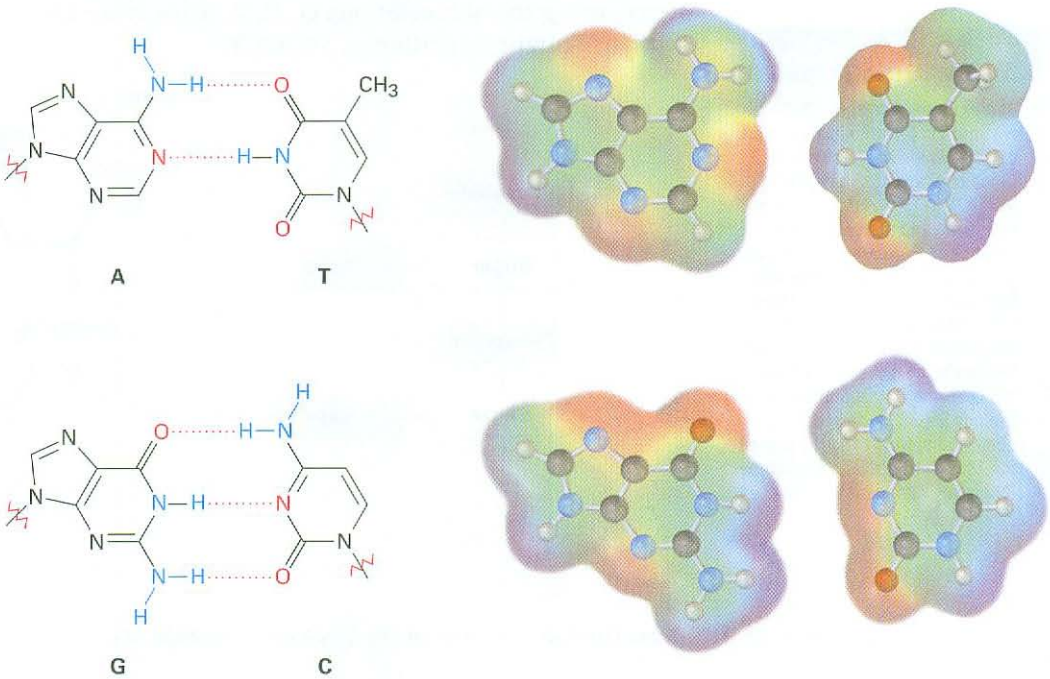
Francis Harry Compton Crick (1916–2004) was born in Northampton, England, and began his scientific career as a physicist. Following an interruption in his studies caused by World War II, he switched to biology and received his Ph.D. in 1954 at Cambridge University. He then remained at Cambridge University as professor. He shared the 1962 Nobel Prize in medicine.

Samples of DNA isolated from different tissues of the same species have the same proportions of heterocyclic bases, but samples from different species often have greatly different proportions of bases. Human DNA, for example, contains about 30% each of adenine and thymine and about 20% each of guanine and cytosine. The bacterium *Clostridium perfringens*, however, contains about 37% each of adenine and thymine and only 13% each of guanine and cytosine. Note that in both examples the bases occur in pairs. Adenine and thymine are present in equal amounts, as are cytosine and guanine. Why?

In 1953, James Watson and Francis Crick made their classic proposal for the secondary structure of DNA. According to the Watson–Crick model, DNA under physiological conditions consists of two polynucleotide strands, running in opposite directions and coiled around each other in a **double helix** like the handrails on a spiral staircase. The two strands are complementary rather than identical and are held together by hydrogen bonds between specific pairs of

ThomsonNOW Click *Organic Interactive* to **use interactive animations to view aspects of DNA structure.**

bases, A with T and C with G. That is, whenever an A base occurs in one strand, a T base occurs opposite it in the other strand; when a C base occurs in one, a G occurs in the other (Figure 28.2). This complementary base-pairing thus explains why A and T are always found in equal amounts, as are G and C.

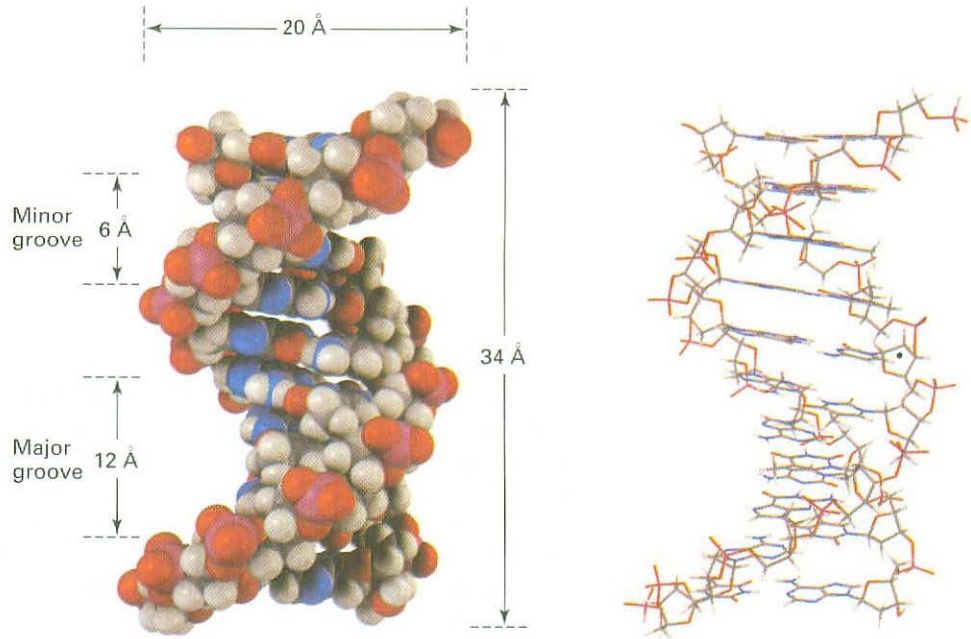


Active Figure 28.2 Hydrogen-bonding between base pairs in the DNA double helix. Electrostatic potential maps show that the faces of the bases are relatively neutral (green), while the edges have positive (blue) and negative (red) regions. Pairing G with C and A with T brings together oppositely charged regions. *Sign in at* **www.thomsonedu.com** *to see a simulation based on this figure and to take a short quiz.*

A full turn of the DNA double helix is shown in Figure 28.3. The helix is 20 Å wide, there are 10 base pairs per turn, and each turn is 34 Å in length. Notice in Figure 28.3 that the two strands of the double helix coil in such a way that two kinds of "grooves" result, a *major groove* 12 Å wide and a *minor groove* 6 Å wide. The major groove is slightly deeper than the minor groove, and both are lined by hydrogen bond donors and acceptors. As a result, a variety of flat, polycyclic aromatic molecules are able to slip sideways, or *intercalate,* between the stacked bases. Many cancer-causing and cancer-preventing agents function by interacting with DNA in this way.

An organism's genetic information is stored as a sequence of deoxyribonucleotides strung together in the DNA chain. For the information to be preserved and passed on to future generations, a mechanism must exist for copying DNA. For the information to be used, a mechanism must exist for decoding the DNA message and implementing the instructions it contains.

What Crick called the "central dogma of molecular genetics" says that the function of DNA is to store information and pass it on to RNA. The function of



Active Figure 28.3 A turn of the DNA double helix in both space-filling and wire-frame formats. The sugar–phosphate backbone runs along the outside of the helix, and the amine bases hydrogen bond to one another on the inside. Both major and minor grooves are visible. *Sign in at* **www.thomsonedu.com** *to see a simulation based on this figure and to take a short quiz.*

RNA, in turn, is to read, decode, and use the information received from DNA to make proteins. Thus, three fundamental processes take place.

- **Replication**—the process by which identical copies of DNA are made so that information can be preserved and handed down to offspring
- **Transcription**—the process by which the genetic messages are read and carried out of the cell nucleus to ribosomes, where protein synthesis occurs
- **Translation**—the process by which the genetic messages are decoded and used to synthesize proteins

WORKED EXAMPLE 28.1 ***Predicting the Complementary Base Sequence in Double-Stranded DNA***

What sequence of bases on one strand of DNA is complementary to the sequence TATGCAT on another strand?

