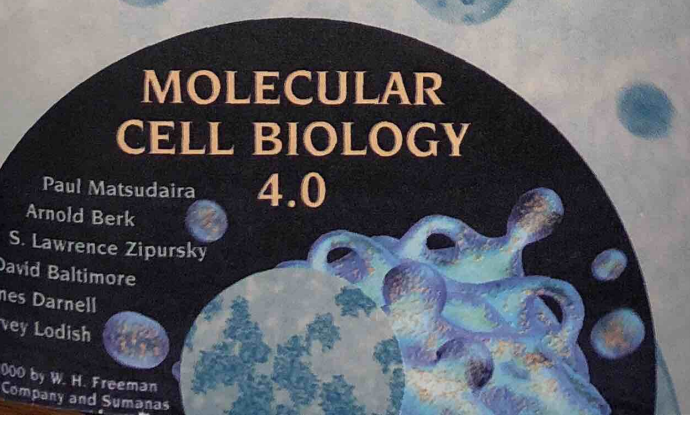
FOURTH EDITION

MOLECULAR CELL BIOLOGY

Harvey Lodish
Arnold Berk
S. Lawrence Zipursky
Paul Matsudaira
David Baltimore
James Darnell



Media Connected

W. H. FREEMAN AND COMPANY

EXECUTIVE EDITOR: Sara Tenney
DEVELOPMENT EDITORS: Katherine Ahr, Ruth Steyn, Kay Ueno
EDITORIAL ASSISTANT: Jessica Olshen
EXECUTIVE MARKETING MANAGER: John A. Britch
PROJECT EDITOR: Katherine Ahr
TEXT AND COVER DESIGNER: Victoria Tomaselli
PAGE MAKEUP: Michael Mendelsohn, Design 2000, Inc.
COVER ILLUSTRATION: Kenneth Eward
ILLUSTRATION COORDINATOR: John Smith, Network Graphics; Tamara Goldman, Bill Page
ILLUSTRATIONS: Network Graphics
PHOTO RESEARCHER: Jennifer MacMillan
PRODUCTION COORDINATOR: Paul W. Rohloff
MEDIA AND SUPPLEMENTS EDITORS: Tanya Awabdy, Adrie Kornasiewicz, Debra Siegel
MEDIA DEVELOPERS: Sumanas, Inc.
COMPOSITION: York Graphics Services, Inc.
MANUFACTURING: Von Hoffman Press

Library of Congress Cataloging-in-Publication Data

Molecular cell biology / Harvey Lodish p [et al.] – 4th ed.
p. cm.
Includes bibliographical references.
ISBN 0-7167-3136-3
1. Cytology. 2. Molecular biology. I. Lodish, Harvey F.
QH581.2.M655 1999
571.6–dc21 99-30831
CIP

Printed in the United States of America

W. H. Freeman and Company
41 Madison Avenue, New York, New York 10010
Houndsmills, Basingstoke RG21 6XS, England

Third Printing, 2001

Contents in Brief

Contents

PART II: Nuclear Control of Cellular Activity

9 Molecular Structure of Genes and Chromosomes

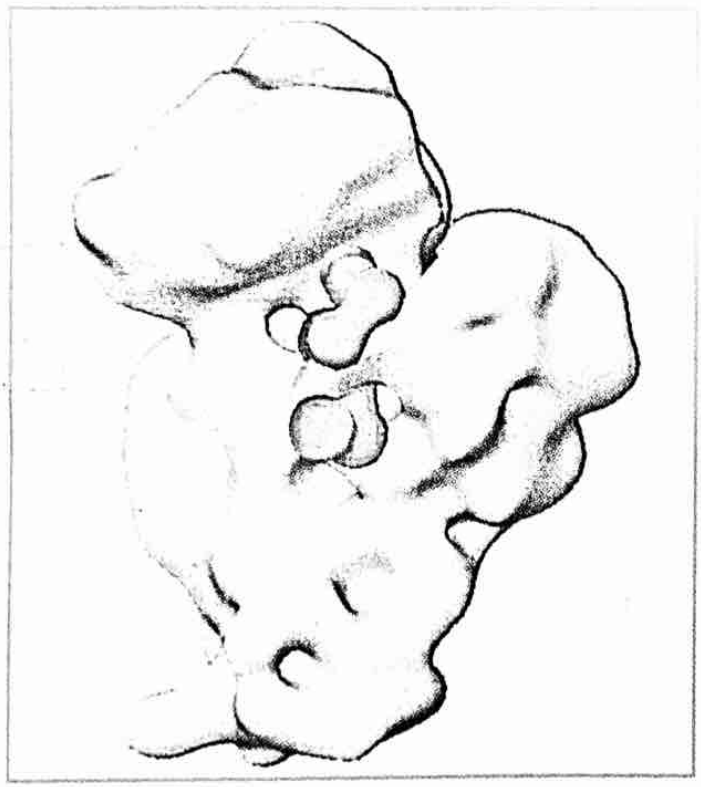
10 Regulation of Transcription Initiation

19 Cell Motility and Shape II: Microtubules and Intermediate Filaments

PART IV: Cell Interactions

20 Cell-to-Cell Signaling: Hormones and Receptors

4 Nucleic Acids, the Genetic Code, and the Synthesis of Macromolecules



Model of small ribosomal subunit (yellow) showing contacts with two tRNAs (red and green) at one point in the protein chain elongation cycle.

The extraordinary versatility of proteins in catalyzing chemical reactions and building cellular structures was featured in the previous chapter. In this chapter we consider the **nucleic acids,** the molecules that (1) contain the information prescribing amino acid sequence in proteins and (2) serve in the several cellular structures that choose, and then link into the correct order, the amino acids of a protein chain. **Deoxyribonucleic acid** (DNA) is the storehouse, or cellular library, that contains all the information required to build the cells and tissues of an organism. The exact duplication of this information in any species from generation to generation assures the genetic continuity of that species. The information is arranged in units identified by classical geneticists from Gregor Mendel through Thomas Hunt Morgan, and known now as **genes,** hereditary units controlling identifiable traits of an organism. In the process of **transcription,** the information stored in DNA is copied into **ribonucleic acid** (RNA), which has three distinct roles in protein synthesis. **Messenger RNA** (mRNA) carries the instructions from DNA that specify the correct order of amino acids during protein synthesis. The remarkably accurate, stepwise assembly of amino acids into proteins occurs by **translation** of mRNA. In this process, the information in mRNA is interpreted by a second type of RNA called **transfer RNA** (tRNA) with the aid of a third type of RNA, **ribosomal RNA** (rRNA), and its associated proteins. As the correct

OUTLINE

MEDIA CONNECTIONS

Focus: Basic Transcriptional Mechanism

Overview: Life Cycle of an mRNA

Focus: Protein Synthesis

Classic Experiment 4.1: Cracking the Genetic Code

amino acids are brought into sequence by tRNAs, they are linked by peptide bonds to make proteins.

Discovery of the structure of DNA in 1953 and the subsequent elucidation of the steps in the synthesis of DNA, RNA, and protein are the monumental achievements of the early days of molecular biology. To understand how DNA directs synthesis of RNA, which then directs assembly of proteins—the so-called *central dogma* of molecular biology—we first discuss the building blocks composing DNA and RNA, the cell's two primary nucleic acids, and their general structures. We then introduce the basic mechanisms of DNA and RNA chain synthesis, including a brief discussion of gene structure, which reveals why molecular processing is required to make functional RNA molecules. Next we outline the roles of mRNA, tRNA, and rRNA in protein synthesis. The chapter closes with a detailed description of the components and biochemical steps in the formation of proteins. Since the events of macromolecular synthesis are so central to all biological functions—growth control, differentiation, and the specialized chemical and physical properties of cells—they will arise again and again in later chapters. A firm grasp of the fundamentals of DNA, RNA, and protein synthesis is necessary to follow the subsequent discussions without difficulty.

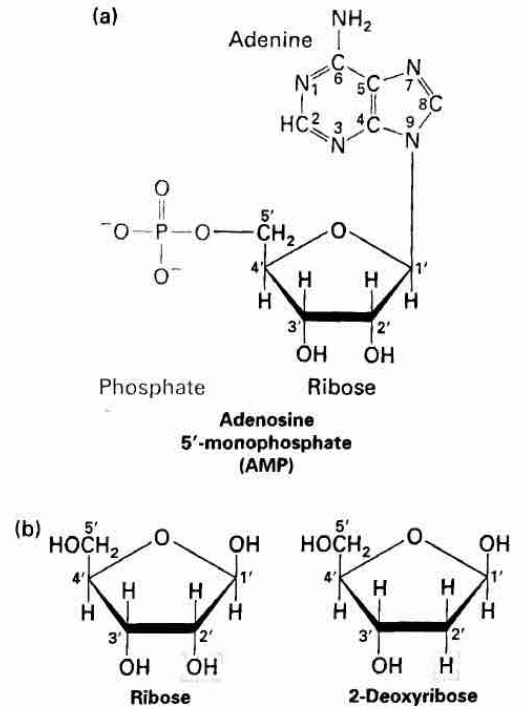
4.1 Structure of Nucleic Acids

DNA and RNA have great chemical similarities. In their *primary structures* both are linear **polymers** (multiple chemical units) composed of **monomers** (single chemical units), called **nucleotides.** Cellular RNAs range in length from less than one hundred to many thousands of nucleotides. Cellular DNA molecules can be as long as several hundred million nucleotides. These large DNA units in association with proteins can be stained with dyes and visualized in the light microscope as chromosomes.

Polymerization of Nucleotides Forms Nucleic Acids

DNA and RNA each consists of only four different nucleotides. All nucleotides have a common structure: a *phosphate* group linked by a phosphoester bond to a *pentose* (a five-carbon sugar molecule) that in turn is linked to an organic *base* (Figure 4-1a). In RNA, the pentose is *ribose;* in DNA, it is *deoxyribose* (Figure 4-1b). The only other difference in the nucleotides of DNA and RNA is that one of the four organic bases differs between the two polymers. The bases adenine, guanine, and cytosine are found in both DNA and RNA; thymine is found only in DNA, and uracil is found only in RNA. The bases are often abbreviated A, G, C, T, and U, respectively. For convenience the single letters are also used when long sequences of nucleotides are written out.

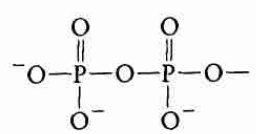
The base components of nucleic acids are heterocyclic compounds with the rings containing nitrogen and carbon. Adenine and guanine are **purines,** which contain a pair of fused rings;

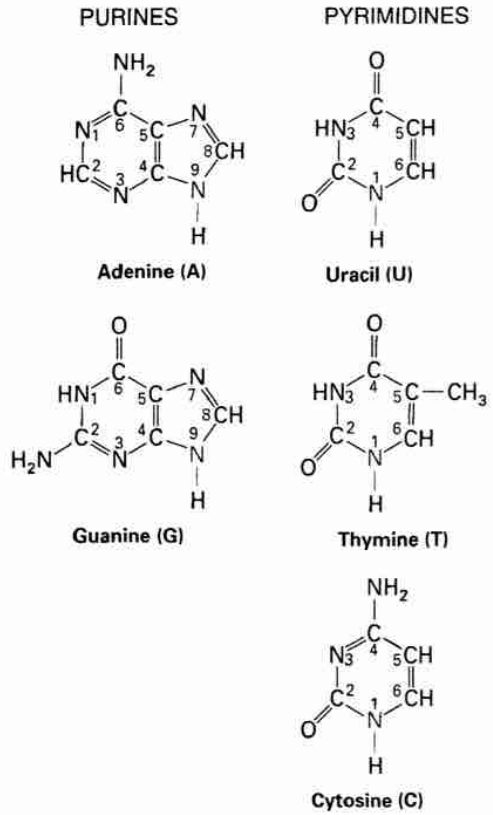


▲ **FIGURE 4-1 All nucleotides have a common structure.** (a) Chemical structure of adenosine 5′-monophosphate (AMP), a nucleotide that is present in RNA. All nucleotides are composed of a phosphate moiety, containing up to three phosphate groups, linked to the 5′ hydroxyl of a pentose sugar, whose 1′ carbon is linked to an organic base. By convention, the carbon atoms of the pentoses are numbered with primes. In natural nucleotides, the 1′ carbon is joined by a β linkage to the base, which is in the plane above the furanose ring, as is the phosphate. (b) Haworth projections of ribose and deoxyribose, the pentoses in nucleic acids.

cytosine, thymine, and uracil are **pyrimidines,** which contain a single ring (Figure 4-2). The acidic character of nucleotides is due to the presence of phosphate, which dissociates at the pH found inside cells, freeing hydrogen ions and leaving the phosphate negatively charged (see Figure 2-22). Because these charges attract proteins, most nucleic acids in cells are associated with proteins. In nucleotides, the 1′ carbon atom of the sugar (ribose or deoxyribose) is attached to the nitrogen at position 9 of a purine (N_9) or at position 1 of a pyrimidine (N_1).

Cells and extracellular fluids in organisms contain small concentrations of **nucleosides,** combinations of a base and a sugar without a phosphate. Nucleotides are nucleosides that have one, two, or three phosphate groups esterified at the 5′ hydroxyl. *Nucleoside monophosphates* have a single esterified phosphate (see Figure 4-1a), *diphosphates* contain a prophosphate group

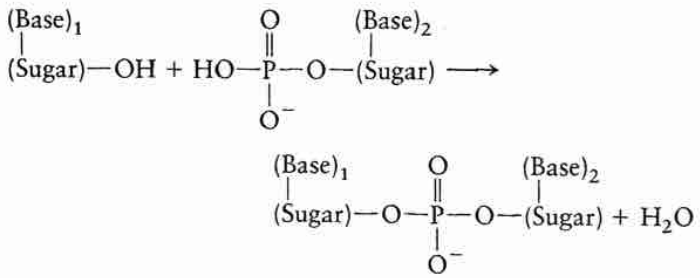




▲ **FIGURE 4-2 The chemical structures of the principal bases in nucleic acids.** In nucleic acids and nucleotides, nitrogen 9 of purines and nitrogen 1 of pyrimidines (red) are bonded to the 1′ carbon of ribose or deoxyribose.

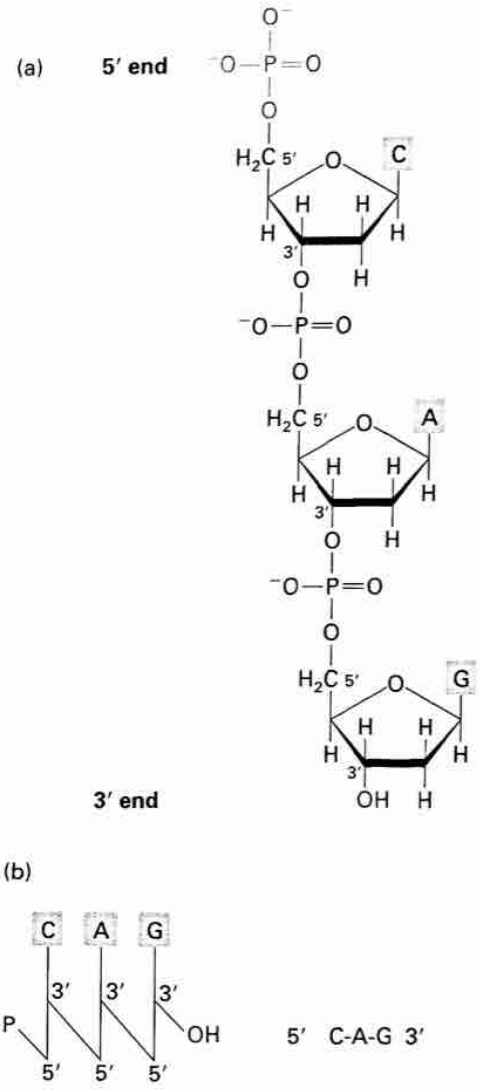
and *triphosphates* have a third phosphate. Table 4-1 lists the names of the nucleosides and nucleotides in nucleic acids and the various forms of nucleoside phosphates. As we will see later, the nucleoside triphosphates are used in the synthesis of nucleic acids. However, these compounds also serve many other functions in the cell: ATP, for example, is the most widely used energy carrier in the cell (see Figure 2-25), and GTP plays crucial roles in intracellular signaling and acts as an energy reservoir, particularly in protein synthesis.

When nucleotides polymerize to form nucleic acids, the hydroxyl group attached to the 3′ carbon of a sugar of one nucleotide forms an ester bond to the phosphate of another nucleotide, eliminating a molecule of water:



This condensation reaction is similar to that in which a peptide bond is formed between two amino acids (Chapter 3).

Thus a single nucleic acid strand is a phosphate-pentose polymer (a polyester) with purine and pyrimidine bases as side groups. The links between the nucleotides are called **phosphodiester bonds.** Like a polypeptide, a nucleic acid strand has an end-to-end chemical orientation: the *5′ end* has a free hydroxyl or phosphate group on the 5′ carbon of its terminal sugar; the *3′ end* has a free hydroxyl group on the 3′ carbon of its terminal sugar (Figure 4-3). This directionality, plus the fact that synthesis proceeds 5′ to 3′, has



▲ **FIGURE 4-3 Alternative ways of representing nucleic acid chains, in this case a single strand of DNA containing only three bases: cytosine (C), adenine (A), and guanine (G).** (a) Chemical structure of the trinucleotide CAG. Note the free hydroxyl group at the 3′ end and free phosphate group at the 5′ end. (b) Two common simplified methods of representing polynucleotides. In the "stick" diagram *(left)*, the sugars are indicated as vertical lines and the phosphodiester bonds as slanting lines; the bases are denoted by their single-letter abbreviations. In the simplest representation *(right)*, the bases are indicated by single letters. By convention, a polynucleotide sequence is always written in the 5′ → 3′ direction (left to right).

TABLE 4-1 Naming Nucleosides and Nucleotides

	Bases			
	Purines		Pyrimidines	
	Adenine (A)	Guanine (G)	Cytosine (C)	Uracil (U) Thymine [T]
Nucleosides in RNA	Adenosine	Guanosine	Cytidine	Uridine
Nucleosides in DNA	Deoxyadenosine	Deoxyguanosine	Deoxycytidine	Deoxythymidine
Nucleotides in RNA	Adenylate	Guanylate	Cytidylate	Uridylate
Nucleotides in DNA	Deoxyadenylate	Deoxyguanylate	Deoxycytidylate	Thymidylate
Nucleoside monophosphates	AMP	GMP	CMP	UMP
Nucleoside diphosphates	ADP	GDP	CDP	UDP
Nucleoside triphosphates	ATP	GTP	CTP	UTP
Deoxynucleoside mono-, di-, and triphosphates	dAMP, etc.			

given rise to the convention that polynucleotide sequences are written and read in the 5′ → 3′ direction (from left to right); for example, the sequence AUG is assumed to be (5′)AUG(3′). (Although, strictly speaking, the letters A, G, C, T, and U stand for bases, they are also often used in diagrams to represent the whole nucleotides containing these bases.) The 5′ → 3′ directionality of a nucleic acid strand is an extremely important property of the molecule.

The linear sequence of nucleotides linked by phosphodiester bonds constitutes the primary structure of nucleic acids. As we discuss in the next section, polynucleotides can twist and fold into three-dimensional conformations stabilized by noncovalent bonds; in this respect, they are similar to polypeptides. Although the primary structures of DNA and RNA are generally similar, their conformations are quite different. Unlike RNA, which commonly exists as a single polynucleotide chain, or strand, DNA contains two intertwined polynucleotide strands. This structural difference is critical to the different functions of the two types of nucleic acids.

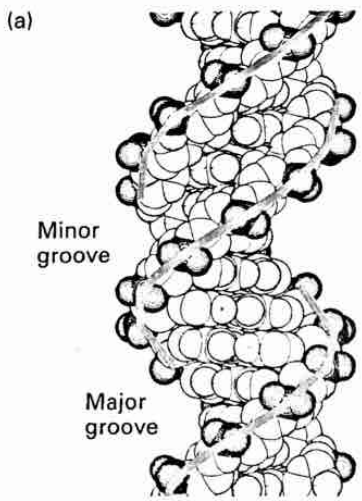
Native DNA Is a Double Helix of Complementary Antiparallel Chains

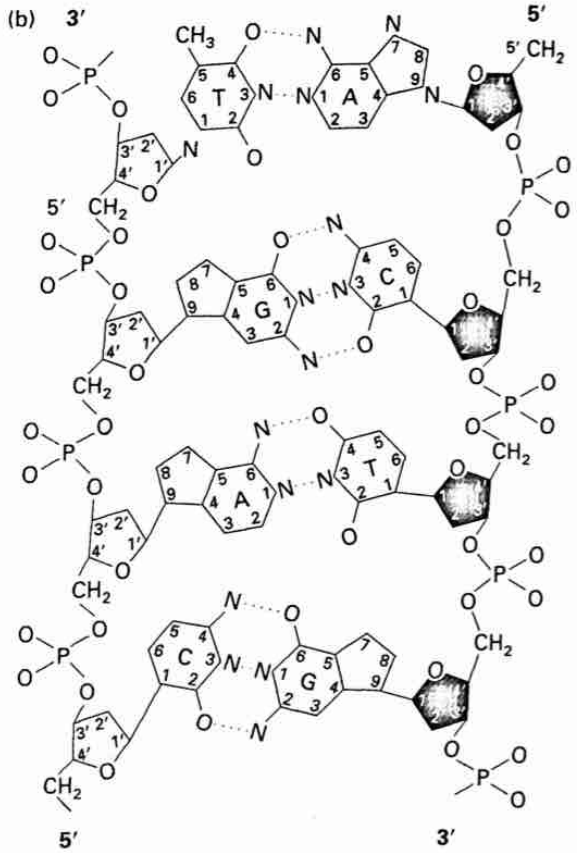
The modern era of molecular biology began in 1953 when James D. Watson and Francis H. C. Crick proposed correctly the double-helical structure of DNA, based on the analysis of x-ray diffraction patterns coupled with careful model building. A closer look at the "thread of life," as the DNA molecule is sometimes called, shows why the discovery of its basic structure suggests its function.

DNA consists of two associated polynucleotide strands that wind *together* through space to form a structure often described as a **double helix.** The two sugar-phosphate backbones are on the outside of the double helix, and the bases project into the interior. The adjoining bases in each strand stack on top of one another in parallel planes (Figure 4-4a). The orientation of the two strands is antiparallel; that is, their 5′ → 3′ directions are opposite. The strands are held in precise register by a regular base-pairing between the two strands: A is paired with T through two hydrogen bonds; G is paired with C through three hydrogen bonds (Figure 4-4b). This *base-pair complementarity* is a consequence of the size, shape, and chemical composition of the bases. The presence of thousands of such hydrogen bonds in a DNA molecule contributes greatly to the stability of the double helix. Hydrophobic and van der Waals interactions between the stacked adjacent base pairs also contribute to the stability of the DNA structure.

To maintain the geometry of the double-helical structure shown in Figure 4-4a, a larger purine (A or G) must pair with a smaller pyrimidine (C or T). In natural DNA, A almost always hydrogen bonds with T and G with C, forming A·T and G·C base pairs often called *Watson-Crick base pairs.* Two polynucleotide strands, or regions thereof, in which all the nucleotides form such base pairs are said to be **complementary.** However, in theory and in synthetic DNAs other interactions can occur. For example, a guanine (a purine) could theoretically form hydrogen bonds with a thymine (a pyrimidine), causing only a minor distortion in the helix. The space available in the helix also would allow pairing between the two pyrimidines cytosine and thymine. Although the nonstandard G·T and C·T base pairs are normally not found in DNA, G·U base pairs are quite common in double-helical regions that form within otherwise single-stranded RNA.

Two polynucleotide strands can, in principle, form either a right-handed or a left-handed helix (Figure 4-5). Because the geometry of the sugar-phosphate backbone is more compatible with the former, natural DNA is a right-handed

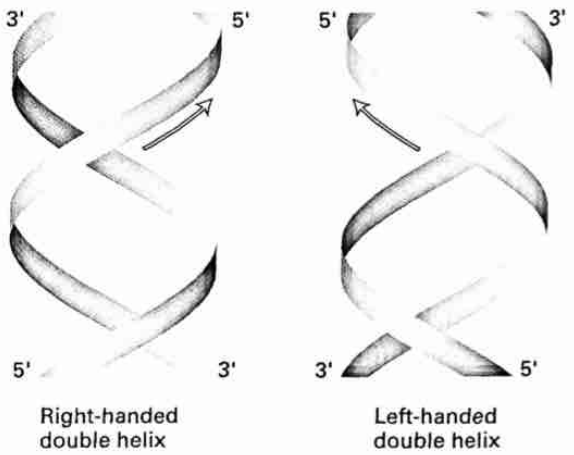




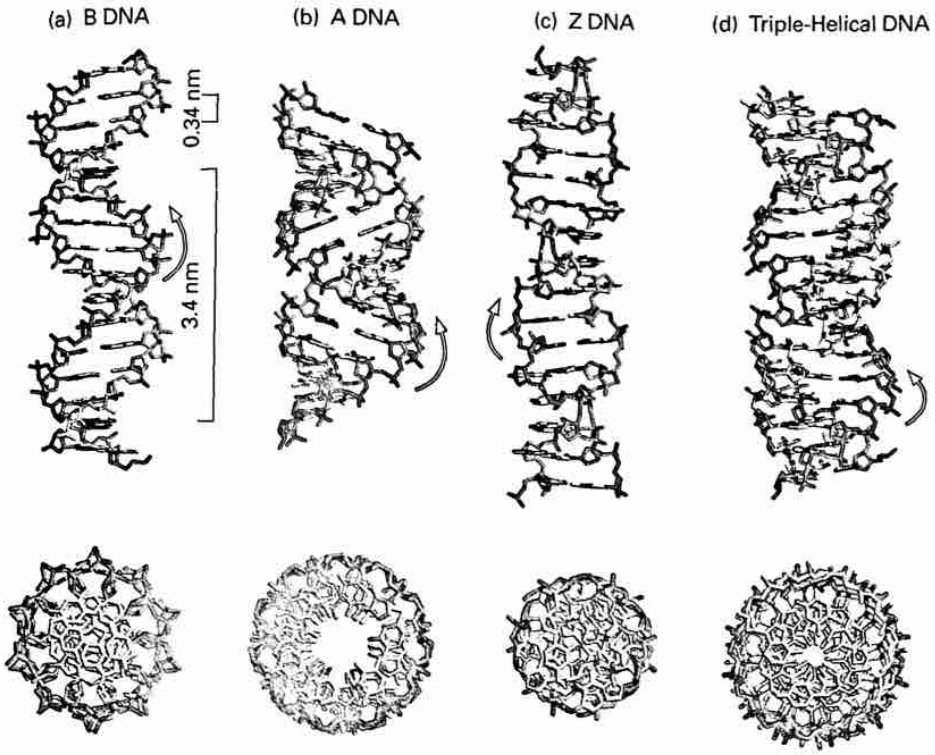
▲ **FIGURE 4-4 Two representations of contacts within the DNA double helix.** (a) Space-filling model of B DNA, the most common form of DNA in cells. The sugar and phosphate residues (gray) in each strand form the backbone, which is traced by a red line, showing the helical twist of the overall molecule. The bases project inward, but are accessible through major and minor grooves; a pair of bases from opposite strands in the major groove are highlighted in light and dark blue. The hydrogen bonds between the bases are in the center of the structure. (b) Stick diagram of the chemical structure of double-helical DNA, unraveled to show the sugar-phosphate backbones (sugar rings in green), base-paired bases (light blue and light red), and hydrogen bonds between the bases (dark red dotted lines). The backbones run in opposite directions; the 5′ and 3′ ends are named for the orientation of the 5′ and 3′ carbon atoms of the sugar rings. Each base pair has one purine base—adenine (A) or guanine (G)—and one pyrimidine base—thymine (T) or cytosine (C)—connected by hydrogen bonds. In this diagram, carbon atoms occur at the junction of every line with another line and no hydrogen atoms are shown. [Part (a) courtesy of A. Rich; part (b) from R. E. Dickerson, 1983, *Sci. Am.* **249**(6):94.]

helix. The x-ray diffraction pattern of DNA indicates that the stacked bases are regularly spaced 0.34 nm apart along the helix axis. The helix makes a complete turn every 3.4 nm; thus there are about 10 pairs per turn. This is referred to as the *B form* of DNA, the normal form present in most DNA stretches in cells (Figure 4-6a). On the outside of B-form DNA, the spaces between the intertwined strands form two helical grooves of different widths described as the *major* groove and the *minor* groove (see Figure 4-4a). Consequently, part of each base is accessible from outside the helix to both small and large molecules that bind to the DNA by contacting chemical groups within the grooves. These two binding surfaces of the DNA molecule are used by different classes of DNA-binding proteins.