Strategy Remember that A and G form complementary pairs with T and C, respectively, and then go through the sequence replacing A by T, G by C, T by A, and C by G. Remember also that the 5′ end is on the left and the 3′ end is on the right in the original strand.

Solution

Original: (5′) TATGCAT (3′)
Complement: (3′) ATACGTA (5′) or (5′) ATGCATA (3′)

Problem 28.3 What sequence of bases on one strand of DNA is complementary to the following sequence on another strand?

(5′) GGCTAATCCGT (3′)

28.3 Replication of DNA

DNA **replication** is an enzyme-catalyzed process that begins with a partial untwisting of the double helix and breaking of the hydrogen bonds between strands, brought about by enzymes called *helicases*. As the strands separate and bases are exposed, new nucleotides line up on each strand in a complementary manner, A to T and G to C, and two new strands begin to grow. Each new strand is complementary to its old template strand, and two identical DNA double helices are produced (Figure 28.4). Because each of the new DNA molecules contains one old strand and one new strand, the process is described as *semiconservative replication*.

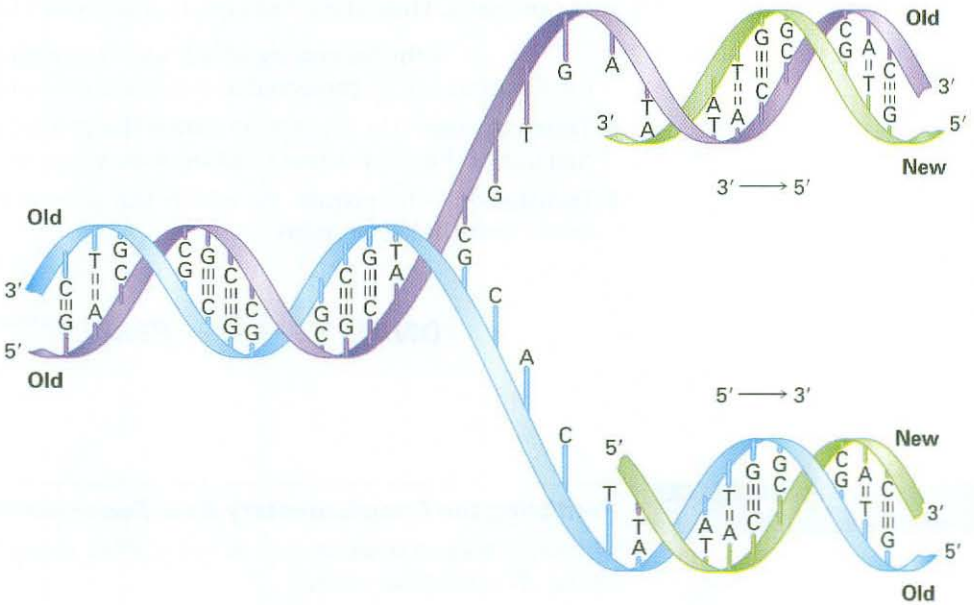
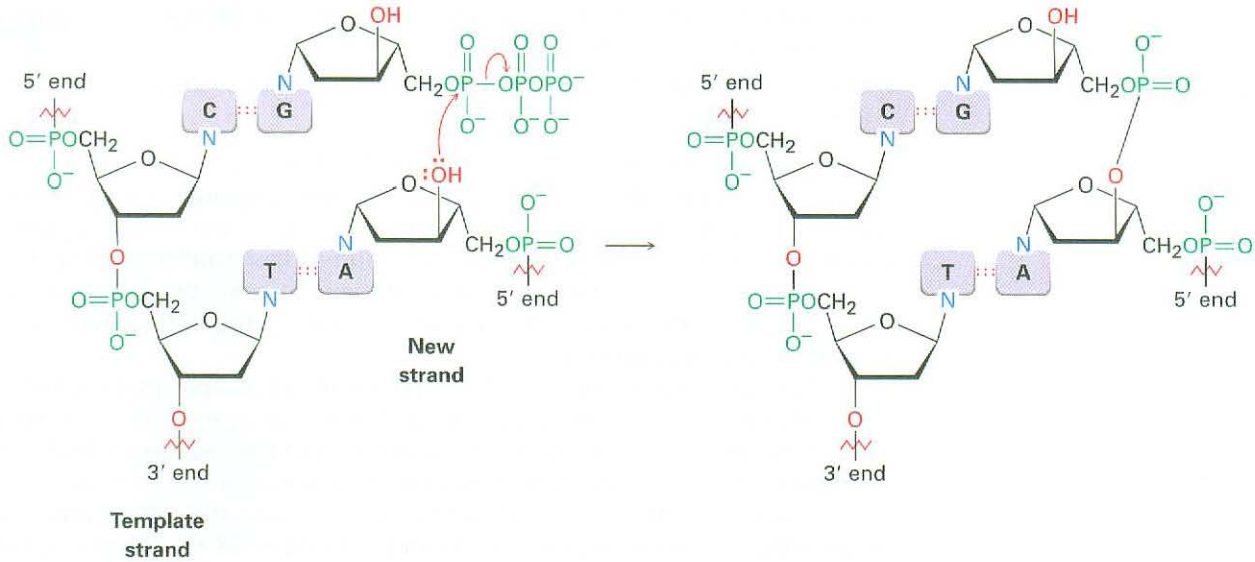


Figure 28.4 A representation of semiconservative DNA replication. The original double-stranded DNA partially unwinds, bases are exposed, nucleotides line up on each strand in a complementary manner, and two new strands begin to grow. Both strands are synthesized in the same 5′ → 3′ direction, one continuously and one in fragments.

Addition of nucleotides to the growing chain takes place in the 5′ → 3′direction and is catalyzed by DNA polymerase. The key step is the addition of a nucleoside 5′-triphosphate to the free 3′-hydroxyl group of the growing chain with loss of a diphosphate leaving group.



Because both new DNA strands are synthesized in the 5′ → 3′ direction, they can't be made in exactly the same way. One new strand must have its 3′ end nearer a point of unraveling (the *replication fork*), while the other new strand has its 5′ end nearer the replication fork. What happens is that the complement of the original 5′ → 3′ strand is synthesized continuously in a single piece to give a newly synthesized copy called the *leading strand*, while the complement of the original 3′ → 5′ strand is synthesized discontinuously in small pieces called *Okazaki fragments* that are subsequently linked by DNA ligases to form the *lagging strand*.

The magnitude of the replication process is staggering. The nucleus of every human cell contains 46 chromosomes (23 pairs), each of which consists of one very large DNA molecule. Each chromosome, in turn, is made up of hundreds of DNA segments called *genes*, and the sum of all genes in a human cell (the human *genome*) is estimated to be 2.9 billion base pairs. Despite the size of these enormous molecules, their base sequence is faithfully copied during replication. The copying process takes only minutes, and an error occurs only about once each 10 to 100 billion bases.

28.4 Transcription of DNA

As noted previously, RNA is structurally similar to DNA but contains ribose rather than deoxyribose and uracil rather than thymine. There are three major kinds of RNA, each of which serves a specific function. All three are much smaller molecules than DNA, and all remain single-stranded rather than double-stranded.

- **Messenger RNA (mRNA)** carries genetic messages from DNA to ribosomes, small granular particles in the cytoplasm of a cell where protein synthesis takes place.

- **Ribosomal RNA (rRNA)** complexed with protein provides the physical makeup of the ribosomes.
- **Transfer RNA (tRNA)** transports amino acids to the ribosomes, where they are joined together to make proteins.

The conversion of the information in DNA into proteins begins in the nucleus of cells with the synthesis of mRNA by **transcription** of DNA. In bacteria, the process begins when RNA polymerase recognizes and binds to a *promoter sequence* on DNA, typically consisting of around 40 base pairs located upstream (5′) of the transcription start site. Within the promoter are two hexameric *consensus sequences*, one located 10 base pairs upstream of the start and the second located 35 base pairs upstream.

Following formation of the polymerase–promoter complex, several turns of the DNA double helix untwist, forming a "bubble" and exposing 14 or so base pairs of the two strands. Appropriate ribonucleotides then line up by hydrogen-bonding to their complementary bases on DNA, bond formation occurs in the 5′ → 3′ direction, the RNA polymerase moves along the DNA chain, and the growing RNA molecule unwinds from DNA (Figure 28.5). At any one time, about 12 base pairs of the growing RNA remain hydrogen-bonded to the DNA template.