In addition to the major B form of DNA, three additional structures have been described. In very low humidity, the crystallographic structure of B DNA changes to the *A form;* RNA-DNA and RNA-RNA helices also exist in this form. The A form is more compact than the B form,



▲ **FIGURE 4-5 Two possible helical forms of DNA are mirror images of each other.** The geometry of the sugar-phosphate backbone of DNA causes natural DNA to be right-handed. (*Right-handed* and *left-handed* are defined by convention.)



◀ **FIGURE 4-6 Models of various DNA structures that are known to exist.** The sugar-phosphate backbone of each chain is on the outside in all structures (one red and one blue) with the bases (silver) oriented inward. Side views are shown at the top, and views along the helical axis at the bottom. (a) The B form of DNA, the usual form in cells, is characterized by a helical turn every 10 base pairs (3.4 nm); adjacent stacked base pairs are 0.34 nm apart. The major and minor grooves are also visible. (b) The more compact A form of DNA has 11 base pairs per turn and exhibits a large tilt of the base pairs with respect to the helix axis. In addition, the A form has a central hole *(bottom)*. This helical form is adopted by RNA-DNA and RNA-RNA helices. (c) Z DNA is a left-handed helix and has a zig-zag (hence "Z") appearance. (d) A triple-helical structure can occur in stretches of DNA where all purines (A, G) in one strand are matched by all pyrimidines (T, C) in the other strand. Such stretches can accommodate a third polypyrimidine strand (yellow). [Courtesy of C. Kielkopf and P. B. Dervan.]

having 11 bases per turn, and the stacked bases are tilted (Figure 4-6b). Short DNA molecules composed of alternating purine-pyrimidine nucleotides (especially Gs and Cs) adopt an alternative left-handed configuration instead of the normal right-handed helix. This structure is called *Z DNA* because the bases seem to zigzag when viewed from the side (Figure 4-6c). It is entirely possible that both A-form and Z-form stretches of DNA exist in cells.

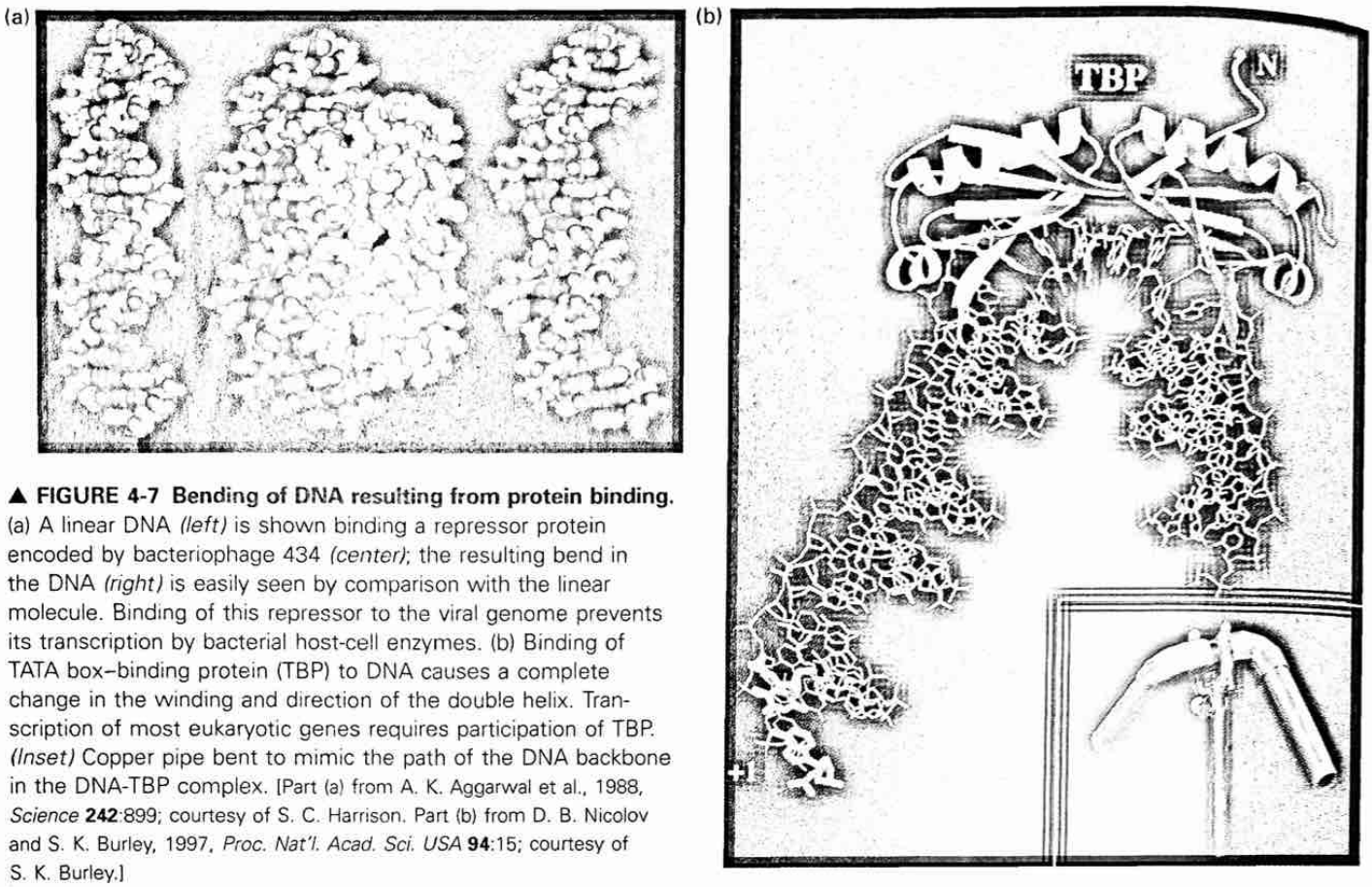
Finally, a triple-stranded DNA structure can also exist at least in the test tube, and possibly during recombination and DNA repair. For example, when synthetic polymers of poly(A) and polydeoxy(U) are mixed, a three-stranded structure is formed (Figure 4-6d). Further, long homopolymeric stretches of DNA composed of C and T residues in one strand and A and G residues in the other can be targeted by short matching lengths of poly(C+T). The synthetic oligonucleotide can insert as a third strand, binding in a sequence-specific manner by so-called *Hoogsteen base pairs*. Specific cleavage of the DNA at the site where the triple helix ends can be achieved by attaching a chemical cleaving agent (e.g., Fe^{2+}-EDTA) to the short oligodeoxynucleotide that makes up the third strand. Such reactions may be useful in studying site-specific DNA damage in cells.

By far the most important modifications in standard B-form DNA come about as a result of protein binding to specific DNA sequences. Although the multitude of hydrogen and hydrophobic bonds between the polynucleotide strands provide stability to DNA, the double helix is somewhat flexible about its long axis. Unlike the α helix in proteins (see Figure 3-6), there are no hydrogen bonds between successive residues in a DNA strand. This property allows DNA to bend when complexed with a DNA-binding protein. Crystallographic analyses of proteins bound to particular regions of DNA have conclusively demonstrated departures from the standard B-DNA structure in protein-DNA complexes. Two examples of DNA deformed by contact with proteins are shown in Figure 4-7. The specific DNA-protein contacts that occur in these tightly bound complexes have the ability both to untwist the DNA and to bend the axis of the helix. Although DNA in cells likely exists in the B form most of the time, particular regions bound to protein clearly depart from the standard conformation.

DNA Can Undergo Reversible Strand Separation

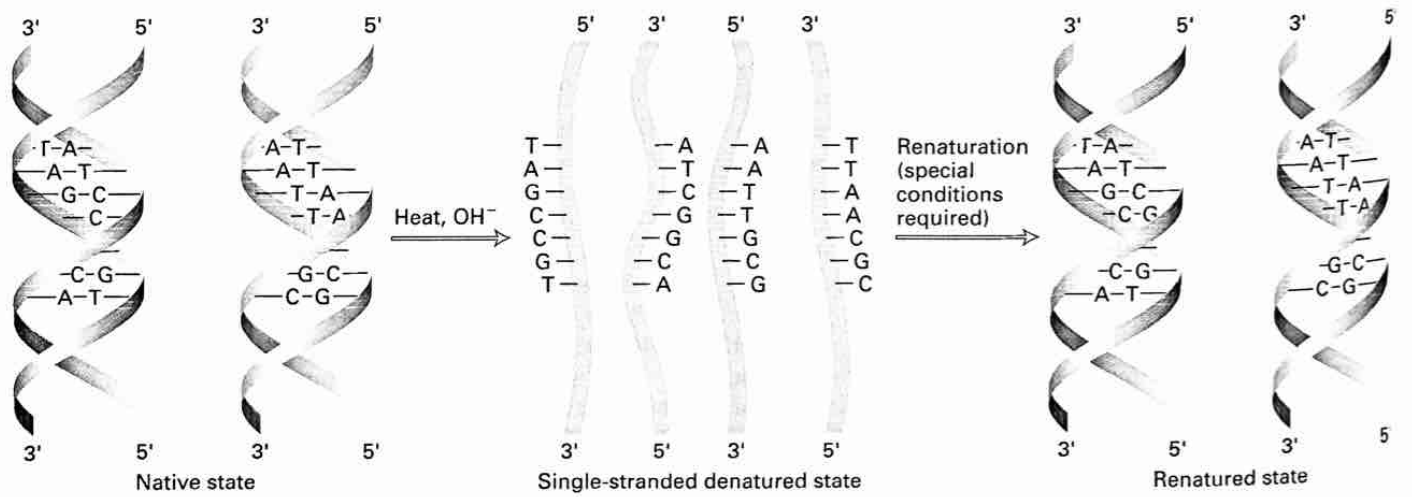
In DNA replication and in the copying of RNA from DNA, the strands of the helix must separate at least temporarily. As we discuss later, during DNA synthesis two new strands are made (one copied from each of the original strands), resulting in two double helices identical with the original one. In the case of copying the DNA template to make RNA, the RNA is released and the two DNA strands reassociate with each other.

The unwinding and separation of DNA strands, referred to as **denaturation**, or "melting," can be induced experimentally. For example, if a solution of DNA is heated, the thermal energy increases molecular motion, eventually breaking the hydrogen bonds and other forces that stabilize the double helix, and the strands separate (Figure 4-8). This

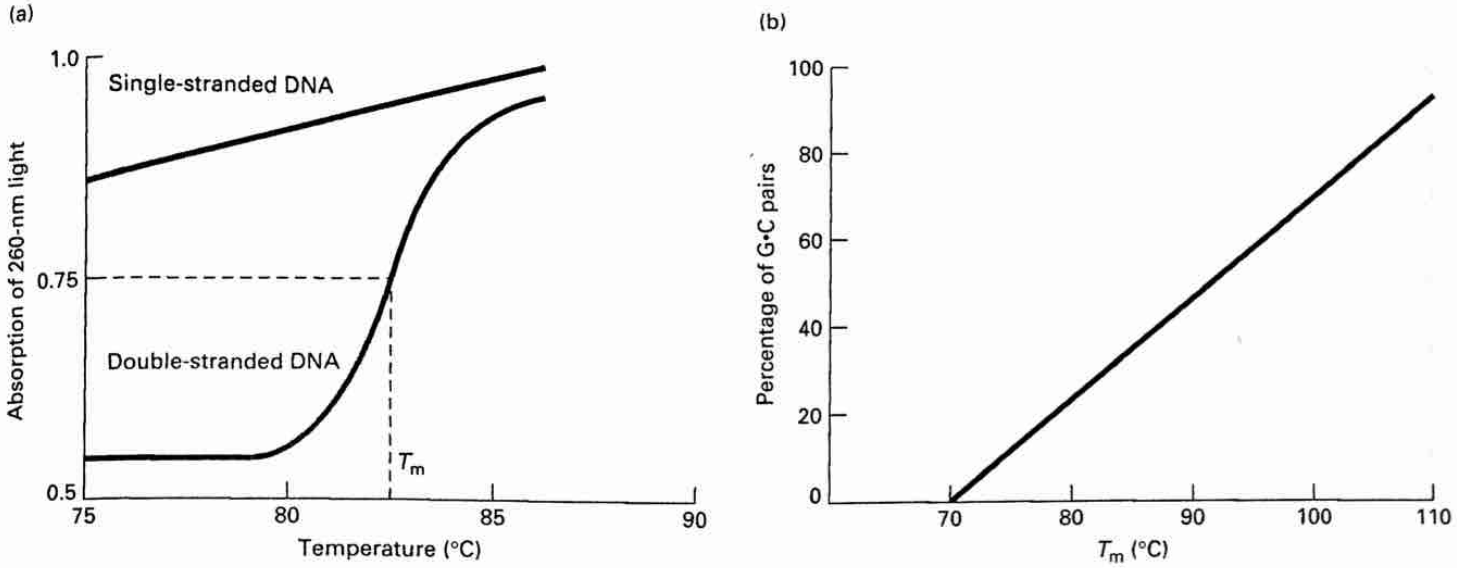


▲ FIGURE 4-7 Bending of DNA resulting from protein binding. (a) A linear DNA *(left)* is shown binding a repressor protein encoded by bacteriophage 434 *(center)*; the resulting bend in the DNA *(right)* is easily seen by comparison with the linear molecule. Binding of this repressor to the viral genome prevents its transcription by bacterial host-cell enzymes. (b) Binding of TATA box–binding protein (TBP) to DNA causes a complete change in the winding and direction of the double helix. Transcription of most eukaryotic genes requires participation of TBP. *(Inset)* Copper pipe bent to mimic the path of the DNA backbone in the DNA-TBP complex. [Part (a) from A. K. Aggarwal et al., 1988, *Science* **242**:899; courtesy of S. C. Harrison. Part (b) from D. B. Nicolov and S. K. Burley, 1997, *Proc. Nat'l. Acad. Sci. USA* **94**:15; courtesy of S. K. Burley.]

melting of DNA changes its absorption of ultraviolet (UV) light (in the 260-nm range), which is routinely used to measure DNA concentration because of the high absorbance of UV light by nucleic acid bases. Native double-stranded DNA absorbs about one-half as much light at 260 nm as does the equivalent amount of single-stranded DNA (Figure 4-9a). Thus, as DNA denatures, its absorption of UV light increases. Near the denaturation temperature, a small increase in temperature causes an abrupt, near simultaneous, loss of the multiple, weak, cooperative interactions holding the two strands together, so that denaturation rapidly occurs throughout the entire length of the DNA.