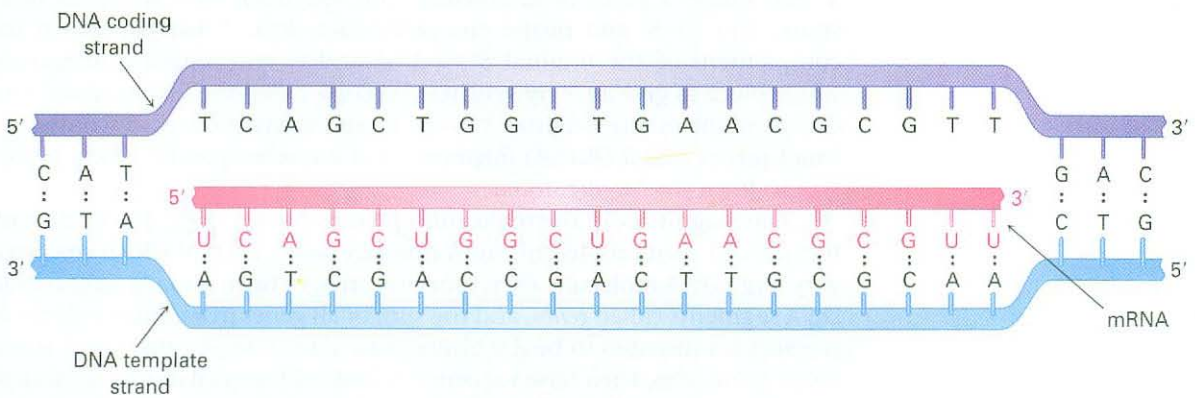


Figure 28.5 Biosynthesis of RNA using a DNA segment as a template.

Unlike what happens in DNA replication, where both strands are copied, only one of the two DNA strands is transcribed into mRNA. The strand that contains the gene is often called the **coding strand**, or *primer strand*, and the strand that gets transcribed is called the **template strand**. Because the template strand and the coding strand are complementary, and because the template strand and the transcribed RNA are also complementary, *the RNA molecule produced during transcription is a copy of the DNA coding strand*. The only difference is that the RNA molecule has a U everywhere the DNA coding strand has a T.

Another part of the picture in vertebrates and flowering plants is that genes are often not continuous segments of the DNA chain. Instead, a gene will begin in one small section of DNA called an *exon*, then be interrupted by a noncoding

section called an *intron*, and then take up again farther down the chain in another exon. The final mRNA molecule results only after the noncoded sections are cut out and the remaining pieces are joined together by spliceosomes. The gene for triose phosphate isomerase in maize, for instance, contains nine exons accounting for approximately 80% of the DNA base pairs and eight introns accounting for only 20% of the base pairs.

Problem 28.4 Show how uracil can form strong hydrogen bonds to adenine.

Problem 28.5 What RNA base sequence is complementary to the following DNA base sequence?

(5′) GATTACCGTA (3′)

Problem 28.6 From what DNA base sequence was the following RNA sequence transcribed?

(5′) UUCGCAGAGU (3′)

28.5 Translation of RNA: Protein Biosynthesis

The primary cellular function of mRNA is to direct biosynthesis of the thousands of diverse peptides and proteins required by an organism—perhaps 100,000 in a human. The mechanics of protein biosynthesis take place on ribosomes, small granular particles in the cytoplasm of a cell that consist of about 60% ribosomal RNA and 40% protein.

The specific ribonucleotide sequence in mRNA forms a message that determines the order in which amino acid residues are to be joined. Each "word," or **codon**, along the mRNA chain consists of a sequence of three ribonucleotides that is specific for a given amino acid. For example, the series UUC on mRNA is a codon directing incorporation of the amino acid phenylalanine into the growing protein. Of the $4^3 = 64$ possible triplets of the four bases in RNA, 61 code for specific amino acids and 3 code for chain termination. Table 28.1 shows the meaning of each codon.

The message embedded in mRNA is read by transfer RNA (tRNA) in a process called **translation**. There are 61 different tRNAs, one for each of the 61 codons that specifies an amino acid. A typical tRNA is single-stranded and has roughly the shape of a cloverleaf, as shown in Figure 28.6 on page 1111. It consists of about 70 to 100 ribonucleotides and is bonded to a specific amino acid by an ester linkage through the 3′ hydroxyl on ribose at the 3′ end of the tRNA. Each tRNA also contains on its middle leaf a segment called an **anticodon**, a sequence of three ribonucleotides complementary to the codon sequence. For example, the codon sequence UUC present on mRNA is read by a phenylalanine-bearing tRNA having the complementary anticodon base sequence GAA. [Remember that nucleotide sequences are written in the 5′ → 3′ direction, so the sequence in an anticodon must be reversed. That is, the complement to (5′)-UUC-(3′) is (3′)-AAG-(5′), which is written as (5′)-GAA-(3′).]

As each successive codon on mRNA is read, different tRNAs bring the correct amino acids into position for enzyme-mediated transfer to the growing

Table 28.1 Codon Assignments of Base Triplets

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

peptide. When synthesis of the proper protein is completed, a "stop" codon signals the end and the protein is released from the ribosome. The process is illustrated in Figure 28.7.

WORKED EXAMPLE 28.2 ***Predicting the Amino Acid Sequence Transcribed from DNA***

What amino acid sequence is coded by the following segment of a DNA coding strand?

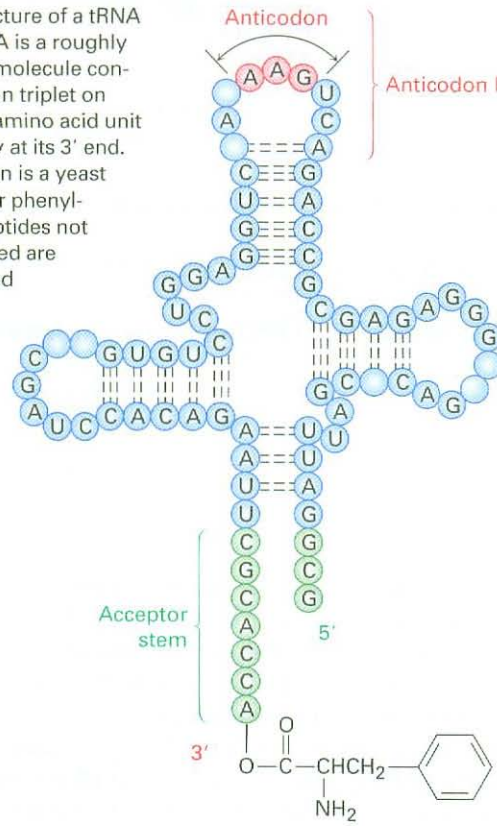
(5′) CTA-ACT-AGC-GGG-TCG-CCG (3′)

Strategy The mRNA produced during translation is a copy of the DNA coding strand, with each T replaced by U. Thus, the mRNA has the sequence

(5′) CUA-ACU-AGC-GGG-UCG-CCG (3′)

Each set of three bases forms a codon, whose meaning can be found in Table 28.1.

Solution Leu-Thr-Ser-Gly-Ser-Pro.



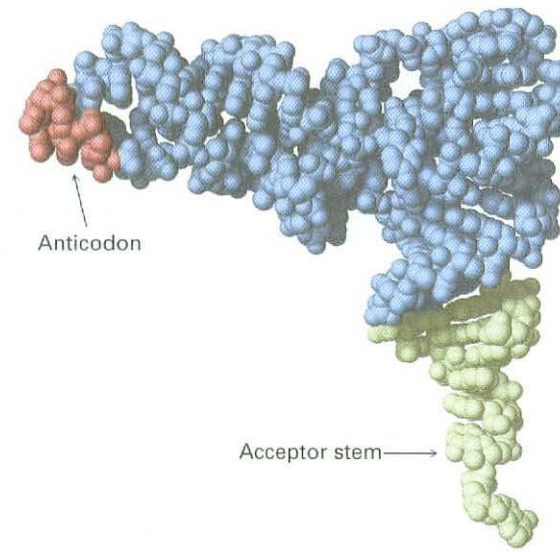


Figure 28.6 Structure of a tRNA molecule. The tRNA is a roughly cloverleaf-shaped molecule containing an anticodon triplet on one "leaf" and an amino acid unit attached covalently at its 3′ end. The example shown is a yeast tRNA that codes for phenylalanine. The nucleotides not specifically identified are chemically modified analogs of the four common nucleotides.