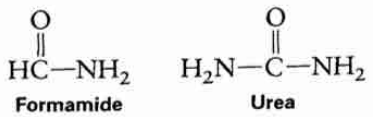


▲ FIGURE 4-8 The denaturation and renaturation of double-stranded DNA molecules.



▲ FIGURE 4-9 Light absorption and temperature in DNA denaturation. (a) Melting of doubled-stranded DNA can be monitored by the absorption of ultraviolet light at 260 nm. As regions of double-stranded DNA unpair, the absorption of light by those regions increases almost twofold. The temperature at which half the bases in a double-stranded DNA sample have denatured is denoted T_m (for temperature of melting). Light absorption by single-stranded DNA changes much less as the temperature is increased. (b) The T_m is a function of the G • C content of the DNA; the higher the G • C percentage, the greater the T_m.

The *melting temperature*, T_m, at which the strands of DNA will separate depends on several factors. Molecules that contain a greater proportion of G·C pairs require higher temperatures to denature because the three hydrogen bonds in G·C pairs make them more stable than A·T pairs with two hydrogen bonds (see Figure 4-4b). Indeed, the percentage of G·C base pairs in a DNA sample can be estimated from its T_m (Figure 4-9b). In addition to heat, solutions of low ion concentration destabilize the double helix, causing it to melt at lower temperatures. DNA is also denatured by exposure to other agents that destabilize hydrogen bonds, such as alkaline solutions and concentrated solutions of formamide or urea:



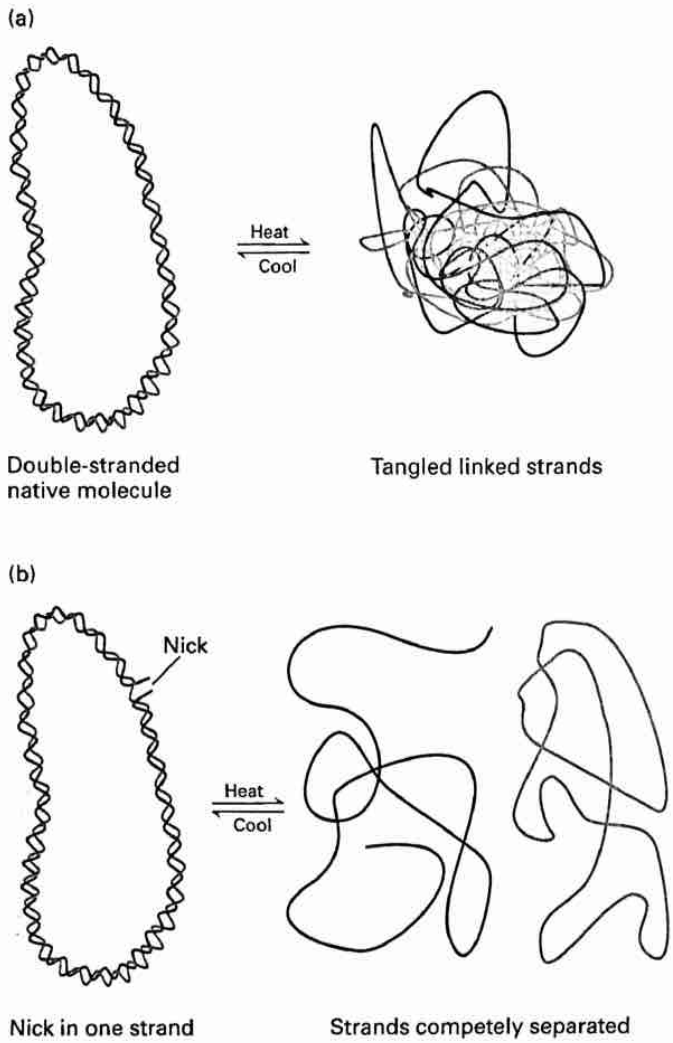
The single-stranded DNA molecules that result from denaturation form random coils without a regular structure. Lowering the temperature or increasing the ion concentration causes the two complementary strands to reassociate into a perfect double helix (see Figure 4-8). The extent of such *renaturation* is dependent on time, the DNA concentration, and the ionic content of the solution. Two DNA strands not related in sequence will remain as random coils and will not renature and, most important, will not greatly inhibit complementary DNA partner strands from finding each other. Denaturation and renaturation of DNA are the basis of nucleic acid **hybridization**, a powerful technique used to study the relatedness of two DNA samples and to detect and isolate specific DNA molecules in a mixture containing numerous different DNA sequences (Chapter 7).

Many DNA Molecules Are Circular

All prokaryotic genomic DNAs and many viral DNAs are circular molecules. Circular DNA molecules also occur in mitochondria, which are present in almost all eukaryotic cells, and in chloroplasts, which are present in plants and some unicellular eukaryotes.

Each of the two strands in a circular DNA molecule forms a closed structure without free ends. Just as is the case for linear DNA, elevated temperatures or alkaline pH destroy the hydrogen bonds and other interactions that stabilize double-helical circular DNA molecules. Unlike linear DNA, however, the two strands of circular DNA cannot unwind and separate; attempts to melt such DNA result in an interlocked, tangled mass of single-stranded DNA (Figure 4-10a).

Only if a native circular DNA is *nicked* (i.e., one of the strands is cut), will the two strands unwind and separate when the molecule is denatured. In this case, one of the separated strands is circular, and the other is linear (Figure 4-10b). Nicking of circular DNA occurs naturally during DNA replication and can be induced experimentally with a low concentration of deoxyribonuclease (a DNA-degrading enzyme), so that only a single phosphodiester bond in the molecule is cleaved. The study of circular DNA molecules lacking free ends first uncovered the complicated geometric shape changes that the double-stranded DNA molecule must undergo when the strands are not free to separate.



▲ **FIGURE 4-10 Denaturation of circular DNA.** (a) If both strands are closed circles, denaturation disrupts the double helix, but the two single strands become tangled about each other and cannot separate. (b) If one or both strands are nicked, however, the two strands will separate on thermal denaturation.

Local Unwinding of DNA Induces Supercoiling

So far we have described DNA as a long regular helical structure that can have local perturbations, especially due to protein binding. In addition, when the two ends of a DNA molecule are fixed, the molecule exhibits a superstructure under certain conditions. This occurs when the base pairing is interrupted and a local region unwinds. The stress induced by unwinding is relieved by twisting of the double helix on itself, forming **supercoils** (Figure 4-11). Unwinding and subsequent supercoiling occurs during replication, transcription, and binding of many proteins to circular DNAs or to long DNA loops whose ends are fixed within eukaryotic chromosomes. Supercoiling is recognized and regulated by enzymes called **topoisomerases.** As discussed in later chapters, these enzymes have an important role in both DNA replication and the transcription of DNA into RNA.

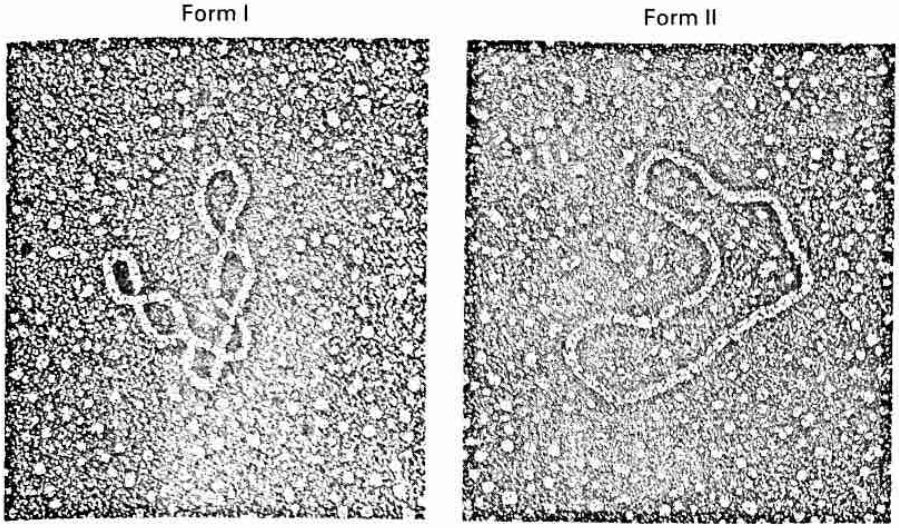
RNA Molecules Exhibit Varied Conformations and Functions

As noted earlier, the primary structure of RNA is generally similar to that of DNA; however, the sugar component (ribose) of RNA has an additional hydroxyl group at the 2′ position (see Figure 4-1b), and thymine in DNA is replaced by uracil in RNA (see Figure 4-2). The hydroxyl group on C_2 of ribose makes RNA more chemically labile than DNA and provides a chemically reactive group that takes part in RNA-mediated enzymatic events. As a result of this lability, RNA is cleaved into mononucleotides by alkaline solution, whereas DNA is not. Like DNA, RNA is a long polynucleotide that can be double-stranded or single-stranded, linear or circular. It can also participate in a hybrid helix composed of one RNA strand and one DNA strand; this hybrid has a slightly different conformation than the common B form of DNA.

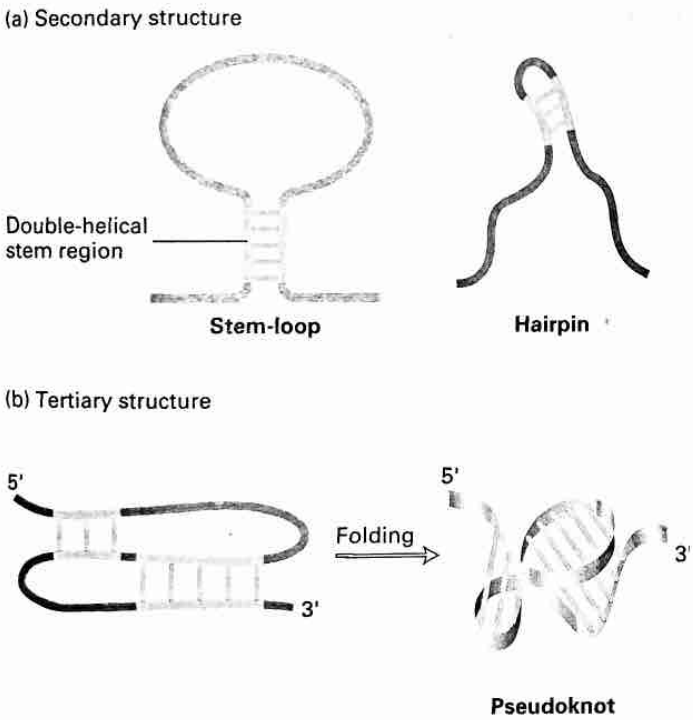
Unlike DNA, which exists primarily in a single, very long three-dimensional structure, the double helix, the various types of RNA exhibit different conformations. Differences in the sizes and conformations of the various types of RNA permit them to carry out specific functions in a cell. The simplest secondary structures in single-stranded RNAs are formed by pairing of complementary bases. "Hairpins" are formed by pairing of bases within ≈5–10 nucleotides of each other, and "stem-loops" by pairing of bases that are separated by ≈50 to several hundred nucleotides (Figure 4-12a). These simple folds can cooperate to form more complicated tertiary structures, one of which is termed a "pseudoknot" (Figure 4-12b).

As discussed in detail later, tRNA molecules adopt a well-defined three-dimensional architecture in solution that is crucial in protein synthesis. Larger rRNA molecules also have locally well defined three-dimensional structures, with more flexible links in between. Secondary and tertiary structures also have been recognized in mRNA, particularly near the ends of molecules. These recently discovered structures are under active study. Clearly, then, RNA molecules are like proteins in that they have structured domains connected by less structured, flexible stretches.

The folded domains of RNA molecules not only are structurally analogous to the α helices and β strands found in proteins, but in some cases also have catalytic capacities. Such catalytic RNAs, called **ribozymes,** can cut RNA chains. Some RNA domains also can catalyze *RNA splicing,* a remarkable process in which an internal RNA sequence, an **intron,** is cut and removed and the two resulting chains, the **exons,** are sealed together. This process occurs during formation of the majority of functional mRNA molecules in eukaryotic cells, and also occurs in bacteria and archaea. Remarkably, some RNAs carry out *self-splicing,* with the catalytic activity residing in the intron sequence. The mechanisms of splicing and self-splicing are discussed in detail in Chapter 11. As noted later in this chapter, rRNA is thought to play a catalytic role in the formation of peptide bonds during protein synthesis.