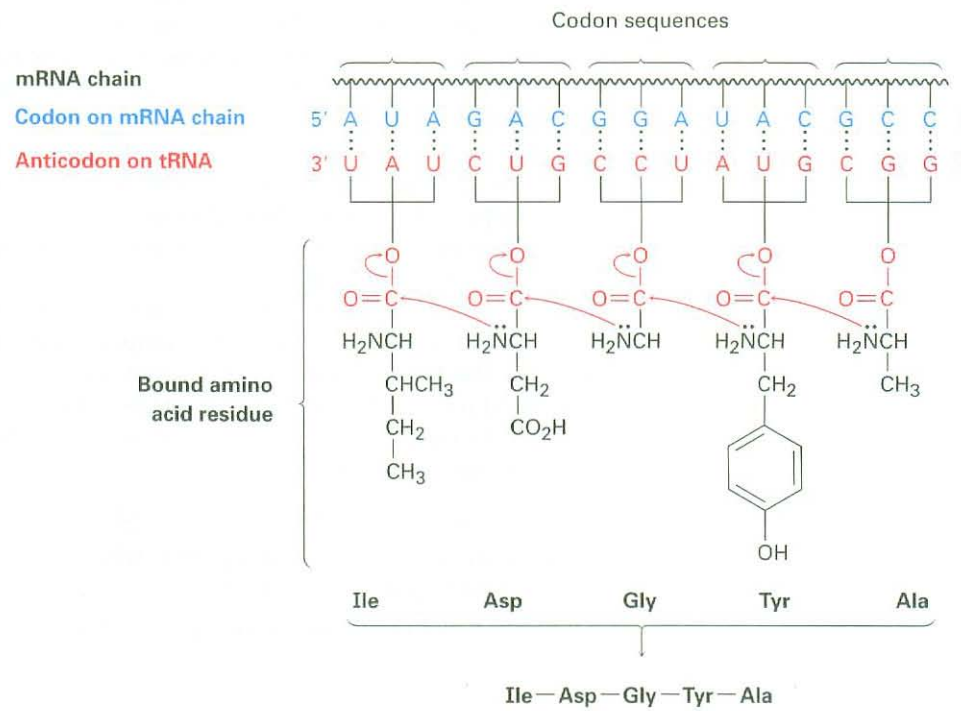


Figure 28.7 A representation of protein biosynthesis. The codon base sequences on mRNA are read by tRNAs containing complementary anticodon base sequences. Transfer RNAs assemble the proper amino acids into position for incorporation into the growing peptide.

Problem 28.7 List codon sequences for the following amino acids:
(a) Ala (b) Phe (c) Leu (d) Tyr

Problem 28.8 List anticodon sequences on the tRNAs carrying the amino acids shown in Problem 28.7.

Problem 28.9 What amino acid sequence is coded by the following mRNA base sequence?

CUU-AUG-GCU-UGG-CCC-UAA

Problem 28.10 What is the base sequence in the original DNA strand on which the mRNA sequence in Problem 28.9 was made?

28.6 DNA Sequencing

One of the greatest scientific revolutions in history is now under way in molecular biology, as scientists are learning how to manipulate and harness the genetic machinery of organisms. None of the extraordinary advances of the past two decades would have been possible, however, were it not for the discovery in 1977 of methods for sequencing immense DNA chains.

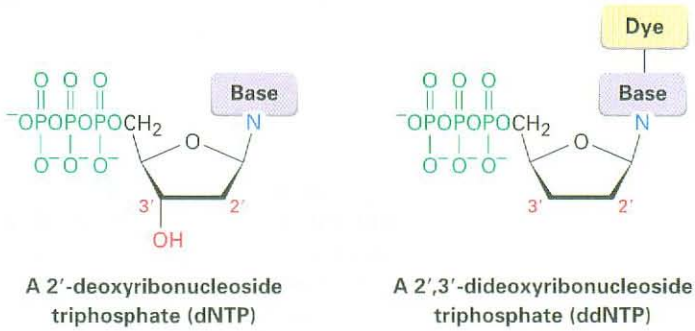
The first step in DNA sequencing is to cleave the enormous chain at known points to produce smaller, more manageable pieces, a task accomplished by the use of *restriction endonucleases*. Each different restriction enzyme, of which more than 3500 are known and approximately 200 are commercially available, cleaves a DNA molecule at a point in the chain where a specific base sequence occurs. For example, the restriction enzyme *Alu*I cleaves between G and C in the four-base sequence AG-CT. Note that the sequence is a *palindrome*, meaning that the *sequence* (5′)-AGCT-(3′) is the same as its *complement* (3′)-TCGA-(5′) when both are read in the same 5′ → 3′ direction. The same is true for other restriction endonucleases.

If the original DNA molecule is cut with another restriction enzyme having a different specificity for cleavage, still other segments are produced whose sequences partially overlap those produced by the first enzyme. Sequencing of all the segments, followed by identification of the overlapping regions, allows complete DNA sequencing.

Two methods of DNA sequencing are available. The *Maxam–Gilbert method* uses chemical techniques, while the **Sanger dideoxy method** uses enzymatic reactions. The Sanger method is the more commonly used of the two and was the method responsible for sequencing the entire human genome of 2.9 billion base pairs. In commercial sequencing instruments, the dideoxy method begins with a mixture of the following:

- The restriction fragment to be sequenced
- A small piece of DNA called a *primer*, whose sequence is complementary to that on the 3′ end of the restriction fragment
- The four 2′-deoxyribonucleoside triphosphates (dNTPs)

- Very small amounts of the four 2′,3′-dideoxyribonucleoside triphosphates (ddNTPs), each of which is labeled with a fluorescent dye of a different color (A 2′,3′-*dideoxy*ribonucleoside triphosphate is one in which both 2′ and 3′ −OH groups are missing from ribose.)



A 2′-deoxyribonucleoside triphosphate (dNTP)

A 2′,3′-dideoxyribonucleoside triphosphate (ddNTP)

DNA polymerase is added to the mixture, and a strand of DNA complementary to the restriction fragment begins to grow from the end of the primer. Most of the time, only normal deoxyribonucleotides are incorporated into the growing chain because of their much higher concentration in the mixture, but every so often, a dideoxyribonucleotide is incorporated. When that happens, DNA synthesis stops because the chain end no longer has a 3′-hydroxyl group for adding further nucleotides.

When reaction is complete, the product consists of a mixture of DNA fragments of all possible lengths, each terminated by one of the four dye-labeled dideoxyribonucleotides. This product mixture is then separated according to the size of the pieces by gel electrophoresis (Section 26.2), and the identity of the terminal dideoxyribonucleotide in each piece—and thus the sequence of the restriction fragment—is identified simply by noting the color with which the attached dye fluoresces. Figure 28.8 shows a typical result.

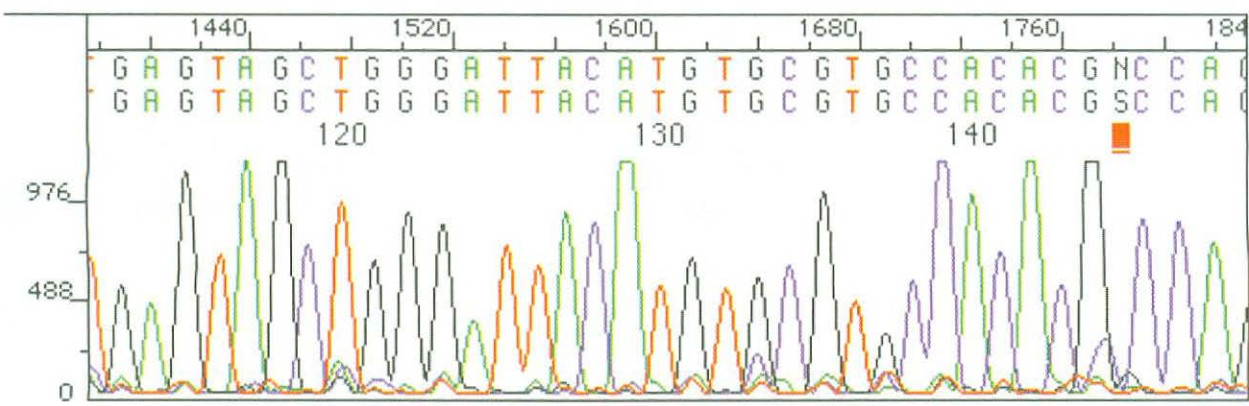


Figure 28.8 The sequence of a restriction fragment determined by the Sanger dideoxy method can be read simply by noting the colors of the dye attached to each of the various terminal nucleotides.

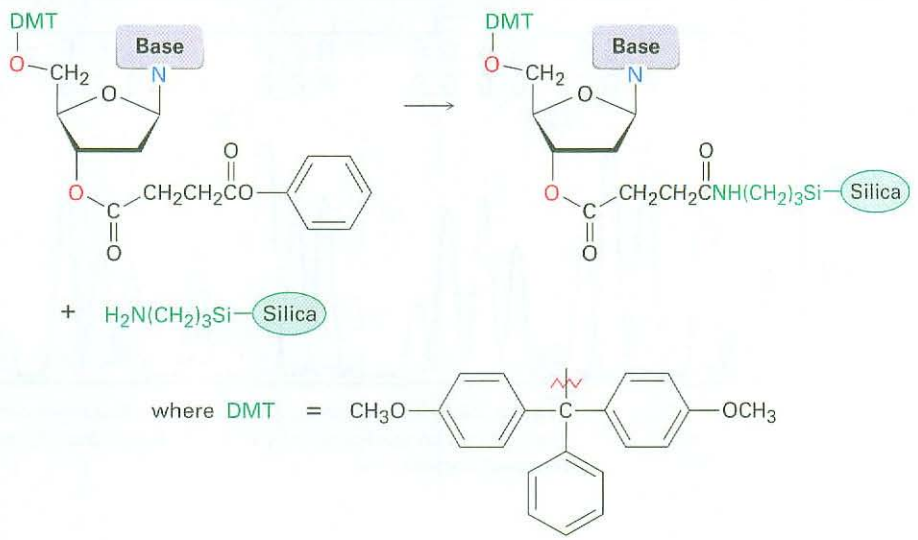
So efficient is the automated dideoxy method that sequences up to 1100 nucleotides in length, with a throughput of up to 19,000 bases per hour, can be sequenced with 98% accuracy. After a decade of work, preliminary sequence information for the entire human genome of 2.9 billion base pairs was announced early in 2001. Remarkably, our genome appears to contain only about 30,000 genes, less than one-third the previously predicted number and only twice the number found in the common roundworm.

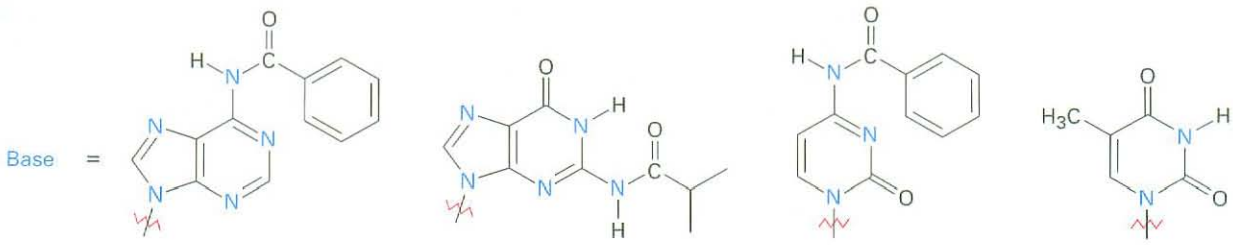
28.7 DNA Synthesis

The ongoing revolution in molecular biology has brought with it an increased demand for the efficient chemical synthesis of short DNA segments, called *oligonucleotides,* or simply *oligos*. The problems of DNA synthesis are similar to those of protein synthesis (Section 26.7) but are more difficult because of the complexity of the nucleotide monomers. Each nucleotide has multiple reactive sites that must be selectively protected and deprotected at the proper times, and coupling of the four nucleotides must be carried out in the proper sequence. Automated DNA synthesizers are available, however, that allow the fast and reliable synthesis of DNA segments up to 200 nucleotides in length.

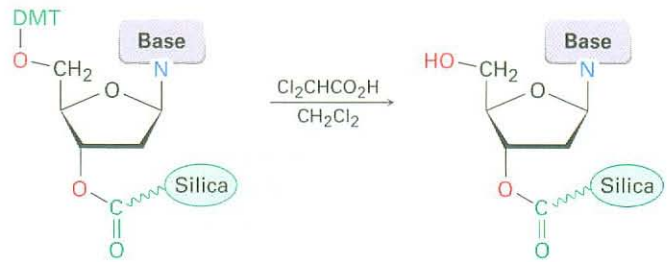
DNA synthesizers operate on a principle similar to that of the Merrifield solid-phase peptide synthesizer (Section 26.8). In essence, a protected nucleotide is covalently bonded to a solid support, and one nucleotide at a time is added to the growing chain by the use of a coupling reagent. After the final nucleotide has been added, all the protecting groups are removed and the synthetic DNA is cleaved from the solid support. Five steps are needed:

Step 1 The first step in DNA synthesis is to attach a protected deoxynucleoside to a silica (SiO_2) support by an ester linkage to the 3′ −OH group of the deoxynucleoside. Both the 5′ −OH group on the sugar and free $-NH_2$ groups on the heterocyclic bases must be protected. Adenine and cytosine bases are protected by benzoyl groups, guanine is protected by an isobutyryl group, and thymine requires no protection. The deoxyribose 5′ −OH is protected as its *p*-dimethoxytrityl (DMT) ether.