◀ **FIGURE 4-11 Supercoiling in electron micrographs of DNA isolated from the SV40 virus.** When isolated SV40 DNA is separated from its associated protein, the DNA duplex is underwound and assumes the supercoiled configuration (form I). If one strand is nicked, the strands can rewind, producing the relaxed-circle configuration (form II), which lacks supercoils. Only a few of the possible supercoils are visualized in the left photograph.



▲ **FIGURE 4-12 RNA secondary and tertiary structures.** (a) Stem-loops, hairpins, and other secondary structures can form by base pairing between distant complementary segments of an RNA molecule. In stem-loops, the single-stranded loop (dark red) between the base-paired helical stem (light red) may be hundreds or even thousands of nucleotides long, whereas in hairpins, the short turn may contain as few as 6–8 nucleotides. (b) Interactions between the flexible loops may result in further folding to form tertiary structures such as the pseudoknot. This tertiary structure resembles a figure-eight knot, but the free ends do not pass through the loops, so no knot is actually formed. [Part (b) adapted from C. W. A. Pleij et al., 1985, *Nucl. Acids Res.* **13**:1717.]

In this chapter, we focus on the functions of mRNA, tRNA, and rRNA in **gene expression**—the process of getting the information in DNA converted into proteins. In later chapters we will encounter other RNAs, often associated with proteins, that participate in other cell functions.

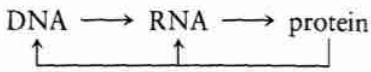
SUMMARY Structure of Nucleic Acids

- Deoxyribonucleic acid (DNA), the genetic material, carries information to specify the amino acid sequences of proteins. It is transcribed into several types of ribonucleic acid (RNA) including messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA), which function in protein synthesis.
- Both DNA and RNA are long, unbranched polymers of nucleotides. Each nucleotide consists of a heterocyclic base linked via a five-carbon sugar (deoxyribose or ribose) to a phosphate group (see Figure 4-1).
- DNA and RNA each contain four different bases (see Figure 4-2). The purines adenine (A) and guanine (G) and the pyrimidine cytosine (C) are present in both DNA and RNA. The pyrimidine thymine (T) present in DNA is replaced by the pyrimidine uracil (U) in RNA.
- The bases in nucleic acids can interact via hydrogen bonds. The standard Watson-Crick base pairs are G·C, A·T (in DNA), and A·U (in RNA). Base pairing stabilizes the native three-dimensional structures of DNA and RNA.
- Adjacent nucleotides in a polynucleotide are linked by phosphodiester bonds. The entire strand has a chemical directionality: the 5′ end with a free hydroxyl or phosphate group on the 5′ carbon of the sugar, and the 3′ end with a free hydroxyl group on the 3′ carbon of the sugar (see Figure 4-3). Polynucleotide sequences are always written in the 5′ → 3′ direction (left to right).

- Natural DNA (B DNA) contains two complementary polynucleotide strands wound together into a regular right-handed double helix with the bases on the inside and the two sugar-phosphate backbones on the outside (see Figure 4-6a). Base pairing (A·T and G·C) and hydrophobic interactions between adjacent bases in the same strand stabilize this native structure.
- Binding of protein to DNA can deform its helical structure, causing local bending or unwinding of the DNA molecule.
- Heat causes the DNA strands to separate (denature). The melting temperature of DNA increases with the percentage of G·C base pairs. Under suitable conditions, separated complementary nucleic acid strands will renature.
- Local unwinding of the DNA helix induces stress, which is relieved by twisting of the molecule on itself, forming supercoils. This process is regulated by topoisomerases, which can add or remove supercoils.
- Natural RNAs are single-stranded polynucleotides that form well-defined secondary and tertiary structures (see Figure 4-12). Some RNAs, called *ribozymes*, have catalytic activity.

4.2 Synthesis of Biopolymers: Rules of Macromolecular Carpentry

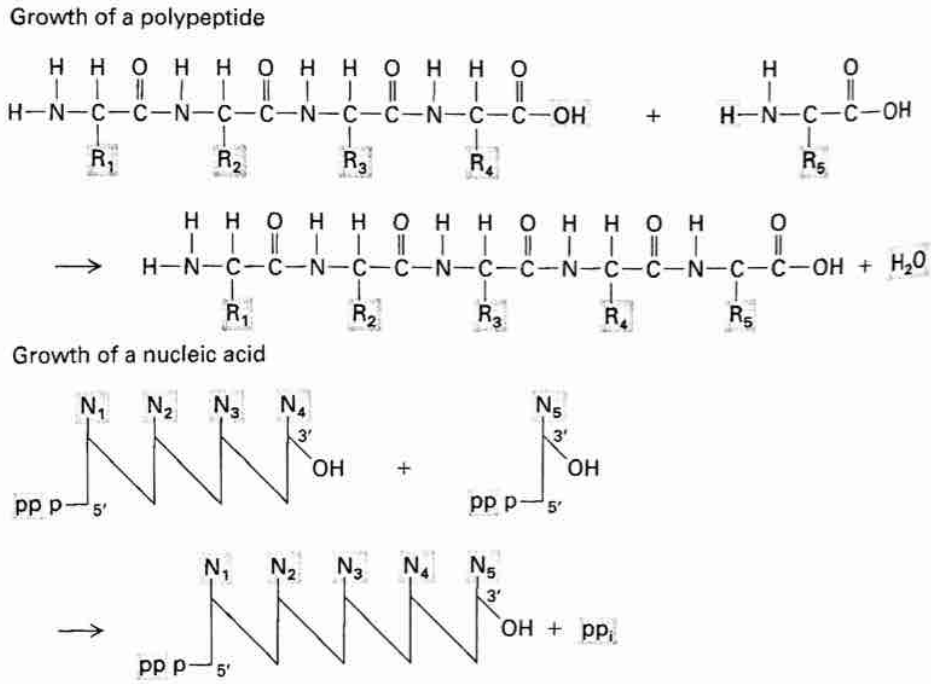
According to the central dogma of molecular biology, DNA directs the synthesis of RNA, and RNA then directs the synthesis of proteins. However, the one-way flow of information posited by the central dogma—that is, DNA → RNA → protein—does not reflect the role of proteins in facilitating the information flow. A more accurate way of representing the relationship between the synthesis of DNA, RNA, and proteins in all cells would look like



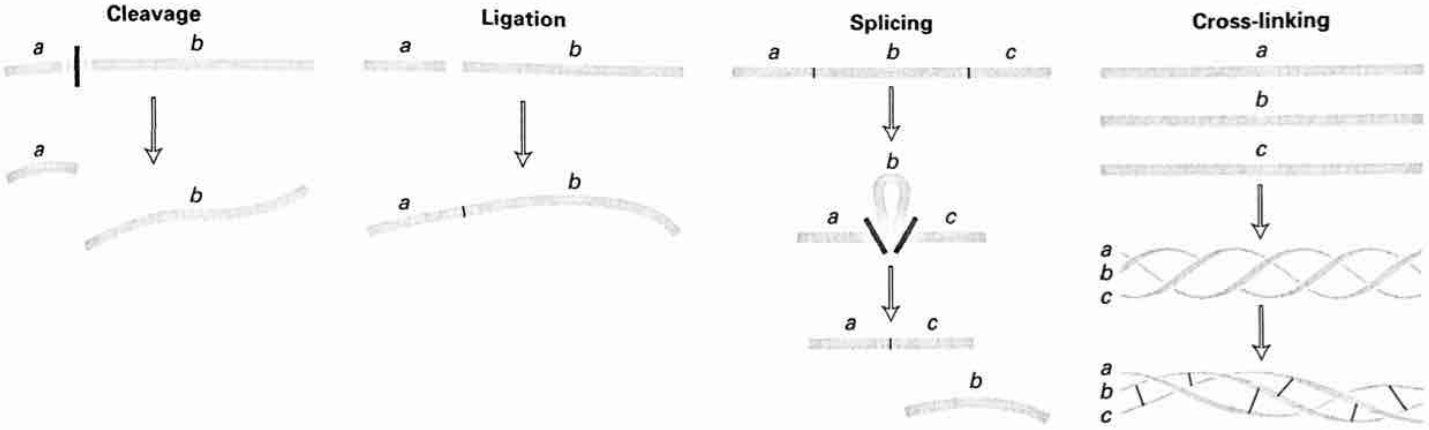
indicating that special proteins catalyze the synthesis of both RNA and DNA.

Interestingly, although the primary and higher-order structures of nucleic acids and proteins differ radically, several common principles apply to the synthesis of both types of macromolecules. Research over the past 30 years has revealed that giant molecular machines, composed largely of proteins, assemble nucleotides into RNA and DNA and amino acids into proteins. Later in this chapter and in subsequent chapters, we examine in detail the molecular machines that synthesize nucleic acids and proteins. To set the stage for these more detailed discussions, we review four general rules that govern the synthesis of both polypeptide and polynucleotide chains:

- *Proteins and nucleic acids are made up of a limited number of different monomeric building blocks.* Although the number of amino acids is theoretically limitless, and several dozen have been identified as metabolic products in various organisms, only 20 different amino acids are used in making proteins. Likewise, only five nitrogenous bases are used to construct RNA and DNA in cells. Cell-free enzyme preparations and cells in culture can be "fooled" into incorporating chemical relatives of the five bases or 20 amino acids, but this almost never happens in nature.



▶ FIGURE 4-13 Chain elongation in the in vivo synthesis of both proteins and nucleic acids proceeds by sequential addition of monomeric units—amino acids and nucleotides, respectively. Elongation of a polypeptide chain proceeds from the N-terminus to the C-terminus. (R_1, R_2, etc., denote side chains of amino acids.) Elongation of a polynucleotide chain proceeds from the 5′ end to the 3′ end. As each nucleotide is added, a pyrophosphate group (PP_i) is released, but the 5′ end of the first nucleotide in the chain retains its triphosphate group. (N_1, N_2, etc., denote purine and pyrimidine bases.)



▲ **FIGURE 4-14 Proteins and nucleic acids typically undergo one or more modifications after the basic polymeric chains are formed.** RNA, DNA, and protein can be cleaved; DNA pieces must be ligated during synthesis. Splicing, the removal of an internal sequence and reunion of the end pieces, often occurs in RNA synthesis; a similar reaction (intein removal) occurs in the formation of some proteins. Covalent chemical cross-linking between protein chains is common, but cross-linking is rare in nucleic acids. Besides the modifications to the polymeric chains illustrated here, individual chemical groups (e.g., phosphate, nitrate, sulfate, methyl, and acetyl groups) may be added. Various types of postpolymerization changes are discussed in later chapters.

- *The monomers are added one at a time.* Biological polymers theoretically could be built by aligning their component monomeric units in the correct order on a template, or mold, and *simultaneously* fusing them all. But cells do not use this mechanism. Rather, the assembly of proteins and nucleic acids occurs through the step-by-step addition of monomers.
- *Each polypeptide and polynucleotide chain has a specific starting point, and growth proceeds in one direction to a fixed terminus.* Synthesis of chains begins and ends at well-defined "start" and "stop" signals. Monomer addition proceeds from the amino (NH_2-) terminus to the carboxyl (COOH-) terminus in proteins, and from the 5′ end to the 3′ end in nucleic acids (Figure 4-13).
- *The primary synthetic product is often modified.* The initially synthesized form of a polypeptide or polynucleotide chain is often inactive or incomplete and undergoes modification to yield an active molecule (Figure 4-14). Protein chains are often shortened, and RNA chains almost always are; DNA chains are made in pieces and then linked together. Distant segments of long RNA chains may be spliced together, eliminating the intervening sequence (intron). A similar process occurs in some proteins (see Figure 3-17). Polypeptide chains sometimes are cross-linked by covalent bonds to form a functional protein. Finally, certain chemical groups may be added either during the formation of the chain or after its synthesis is complete: e.g., methyl groups to specific sites in DNA, RNA, and proteins; phosphate groups and a wide variety of oligosaccharides to proteins. The details of this macromolecular carpentry are discussed in later chapters.

4.3 Nucleic Acid Synthesis

The ordered assembly of deoxyribonucleotides into DNA and of ribonucleotides into RNA involves somewhat simpler cellular mechanisms than the correct assembly of the amino acids in a protein chain. Here we consider a few general principles governing the formation of polynucleotide chains in cells and briefly discuss some properties of the enzymes that carry out such synthesis. We also describe the steps in the production of mRNA and examine how and why this process differs in bacteria and eukaryotes. Later chapters cover the mechanism of DNA replication and its control during cell growth and division, and the mechanism and the control of the synthesis of specific mRNAs during differentiation (Chapters 10 and 12).

Both DNA and RNA Chains Are Produced by Copying of Template DNA Strands

The regular pairing of bases in the double-helical DNA structure suggested to Watson and Crick a mechanism of DNA synthesis. Their proposal that new strands of DNA are synthesized by copying of parental strands of DNA has proved to be correct.

The DNA strand that is copied to form a new strand is called a **template.** The information in the template is preserved: although the first copy has a complementary sequence, not an identical one, a copy of the copy produces the original (template) sequence again. In the replication of a double-stranded, or *duplex,* DNA molecule, both original (parental) DNA strands are copied. When copying is finished, the two new duplexes, each consisting of one of the two original strands plus its copy, separate from each other.

In some viruses, single-stranded RNA molecules function as templates for synthesis of complementary RNA or DNA chains (Chapter 7). However, the vast majority of RNA and DNA in cells is synthesized from preexisting duplex DNA.