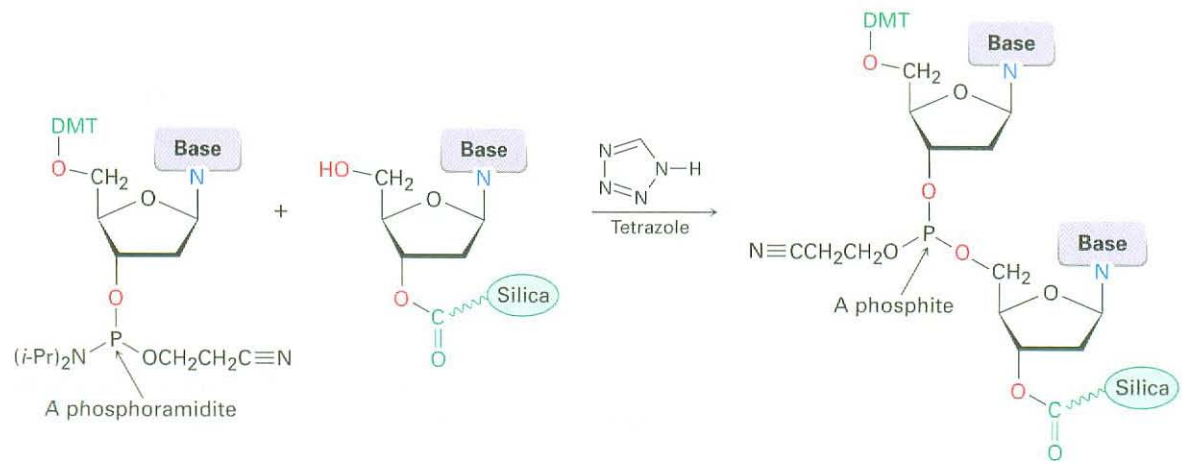




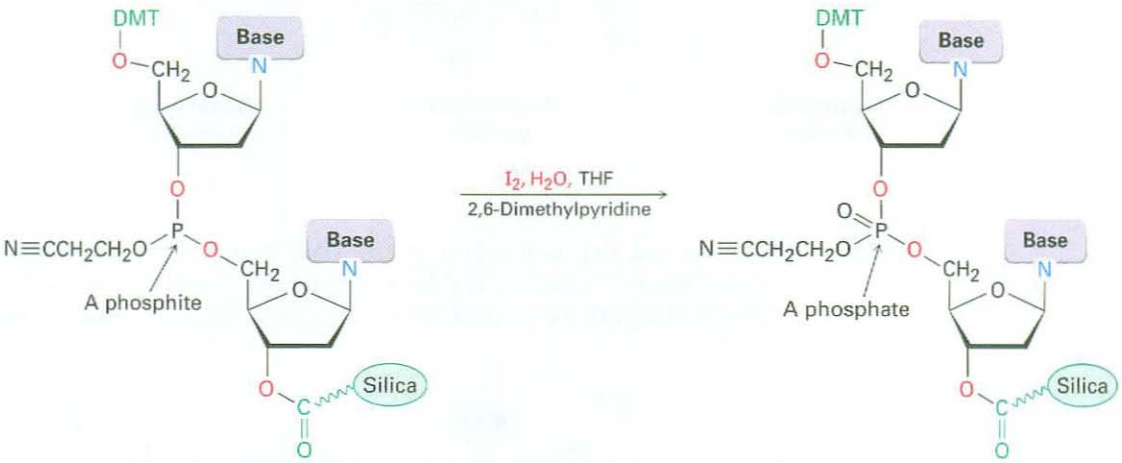
Step 2 The second step is removal of the DMT protecting group by treatment with dichloroacetic acid in CH_2Cl_2. The reaction occurs by an S_N1 mechanism and proceeds rapidly because of the stability of the tertiary, benzylic dimethoxytrityl cation.



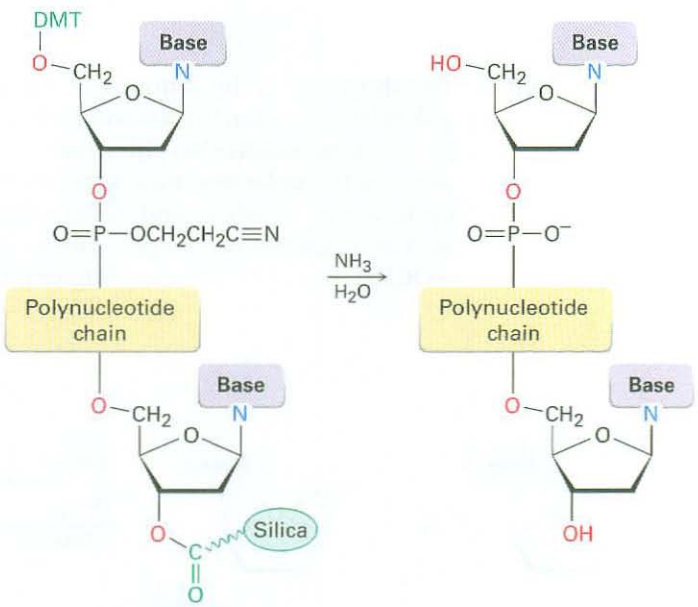
Step 3 The third step is the coupling of the polymer-bonded deoxynucleoside with a protected deoxynucleoside containing a *phosphoramidite* group at its 3′ position. [A phosphoramidite has the structure $R_2NP(OR)_2$.] The coupling reaction takes place in the polar aprotic solvent acetonitrile; requires catalysis by the heterocyclic amine tetrazole; and yields a *phosphite*, $P(OR)_3$, as product. Note that one of the phosphorus oxygen atoms is protected by a β-cyanoethyl group, $-OCH_2CH_2C{\equiv}N$. The coupling step takes place in better than 99% yield.



Step 4 With the coupling accomplished, the phosphite product is oxidized to a phosphate by treatment with iodine in aqueous tetrahydrofuran in the presence of 2,6-dimethylpyridine. The cycle (1) deprotection, (2) coupling, and (3) oxidation is then repeated until an oligonucleotide chain of the desired sequence has been built.



Step 5 The final step is removal of all protecting groups and cleavage of the ester bond holding the DNA to the silica. All these reactions are done at the same time by treatment with aqueous NH_3. Purification by electrophoresis then yields the synthetic DNA.



Problem 28.11 | *p*-Dimethoxytrityl (DMT) ethers are easily cleaved by mild acid treatment. Show the mechanism of the cleavage reaction.

Problem 28.12 Propose a mechanism to account for cleavage of the β-cyanoethyl protecting group from the phosphate groups on treatment with aqueous ammonia. (Acrylonitrile, $H_2C{=}CHCN$, is a by-product.) What kind of reaction is occurring?

28.8 The Polymerase Chain Reaction

Kary Banks Mullis

Kary Banks Mullis (1944–) was born in rural Lenoir, North Carolina; did undergraduate work at Georgia Tech.; and received his Ph.D. at the University of California, Berkeley, in 1973. From 1979 to 1986 he worked at Cetus Corp., where his work on developing PCR was carried out. Since 1988, he has followed his own drummer as self-employed consultant and writer. He received the 1993 Nobel Prize in chemistry.

It often happens that only tiny amounts of a gene sequence can be obtained directly from an organism's DNA, so methods for obtaining larger amounts are sometimes needed to carry out the sequencing. The invention of the **polymerase chain reaction (PCR)** by Kary Mullis in 1986 has been described as being to genes what Gutenberg's invention of the printing press was to the written word. Just as the printing press produces multiple copies of a book, PCR produces multiple copies of a given DNA sequence. Starting from less than 1 *picogram* of DNA with a chain length of 10,000 nucleotides (1 pg = 10^{-12} g; about 100,000 molecules), PCR makes it possible to obtain several micrograms (1 μg = 10^{-6} g; about 10^{11} molecules) in just a few hours.

The key to the polymerase chain reaction is *Taq* DNA polymerase, a heat-stable enzyme isolated from the thermophilic bacterium *Thermus aquaticus* found in a hot spring in Yellowstone National Park. *Taq* polymerase is able to take a single strand of DNA that has a short, primer segment of complementary chain at one end and then finish constructing the entire complementary strand. The overall process takes three steps, as shown schematically in Figure 28.9. (More recently, improved heat-stable DNA polymerase enzymes have become available, including Vent polymerase and *Pfu* polymerase, both isolated from bacteria growing near geothermal vents in the ocean floor. The error rate of both enzymes is substantially less than that of *Taq*.)

Step 1 The double-stranded DNA to be amplified is heated in the presence of *Taq* polymerase, Mg^{2+} ion, the four deoxynucleotide triphosphate monomers (dNTPs), and a large excess of two short oligonucleotide primers of about 20 bases each. Each primer is complementary to the sequence at the end of one of the target DNA segments. At a temperature of 95 °C, double-stranded DNA denatures, spontaneously breaking apart into two single strands.

Step 2 The temperature is lowered to between 37 and 50 °C, allowing the primers, because of their relatively high concentration, to anneal by hydrogen-bonding to their complementary sequence at the end of each target strand.

Step 3 The temperature is then raised to 72 °C, and *Taq* polymerase catalyzes the addition of further nucleotides to the two primed DNA strands. When replication of each strand is finished, *two* copies of the original DNA now exist. Repeating the denature–anneal–synthesize cycle a second time yields four DNA copies, repeating a third time yields eight copies, and so on, in an exponential series.

PCR has been automated, and 30 or so cycles can be carried out in an hour, resulting in a theoretical amplification factor of 2^{30} ($\sim 10^9$). In practice, however, the efficiency of each cycle is less than 100%, and an experimental amplification of about 10^6 to 10^8 is routinely achieved for 30 cycles.

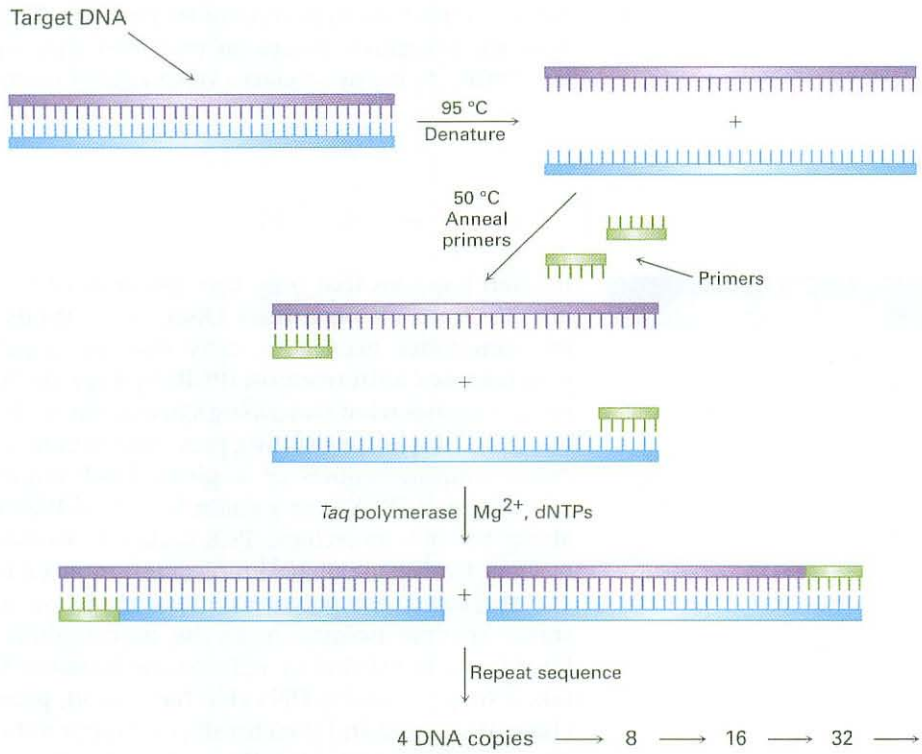


Figure 28.9 The polymerase chain reaction. Details are explained in the text.

Focus On . . .

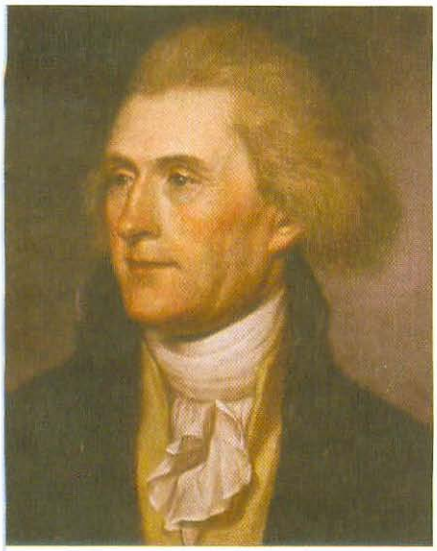
DNA Fingerprinting

The invention of DNA sequencing has affected society in many ways, few more dramatic than those stemming from the development of *DNA fingerprinting*. DNA fingerprinting arose from the discovery in 1984 that human genes contain short, repeating sequences of noncoding DNA, called *short tandem repeat* (STR) loci. Furthermore, the STR loci are slightly different for every individual, except identical twins. By sequencing these loci, a pattern unique to each person can be obtained.

Perhaps the most common and well-publicized use of DNA fingerprinting is that carried out by crime laboratories to link suspects to biological evidence—blood, hair follicles, skin, or semen—found at a crime scene. Thousands of court cases have now been decided based on DNA evidence.

For use in criminal cases, forensic laboratories in the United States have agreed on 13 core STR loci that are most accurate for identification of an individual. Based on these 13 loci, a Combined DNA Index System (CODIS) has

(continued)



Historians have wondered for many years whether Thomas Jefferson fathered a child by Sally Hemings. DNA fingerprinting evidence obtained in 1998 is inconclusive but strongly suggestive.

been established to serve as a registry of convicted offenders. When a DNA sample is obtained from a crime scene, the sample is subjected to cleavage with restriction endonucleases to cut out fragments containing the STR loci, the fragments are amplified using the polymerase chain reaction, and the sequences of the fragments are determined.

If the profile of sequences from a known individual and the profile from DNA obtained at a crime scene match, the probability is approximately 82 billion to 1 that the DNA is from the same individual. In paternity cases, where the DNA of father and offspring are related but not fully identical, the identity of the father can be established with a probability of 100,000 to 1. Even after several generations have passed, paternity can still be implied by DNA analysis of the Y chromosome of direct male-line descendants. The most well-known such case is that of Thomas Jefferson, who may have fathered a child by his slave Sally Hemings. Although Jefferson himself has no male-line descendants, DNA analysis of the male-line descendants of Jefferson's paternal uncle contained the same Y chromosome as a male-line descendant of Eston Hemings, youngest son of Sally Hemings. Thus, a mixing of the two genomes is clear, although the male individual responsible for that mixing can't be conclusively identified.

Among its many other applications, DNA fingerprinting is widely used for the diagnosis of genetic disorders, both prenatally and in newborns. Cystic fibrosis, hemophilia, Huntington's disease, Tay–Sachs disease, sickle cell anemia, and thalassemia are among the many diseases that can be detected, enabling early treatment of an affected child. Furthermore, by studying the DNA fingerprints of relatives with a history of a particular disorder, it's possible to identify DNA patterns associated with the disease and perhaps obtain clues for eventual cure. In addition, the U.S. Department of Defense now requires blood and saliva samples from all military personnel. The samples are stored, and DNA is extracted should the need for identification of a casualty arise.

SUMMARY AND KEY WORDS

The nucleic acids **DNA (deoxyribonucleic acid)** and **RNA (ribonucleic acid)** are biological polymers that act as chemical carriers of an organism's genetic information. Enzyme-catalyzed hydrolysis of nucleic acids yields **nucleotides**, the monomer units from which RNA and DNA are constructed. Further enzyme-catalyzed hydrolysis of the nucleotides yields **nucleosides** plus phosphate. Nucleosides, in turn, consist of a purine or pyrimidine base linked to C1 of an aldopentose sugar—ribose in RNA and 2-deoxyribose in DNA. The nucleotides are joined by phosphate links between the 5′ phosphate of one nucleotide and the 3′ hydroxyl on the sugar of another nucleotide.

Molecules of DNA consist of two complementary polynucleotide strands held together by hydrogen bonds between heterocyclic bases on the different strands and coiled into a **double helix**. Adenine and thymine form hydrogen bonds to each other, as do cytosine and guanine.

polymerase chain reaction (PCR), 1117
replication, 1106
ribonucleic acid (RNA), 1100
ribosomal RNA (rRNA), 1108
Sanger dideoxy method, 1112
template strand, 1108
transcription, 1108
transfer RNA (tRNA), 1108
translation, 1109

Three processes take place in deciphering the genetic information of DNA:

- **Replication** of DNA is the process by which identical DNA copies are made. The DNA double helix unwinds, complementary deoxyribonucleotides line up in order, and two new DNA molecules are produced.
- **Transcription** is the process by which RNA is produced to carry genetic information from the nucleus to the ribosomes. A short segment of the DNA double helix unwinds, and complementary ribonucleotides line up to produce **messenger RNA (mRNA)**.
- **Translation** is the process by which mRNA directs protein synthesis. Each mRNA is divided into **codons**, ribonucleotide triplets that are recognized by small amino acid–carrying molecules of **transfer RNA (tRNA)**, which deliver the appropriate amino acids needed for protein synthesis.

Sequencing of DNA is carried out by the **Sanger dideoxy method**, and small DNA segments can be synthesized in the laboratory by automated instruments. Small amounts of DNA can be amplified by factors of 10^6 using the **polymerase chain reaction (PCR)**.

EXERCISES

Organic KNOWLEDGE TOOLS

ThomsonNOW™ Sign in at **www.thomsonedu.com** to assess your knowledge of this chapter's topics by taking a pre-test. The pre-test will link you to interactive organic chemistry resources based on your score in each concept area.

Online homework for this chapter may be assigned in Organic OWL.

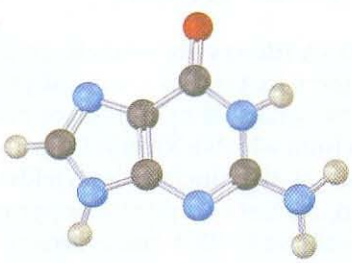
■ indicates problems assignable in Organic OWL.

VISUALIZING CHEMISTRY

(Problems 28.1–28.12 appear within the chapter.)