Nucleic Acid Strands Grow in the 5′ → 3′ Direction

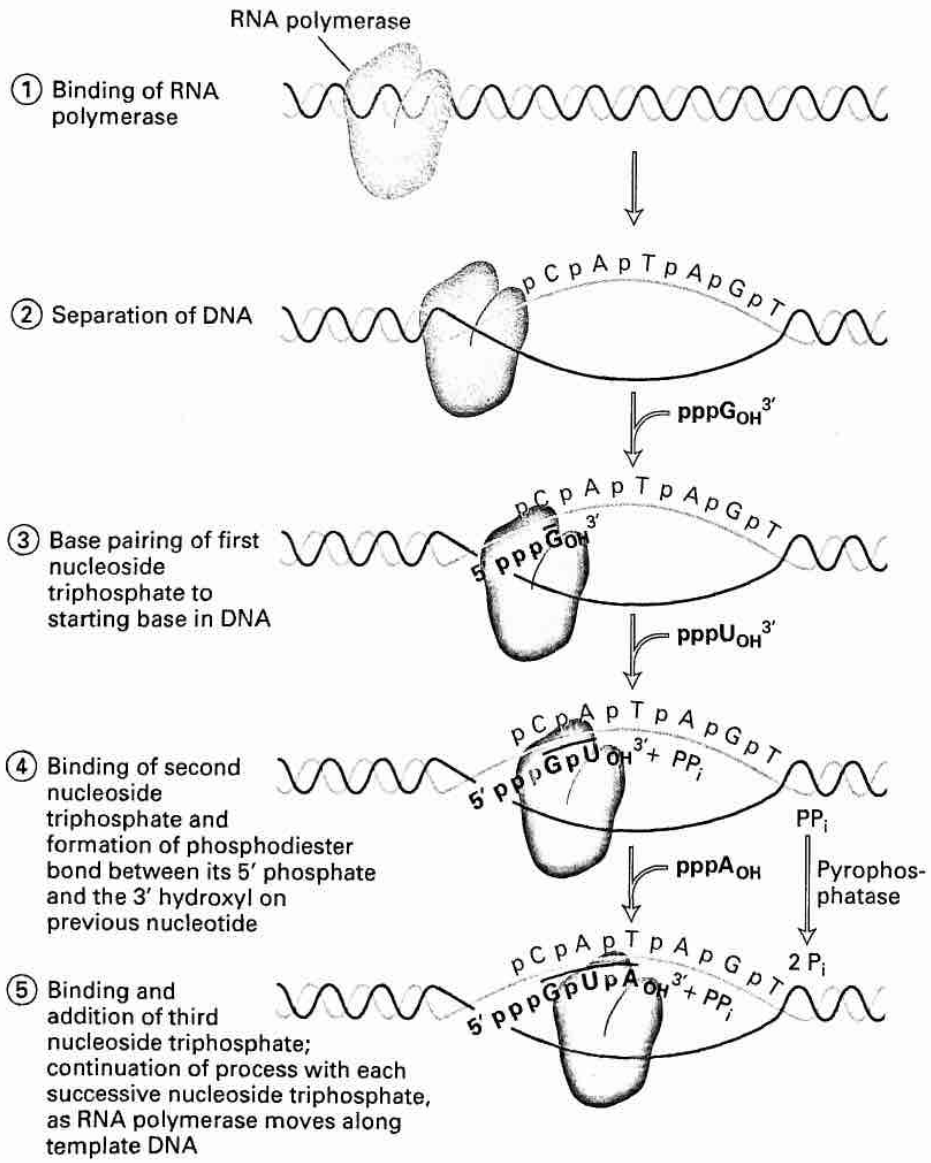
All RNA and DNA synthesis, both cellular and viral, proceeds in the same chemical direction: from the 5′ (phosphate) end to the 3′ (hydroxyl) end (see Figure 4-13). Nucleic acid chains are assembled from 5′ triphosphates of ribonucleosides or deoxyribonucleosides. Strand growth is energetically unfavorable but is driven by the energy available in the triphosphates. The α phosphate of the incoming nucleotide attaches to the 3′ hydroxyl of the ribose (or deoxyribose) of the preceding residue to form a phosphodiester bond, releasing a pyrophosphate (PP_i). The equilibrium of the reaction is driven further toward chain elongation by pyrophosphatase, which catalyzes the cleavage of PP_i into two molecules of inorganic phosphate (see Table 2-7).

RNA Polymerases Can Initiate Strand Growth but DNA Polymerases Cannot

The enzymes that copy (replicate) DNA to make more DNA are **DNA polymerases**; those that copy (transcribe) DNA to form RNA are **RNA polymerases**. Because the two DNA strands are complementary, rather than identical, transcription of a particular DNA segment theoretically could yield two mRNAs with different sequences and hence different protein-coding potentials. Generally, only one strand of the duplex in a particular DNA segment gives rise to usable information when transcribed into mRNA. In unusual cases, though, limited sections of DNA encode proteins on both strands.

An RNA polymerase can find an appropriate initiation site on duplex DNA; bind the DNA; temporarily "melt," or separate, the two strands in that region; and begin generating a new RNA strand (Figure 4-15). As discussed in Chapter 10, the location and regulated use of transcription start sites to produce mRNA requires many dozens of proteins in eukaryotes and several proteins

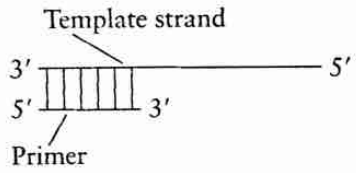
MEDIA CONNECTIONS
Basic Transcriptional Mechanism



◀ FIGURE 4-15 Transcription of DNA into RNA is catalyzed by RNA polymerase, which can initiate the synthesis of strands de novo on DNA templates. The nucleotide at the 5′ end of an RNA strand retains all three of its phosphate groups; all subsequent nucleotides release pyrophosphate (PP_i) when added to the chain and retain only their α phosphate (red). The released PP_i is subsequently hydrolyzed by pyrophosphatase to P_i, driving the equilibrium of the overall reaction toward chain elongation. In most cases, only one DNA strand is transcribed into RNA.

even in bacteria. The nucleotide at the 5′ terminus of a growing RNA strand is chemically distinct from the nucleotides within the strand in that it retains all three phosphate groups. When an additional nucleotide is added to the 3′ end of the growing strand, only the α-phosphate is retained; the β and γ phosphates are lost as pyrophosphate, which is subsequently hydrolyzed to yield 2 molecules of inorganic phosphate.

Unlike RNA polymerases, DNA polymerases cannot initiate chain synthesis de novo; instead, they require a short, preexisting RNA or DNA strand, called a primer, to begin chain growth. With a primer base-paired to the template strand, a DNA polymerase adds nucleotides to the free hydroxyl group at the 3′ end of the primer:



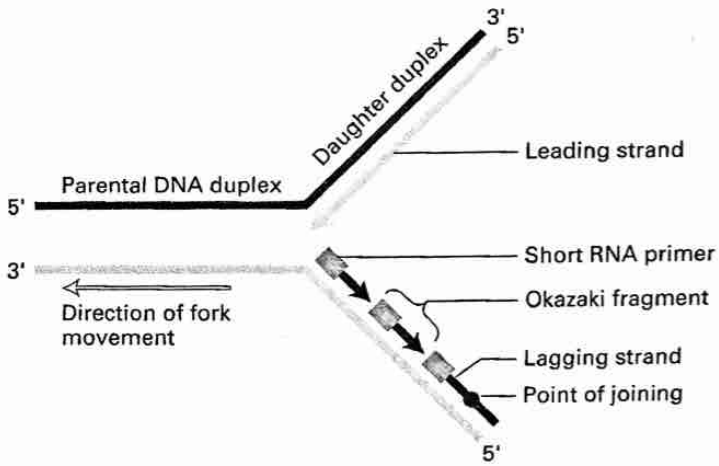
If RNA is the primer, the polynucleotide copied from the template is RNA at the 5′ end and DNA at the 3′ end.

Both prokaryotic and eukaryotic cells have several different types of DNA polymerases. Some polymerases participate in making new DNA to prepare for cell division; other polymerases serve in the repair and recombination of DNA molecules. The structure, mechanism, and physiological role of these enzymes are described in Chapter 12.

Replication of Duplex DNA Requires Assembly of Many Proteins at a Growing Fork

Because duplex DNA consists of two intertwined strands, the base-pair copying of each strand requires unwinding of the original duplex, which is accomplished by specific "unwinding proteins" called **helicases.** As noted earlier, local unwinding of duplex DNA produces torsional stress, leading to formation of supercoils, which are removed by topoisomerases. The action of all these proteins produces a moving, highly specialized region of the DNA called the **growing fork,** at which DNA polymerase carries out nucleotide addition. In order for DNA polymerase to move along and copy a duplex DNA, helicase must sequentially unwind the duplex and topoisomerase must remove the supercoils that form.

DNA replication begins with creation of a growing fork by a protein or proteins that have helicase activity and unwind a short section of parental DNA. A specialized RNA polymerase then forms short RNA primers complementary to the unwound template strands. Each such primer, still bound to its complementary DNA strand, is then elongated by DNA polymerase, thereby forming a new daughter strand. One final major complication in the operation of a DNA growing fork is that although the two strands of the parental duplex are antiparallel, nucleotides can be added to the growing new strands only in the 5′ → 3′ direction. As diagrammed in Figure 4-16, synthesis of one daughter strand, called the **leading strand,** proceeds continuously from a single RNA primer in the 5′ → 3′ direction, *the same direction as movement of the growing fork.* Because growth of the other daughter strand, called the **lagging strand,** also must occur in the 5′ → 3′ direction, copying of its template strand must somehow occur in the *opposite* direction from the movement of the growing fork. A cell accomplishes this feat by producing additional short RNA primers every 1000 bases or so on the second parental strand, as more of the strand is exposed by unwinding. Each of these primers, base-paired to their template strand, is elongated in the 5′ → 3′ direction, forming discontinuous segments called **Okazaki fragments** after their discoverer Reiji Okazaki. The RNA primer of each Okazaki fragment is removed and replaced by DNA chain growth from the neighboring Okazaki fragment; finally an enzyme called *DNA ligase* joins the adjacent fragments. At least 30 proteins participate in the formation and operation of a DNA growing fork; this DNA-replication machine is discussed in detail in Chapter 12.



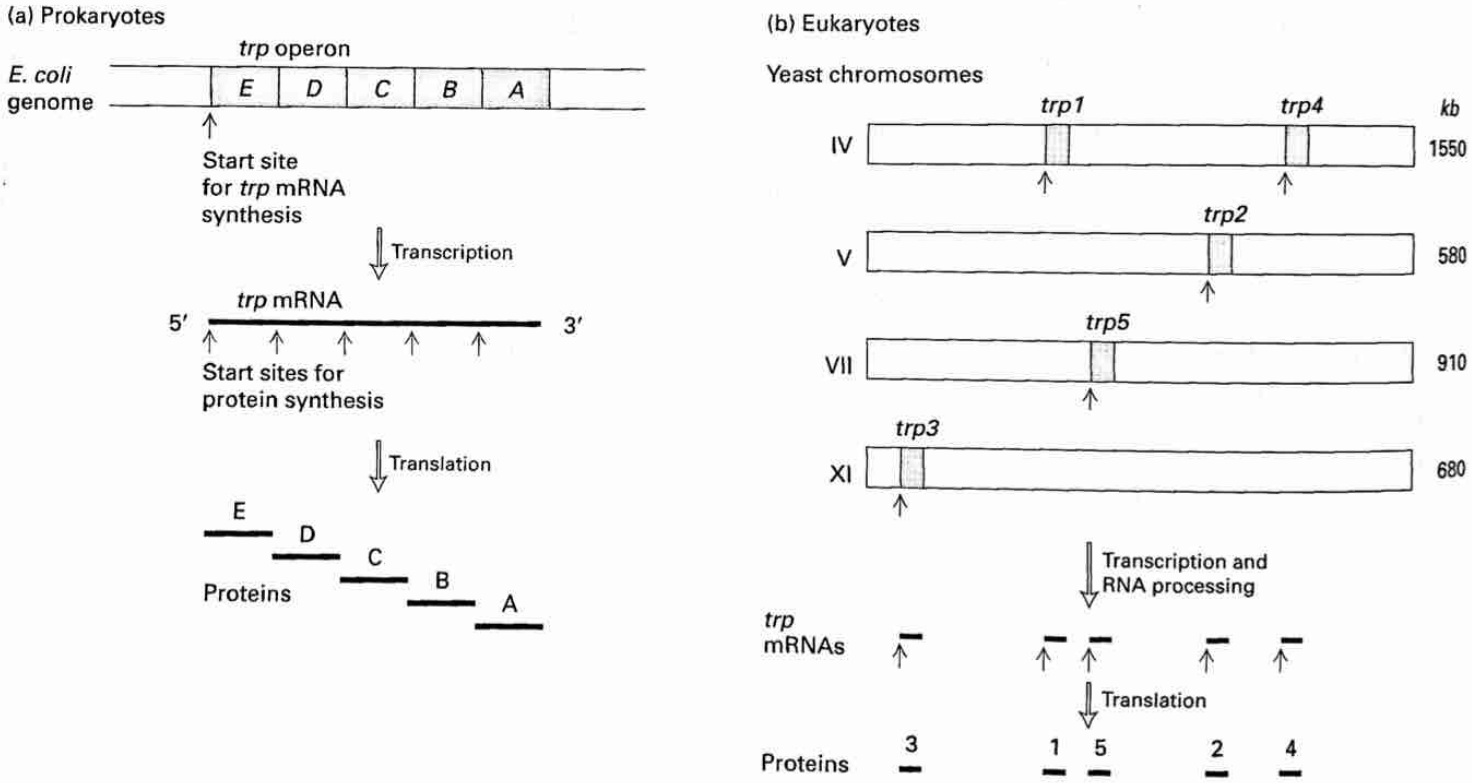
▲ FIGURE 4-16 Schematic diagram of DNA replication at a growing fork. Nucleotides are added by DNA polymerase to each daughter strand in the 5′ → 3′ direction (indicated by arrowheads). Synthesis of the leading strand occurs continuously from a single RNA primer at its 5′ end (not shown). Synthesis of the other new strand—the lagging strand—proceeds discontinuously, initially forming Okazaki fragments, from multiple RNA primers that are formed on the parental strand as each new region of DNA is exposed at the growing fork. The RNA primers are elongated by DNA polymerase. As each growing fragment approaches the previous primer, that primer is removed by another enzyme and the fragments are joined by DNA ligase to form a continuous DNA strand. By repetition of this process, the entire lagging strand eventually is completed.

Organization of Genes in DNA Differs in Prokaryotes and Eukaryotes

Having outlined the principles governing the stepwise assembly of polynucleotides, we now focus briefly on the large-scale arrangement of information in DNA and how this arrangement dictates the requirements for RNA manufacture so that information transfer goes smoothly. The simplest definition of a gene is a "unit of DNA that contains the information to specify synthesis of a single polypeptide chain." The number of genes in cells varies widely, with the simpler non-nucleated prokaryotic cells having far fewer genes than eukaryotic cells. The vast majority of genes carry information to build protein molecules, and it is the RNA copies of such *protein-coding genes* that are the mRNA molecules of cells. In recent years, the entire sequence of the DNA genome of several organisms has been determined, providing direct evidence for large differences in their protein-coding capacity (Chapter 7).

The most common arrangement of protein-coding genes in all prokaryotes has a powerful and appealing logic: genes devoted to a single metabolic goal, say, the synthesis of the amino acid tryptophan, are most often found in a contiguous array in the DNA. This gene order makes it possible to produce a continuous strand of mRNA that carries the message for a related series of enzymes devoted to making tryptophan (Figure 4-17a). Each section of the mRNA represents the unit (or gene) that instructs the protein-synthesizing apparatus to make a particular protein. Such an arrangement of genes in a functional group is called an **operon,** because it operates as a unit from a single transcription start site. In prokaryotic DNA the genes are closely packed with very few noncoding gaps, and the DNA is transcribed directly into colinear mRNA, which then is translated into protein, even while stretches of the mRNA closer to the 3′ end are still being produced.

This economic clustering of genes devoted to a single metabolic function does not occur in eukaryotes, even simple ones like yeasts that can be metabolically similar to bacteria. Rather, eukaryotic genes, even those devoted to a single pathway, are most often physically separated in the DNA, sometimes even being located on different chromosomes.

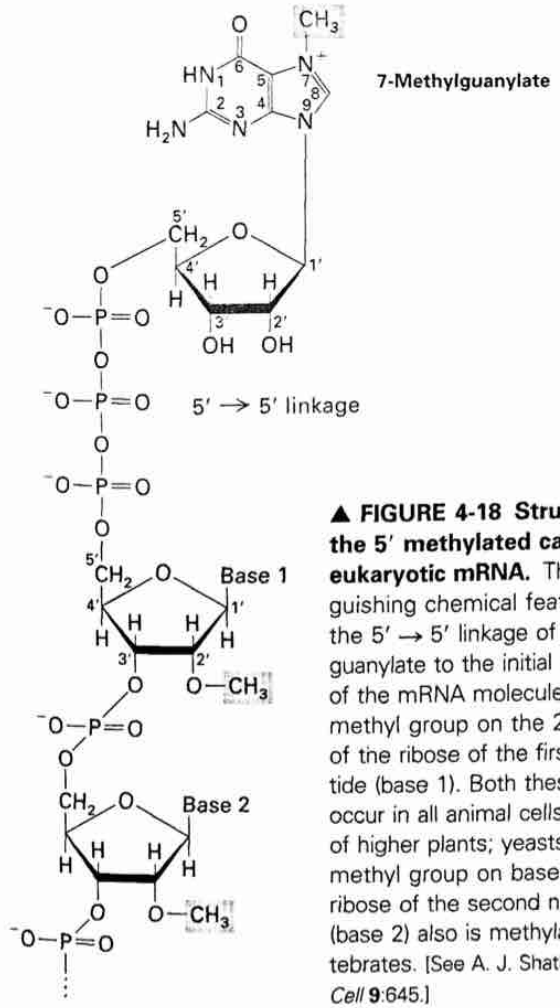


▲ **FIGURE 4-17 Comparison of gene organization, transcription, and translation in prokaryotes and eukaryotes.** (a) The tryptophan *(trp)* operon is a continuous segment of the *E. coli* chromosome, containing five genes (blue) that encode the enzymes necessary for the stepwise synthesis of tryptophan. The entire operon is transcribed from one start site (blue arrow) into one long continuous *trp* mRNA (red). Translation of this mRNA begins at five different start sites, yielding five proteins (green). Proteins E and D associate to form the first enzyme in the tryptophan biosynthetic pathway; protein C catalyzes the intermediate step; and proteins A and B form tryptophan synthetase, the final enzyme. Thus the order of the genes in the bacterial genome parallels the sequential function of the encoded proteins in the tryptophan pathway. (b) The five genes encoding the enzymes required for tryptophan synthesis in yeast *(Saccharomyces cerevisiae)* are carried on four different chromosomes. Each gene is transcribed from its own start site to yield a primary transcript that is processed into a functional mRNA encoding a single protein (see Figure 4-19). The length of the yeast chromosomes is given in kilobases (10^3 bases), with all drawn to the same length.

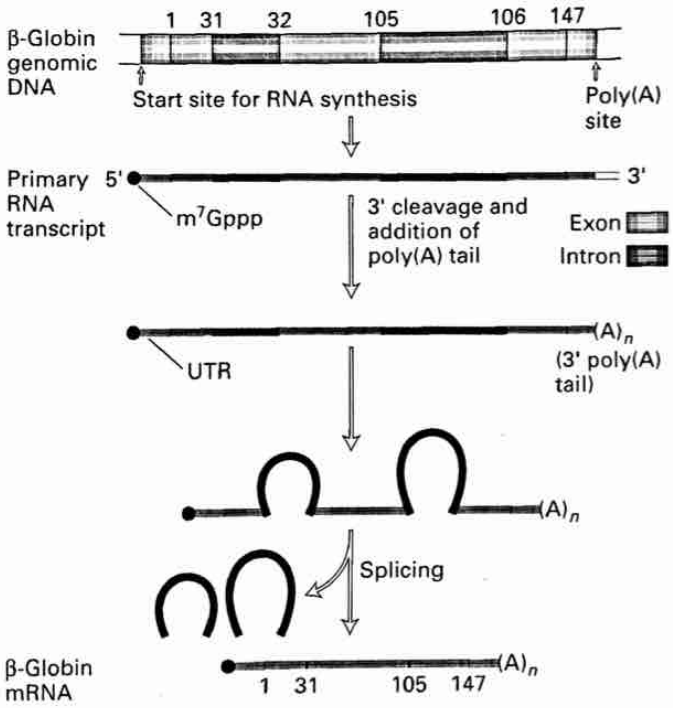
Each gene is transcribed from its own start site, producing one mRNA, which generally is translated to yield a single protein (Figure 4-17b). Moreover, when researchers first compared the nucleotide sequences of eukaryotic mRNAs with the DNAs encoding them, they were astounded to find that the uninterrupted protein-coding sequence of a given mRNA was broken up (discontinuous) in its corresponding section of DNA. They concluded that the eukaryotic gene existed in pieces of coding sequence, the *exons*, separated by non-protein-coding segments, the *introns*. This astonishing finding, first discovered in viruses that infect eukaryotic cells, implied that the long initial RNA copy, called the **primary transcript**, the entire copied DNA sequence, had to be clipped apart to remove the introns and then carefully stitched back together to produce many mRNAs of eukaryotic cells.

Eukaryotic Primary RNA Transcripts Are Processed to Form Functional mRNAs

In prokaryotic cells, which have no nuclei, translation of an mRNA into protein can begin from the 5′ end of the mRNA even while the 3′ end is still being copied from DNA. Thus, transcription and translation can occur concurrently.



▲ FIGURE 4-18 Structure of the 5′ methylated cap of eukaryotic mRNA. The distinguishing chemical features are the 5′ → 5′ linkage of 7-methylguanylate to the initial nucleotide of the mRNA molecule and the methyl group on the 2′ hydroxyl of the ribose of the first nucleotide (base 1). Both these features occur in all animal cells and in cells of higher plants; yeasts lack the methyl group on base 1. The ribose of the second nucleotide (base 2) also is methylated in vertebrates. [See A. J. Shatkin, 1976, *Cell* **9**:645.]



▲ FIGURE 4-19 Overview of RNA processing in eukaryotes using β-globin gene as an example. The β-globin gene contains three protein-coding exons (red) and two intervening noncoding introns (blue). The introns interrupt the protein-coding sequence between the codons for amino acids 31 and 32 and 105 and 106. Transcription of this and many other genes starts slightly upstream of the 5′ exon and extends downstream of the 3′ exon, resulting in noncoding regions (gray) at the ends of the primary transcript. These regions, referred to as *untranslated regions (UTRs)*, are retained during processing. The 5′ 7-methylguanylate cap (m^7Gppp; green dot) is added during formation of the primary RNA transcript, which extends beyond the poly(A) site. After cleavage at the poly(A) site and addition of multiple A residues to the 3′ end, splicing removes the introns and joins the exons. The small numbers refer to positions in the 147-aa sequence of β-globin.

MEDIA CONNECTIONS
Life Cycle of an mRNA

In eukaryotic cells, however, not only is the nucleus separated from the cytoplasm where protein synthesis occurs, but the primary RNA transcript of a protein-coding gene must undergo several modifications, collectively termed **RNA processing**, that yield a functional mRNA. This mRNA then must be transported to the cytoplasm before it can be translated into protein. Thus, transcription and translation cannot occur concurrently in eukaryotic cells.

The initial steps in processing of all eukaryotic primary RNA transcripts occur at the two ends, and these modifications are retained in mRNAs. To the initiating (5′) nucleotide of the primary transcript is added the *5′ cap*, which may serve to protect mRNA from enzymatic degradation (Figure 4-18). This modification occurs before transcription is complete, so the 5′ cap is present in the primary transcript. Processing at the 3′ end of the primary transcript involves cleavage by an endonuclease to yield a free 3′-hydroxyl

group to which a string of adenylic acid residues is added by an enzyme called *poly(A) polymerase*. The resulting poly(A) tail contains 100–250 bases, being shorter in yeasts and invertebrates than in vertebrates. Poly(A) polymerase is part of a complex of proteins that adds the poly(A) tail. This complex does not require a template and can determine the correct number of A residues to add in each species.

The final step in the processing of many different eukaryotic mRNA molecules is **splicing:** the internal cleavage of the RNA transcript to excise the introns, followed by ligation of the coding exons. Many eukaryotic mRNAs also contain noncoding regions at each end; these are referred to as the 5′ and 3′ *untranslated regions* (UTRs). Figure 4-19 summarizes the basic steps in RNA processing. We examine the cellular machinery for carrying out processing of mRNA, as well as tRNA and rRNA, in Chapter 11.

SUMMARY Nucleic Acid Synthesis

- The transfer of information from genes to proteins is assisted by proteins that participate in the synthesis of DNA and RNA.
- Polynucleotide and polypeptide chains are assembled from a limited number of monomeric units that are added one at a time, beginning at the 5′ end in nucleic acids and the amino-terminal end in proteins (see Figure 4-13). In both cases, the initial polymeric product generally is modified in some fashion to produce a functional molecule.
- A polynucleotide chain is synthesized by copying of a complementary template strand (usually DNA). In this process, the duplex DNA is locally unwound, revealing the unpaired template strand, and nucleotides are added to the 3′-hydroxyl end of the growing strand by RNA or DNA polymerase.
- RNA polymerase can initiate transcription of DNA into RNA by binding to a specific start site and unwinding the duplex. As the enzyme moves along the DNA, it unwinds sequential segments of the DNA and adds nucleotides to the growing RNA strand (see Figure 4-15). Most commonly, only one DNA strand in any one locus is transcribed into RNA.
- Replication of DNA requires the assistance of helicase to unwind the duplex, topoisomerase to remove supercoils, and a specialized RNA polymerase to form RNA primers because DNA polymerase cannot start chains. Nucleotide addition at the growing fork, a moving region of strand separation produced by sequential unwinding of the duplex, is catalyzed by one type of DNA polymerase.
- During DNA replication, the two new daughter strands are assembled somewhat differently because DNA polymerase can add nucleotides only in the 5′ → 3′ direction (see Figure 4-16). One new strand, the leading strand, is elongated continuously from a single primer. The other new strand, the lagging strand, is synthesized discontinuously as a series of short segments, called Okazaki fragments, initiated from multiple RNA primers. After removal of the intervening primer, adjacent Okazaki fragments are joined by DNA ligase.
- In prokaryotic DNA, related protein-coding genes are clustered into a functional region, an operon, which is transcribed from a single start site into one mRNA encoding multiple proteins (see Figure 4-17a). Translation of a mRNA can begin before synthesis of the mRNA is complete.
- In eukaryotic DNA, each protein-coding gene is transcribed from its own start site, and very often the coding regions (exons) are separated by noncoding regions (introns). The primary RNA transcript produced from such a gene must undergo processing to yield a functional mRNA. During processing, the ends of all primary transcripts are modified by addition of a 5′ cap and 3′ poly(A) tail; many transcripts also undergo splicing—removal of the introns and joining of the exons (see Figure 4-19).