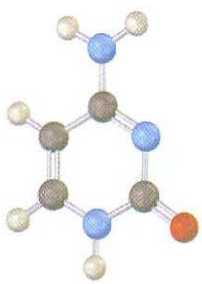
28.13 ■ Identify the following bases, and tell whether each is found in DNA, RNA, or both:

(a)

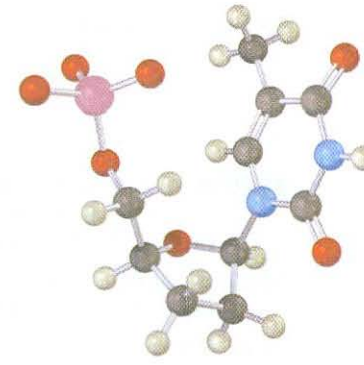


(b)

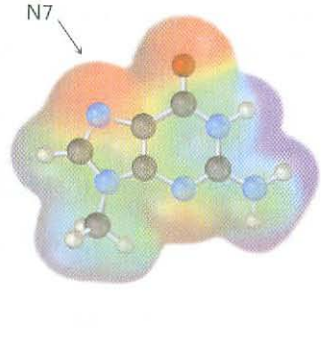
(c)



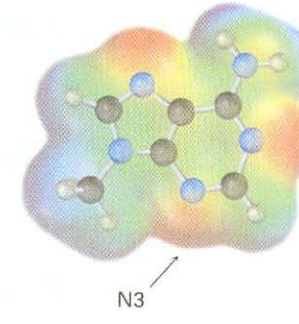
28.14 ■ Identify the following nucleotide, and tell how it is used:



28.15 Amine bases in nucleic acids can react with alkylating agents in typical S_N2 reactions. Look at the following electrostatic potential maps, and tell which is the better nucleophile, guanine or adenine. The reactive positions in each are indicated.



9-Methylguanine



9-Methyladenine

ADDITIONAL PROBLEMS

28.16 Human brain natriuretic peptide (BNP) is a small peptide of 32 amino acids used in the treatment of congestive heat failure. How many nitrogen bases are present in the DNA that codes for BNP?

28.17 Human and horse insulin both have two polypeptide chains, with one chain containing 21 amino acids and the other containing 30 amino acids. They differ in primary structure at two places. At position 9 in one chain, human insulin has Ser and horse insulin has Gly; at position 30 in the other chain, human insulin has Thr and horse insulin has Ala. How must the DNA for the two insulins differ?

28.18 ■ The DNA of sea urchins contains about 32% A. What percentages of the other three bases would you expect in sea urchin DNA? Explain.

28.19 The codon UAA stops protein synthesis. Why does the sequence UAA in the following stretch of mRNA not cause any problems?

-GCA-UUC-GAG-GUA-ACG-CCC-

28.20 ■ Which of the following base sequences would most likely be recognized by a restriction endonuclease? Explain.
(a) GAATTC (b) GATTACA (c) CTCGAG

28.21 ■ For what amino acids do the following ribonucleotide triplets code?
(a) AAU (b) GAG (c) UCC (d) CAU

28.22 ■ From what DNA sequences were each of the mRNA codons in Problem 28.21 transcribed?

28.23 ■ What anticodon sequences of tRNAs are coded for by the codons in Problem 28.21?

28.24 Draw the complete structure of the ribonucleotide codon UAC. For what amino acid does this sequence code?

28.25 Draw the complete structure of the deoxyribonucleotide sequence from which the mRNA codon in Problem 28.24 was transcribed.

28.26 Give an mRNA sequence that will code for synthesis of metenkephalin.

Tyr-Gly-Gly-Phe-Met

28.27 Give an mRNA sequence that will code for the synthesis of angiotensin II.

Asp-Arg-Val-Tyr-Ile-His-Pro-Phe

28.28 ■ What amino acid sequence is coded for by the following DNA coding strand?

(5′) CTT-CGA-CCA-GAC-AGC-TTT (3′)

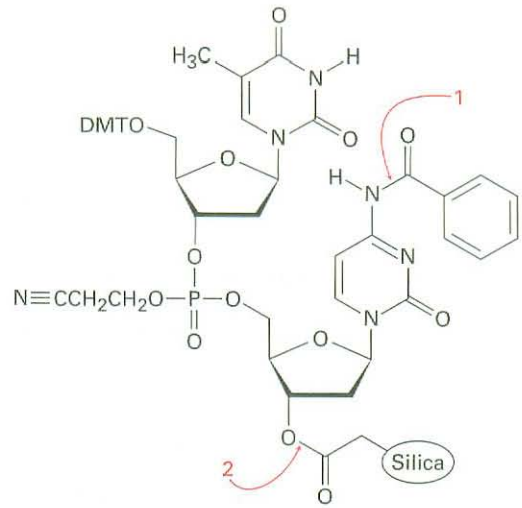
28.29 ■ What amino acid sequence is coded for by the following mRNA base sequence?

(5′) CUA-GAC-CGU-UCC-AAG-UGA (3′)

28.30 If the DNA coding sequence -CAA-CCG-GAT- were miscopied during replication and became -CGA-CCG-GAT-, what effect would there be on the sequence of the protein produced?

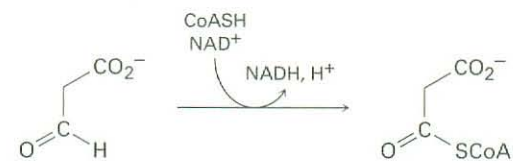
28.31 ■ Show the steps involved in a laboratory synthesis of the DNA fragment with the sequence CTAG.

28.32 ■ The final step in DNA synthesis is deprotection by treatment with aqueous ammonia. Show the mechanisms by which deprotection occurs at the points indicated in the following structure:



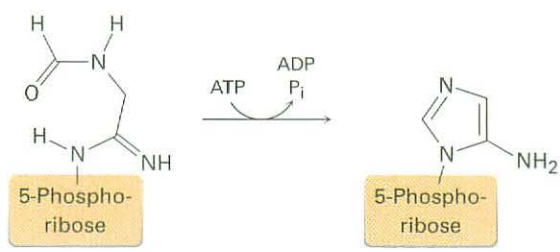
28.33 ■ Draw the structure of cyclic adenosine monophosphate (cAMP), a messenger involved in the regulation of glucose production in the body. Cyclic AMP has a phosphate ring connecting the 3′ and 5′ hydroxyl groups on adenosine.

28.34 The final step in the metabolic degradation of uracil is the oxidation of malonic semialdehyde to give malonyl CoA. Propose a mechanism.



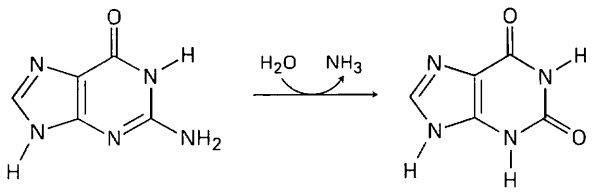
Malonic semialdehyde **Malonyl CoA**

28.35 One of the steps in the biosynthesis of a nucleotide called inosine monophosphate is the formation of aminoimidazole ribonucleotide from formylglycinamidine ribonucleotide. Propose a mechanism.



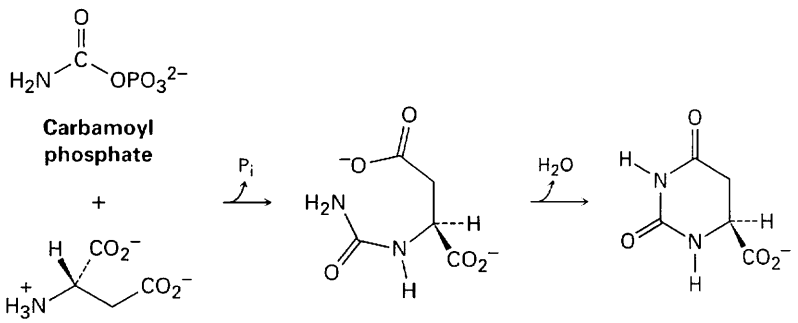
Formylglycinamidine ribonucleotide **Aminoimidazole ribonucleotide**

28.36 One of the steps in the metabolic degradation of guanine is hydrolysis to give xanthine. Propose a mechanism.



Guanine **Xanthine**

28.37 One of the steps in the biosynthesis of uridine monophosphate is the reaction of aspartate with carbamoyl phosphate to give carbamoyl aspartate followed by cyclization to form dihydroorotate. Propose mechanisms for both steps.



Carbamoyl phosphate + **Aspartate** **Carbamoyl aspartate** **Dihydroorotate**