4.4 The Three Roles of RNA in Protein Synthesis

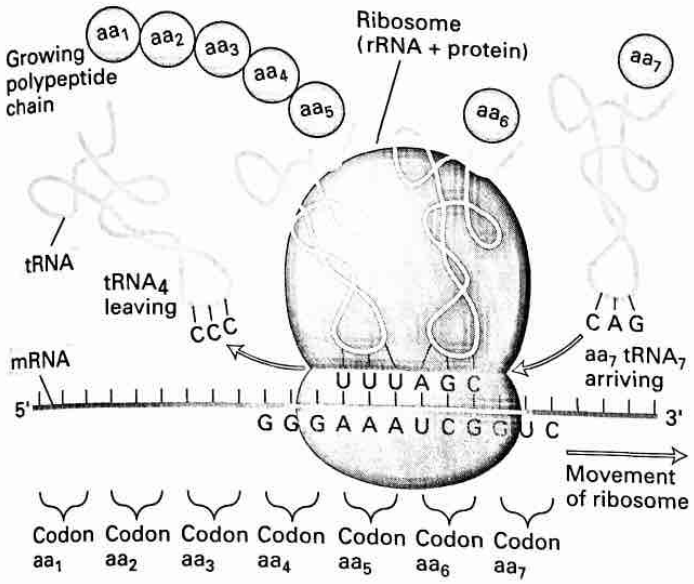
Although DNA stores the information for protein synthesis and RNA carries out the instructions encoded in DNA, most biological activities are carried out by proteins. The accurate synthesis of proteins thus is critical to the proper functioning of cells and organisms. We saw in Chapter 3 that the linear order of amino acids in each protein determines its three-dimensional structure and activity. For this reason, assembly of amino acids in their correct order, as encoded in DNA, is the key to production of functional proteins.

Three kinds of RNA molecules perform different but cooperative functions in protein synthesis (Figure 4-20):

1. Messenger RNA (mRNA) carries the genetic information copied from DNA in the form of a series of three-base code "words," each of which specifies a particular amino acid.

2. Transfer RNA (tRNA) is the key to deciphering the code words in mRNA. Each type of amino acid has its own type of tRNA, which binds it and carries it to the growing end of a polypeptide chain if the next code word on mRNA calls for it. The correct tRNA with its attached amino acid is selected at each step because each specific tRNA molecule contains a three-base sequence that can base-pair with its complementary code word in the mRNA.

3. Ribosomal RNA (rRNA) associates with a set of proteins to form **ribosomes.** These complex structures, which physically move along an mRNA molecule, catalyze the assembly



▲ FIGURE 4-20 The three roles of RNA in protein synthesis. Messenger RNA (mRNA) is translated into protein by the joint action of transfer RNA (tRNA) and the ribosome, which is composed of numerous proteins and two major ribosomal RNA (rRNA) molecules. [Adapted from A. J. F. Griffiths et al., 1993, *An Introduction to Genetics Analysis,* 5th ed., W. H. Freeman.]

of amino acids into protein chains. They also bind tRNAs and various accessory molecules necessary for protein synthesis. Ribosomes are composed of a large and small subunit, each of which contains its own rRNA molecule or molecules.

Translation is the whole process by which the base sequence of an mRNA is used to order and to join the amino acids in a protein. The three types of RNA participate in this essential protein-synthesizing pathway in all cells; in fact, the development of the three distinct functions of RNA was probably the molecular key to the origin of life. How each RNA carries out its specific task is discussed in this section, while the biochemical events in protein synthesis and the required protein factors are described in the final section of the chapter.

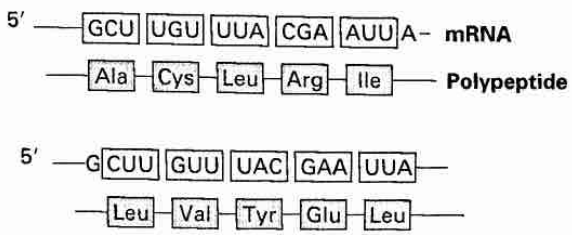
Messenger RNA Carries Information from DNA in a Three-Letter Genetic Code

RNA contains ribonucleotides of adenine, cytidine, guanine, and uracil; DNA contains deoxyribonucleotides of adenine, cytidine, guanine, and thymine. Because 4 nucleotides, taken individually, could represent only 4 of the 20 possible amino acids in coding the linear arrangement in proteins, a *group* of nucleotides is required to represent each amino acid. The code employed must be capable of specifying at least 20 words (i.e., amino acids).

If two nucleotides were used to code for one amino acid, then only 16 (or 4^2) different code words could be formed, which would be an insufficient number. However, if a group of three nucleotides is used for each code word, then 64 (or 4^3) code words can be formed. Any code using groups of three or more nucleotides will have more than enough units to encode 20 amino acids. Many such coding systems are mathematically possible. However, the actual **genetic code** used by cells is a *triplet* code, with every three nucleotides being "read" from a specified starting point in the mRNA. Each triplet is called a **codon.** Of the 64 possible codons in the genetic code, 61 specify individual amino acids and three are stop codons. Table 4-2 shows that most amino acids are encoded by more than one codon. Only two—methionine and tryptophan—have a single codon; at the other extreme, leucine, serine, and arginine are each specified by six different codons. The different codons for a given amino acid are said to be *synonymous.* The code itself is termed *degenerate,* which means that it contains redundancies.

Synthesis of all protein chains in prokaryotic and eukaryotic cells begins with the amino acid methionine. In most mRNAs, the start (initiator) codon specifying this amino-terminal methionine is AUG. In a few bacterial mRNAs, GUG is used as the initiator codon, and CUG occasionally is used as an initiator codon for methionine in eukaryotes. The three codons UAA, UGA, and UAG do not specify amino acids but constitute stop (*terminator*) signals that mark the carboxyl terminus of protein chains in almost all cells. The sequence of codons that runs from a specific start site to a terminating codon is called a **reading frame.** This precise linear array of ribonucleotides in groups of three in mRNA specifies the precise linear sequence of amino acids in a protein and also signals where synthesis of the protein chain starts and stops.

Because the genetic code is a commaless, overlapping triplet code, a particular mRNA theoretically could be translated in three different reading frames. Indeed some mRNAs have been shown to contain overlapping information that can be translated in different reading frames, yielding different polypeptides (Figure 4-21). The vast majority of



▲ FIGURE 4-21 Example of how the genetic code—an overlapping, commaless triplet code—can be read in two different frames. If translation of the mRNA sequence shown begins at two different upstream start sites (not shown), then two overlapping reading frames are possible; in this case, the codons are shifted one base to the right in the lower frame. As a result, different amino acids are encoded by the same nucleotide sequence. Many instances of such overlaps have been discovered in viral and cellular genes of prokaryotes and eukaryotes. It is theoretically possible for the mRNA to have a third reading frame.

TABLE 4-2 The Genetic Code (RNA to Amino Acids)*

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

**"Stop (och)" stands for the ochre termination triplet, and "Stop (amb)" for the amber, named after the bacterial strains in which they were identified. AUG is the most common initiator codon; GUG usually codes for valine, and CUG for leucine, but, rarely, these codons can also code for methionine to initiate an mRNA chain.*

mRNAs, however, can be read in only one frame because stop codons encountered in the other two possible reading frames terminate translation before a functional protein is produced. Another unusual coding arrangement occurs because of *frameshifting*. In this case the protein-synthesizing machinery may read four nucleotides as one amino acid and then continue reading triplets, or it may back up one base and read all succeeding triplets in the new frame until termination of the chain occurs. These frameshifts are not common events, but a few dozen such instances are known.

The meaning of each codon is the same in most known organisms—a strong argument that life on earth evolved only once. Recently the genetic code has been found to differ for a few codons in many mitochondria, in ciliated protozoans, and in *Acetabularia*, a single-celled plant. As shown in Table 4-3, most of these changes involve reading of normal stop codons as amino acids, not an exchange of one

TABLE 4-3 Unusual Codon Usage in Nuclear and Mitochondrial Genes

Codon	Universal Code	Unusual Code	Occurrence*
UGA	Stop	Trp	*Mycoplasma, Spiroplasma,* mitochondria of many species
CUG	Leu	Thr	Mitochondria in yeasts
UAA, UAG	Stop	Gln	*Acetabularia, Tetrahymena,* Paramecium, etc.
UGA	Stop	Cys	*Euplotes*

**"Unusual code" is used in nuclear genes of the listed organisms and in mitochondrial genes as indicated.*

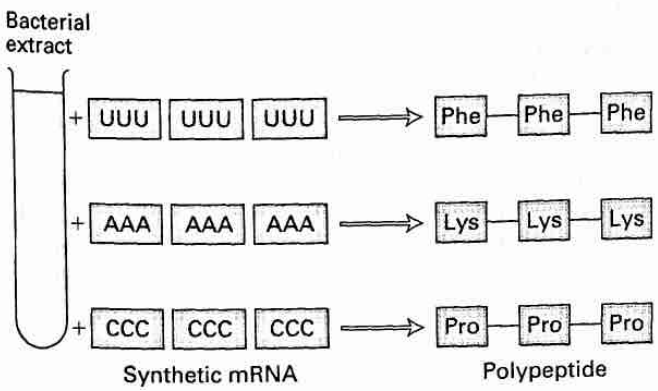
SOURCE: S. Osawa et al., 1992, *Microbiol. Rev.* 56:229.

amino acid for another. It is now thought that these exceptions to the general code are later evolutionary developments; that is, at no single time was the code immutably fixed, although massive changes were not tolerated once a general code began to function early in evolution.

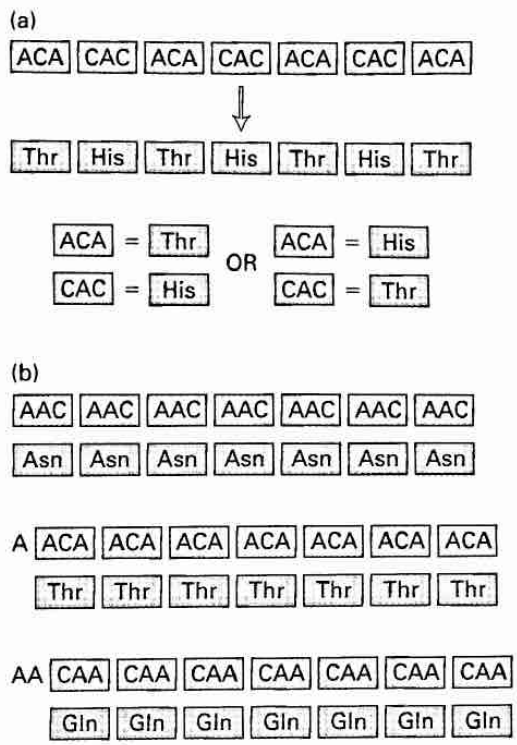
Experiments with Synthetic mRNAs and Trinucleotides Broke the Genetic Code

Having described the genetic code, we briefly recount how it was deciphered—one of the great triumphs of modern biochemistry. The underlying experimental work was carried out largely with cell-free bacterial extracts containing all the necessary components for protein synthesis except mRNA (i.e., tRNAs, ribosomes, amino acids, and the energy-rich nucleotides ATP and GTP).

Initially, researchers added synthetic mRNAs containing a single type of nucleotide to such extracts and then determined the amino acid incorporated into the polypeptide that was formed. In the first successful experiment, synthetic mRNA composed only of U residues [poly(U)] yielded polypeptides made up only of phenylalanine. Thus it was concluded that a codon for phenylalanine consisted entirely of U's. Likewise, experiments with poly(C) and poly(A) showed that a codon for proline contained only C's and a codon for lysine only A's (Figure 4-22). [Poly(G) did not work in this type of experiment because it assumes an unusable stacked structure that is not translated well.] Next, synthetic mRNAs composed of alternating bases were used. The results of these experiments not only revealed more codons but also demonstrated that codons are three bases long. The example of this approach illustrated in Figure 4-23 led to identification of ACA as the codon for threonine and CAC for histidine. Similar experiments with many such mixed polynucleotides revealed a substantial part of the genetic code.



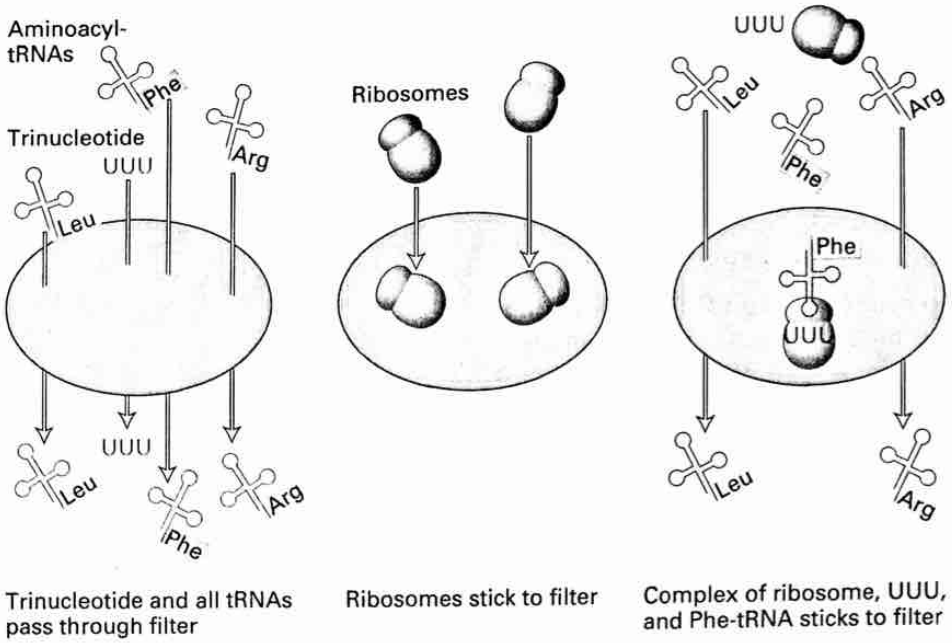
▲ FIGURE 4-22 Assigning codons using synthetic mRNAs containing a single ribonucleotide. Addition of such a synthetic mRNA to a bacterial extract that contained all the components necessary for protein synthesis except mRNA resulted in synthesis of polypeptides composed of a single type of amino acid as indicated. [See M. W. Nirenberg and J. H. Matthei, 1961, *Proc. Nat'l. Acad. Sci. USA* **47**:1588.]



▲ FIGURE 4-23 Assigning codons using mixed polynucleotides. (a) When a synthetic mRNA with alternating A and C residues was added to a protein-synthesizing bacterial extract, the resulting polypeptide contained alternating threonine and histidine residues. This finding is compatible with the two alternative codon assignments shown. (Note that alternating residues yield the same sequence of triplets regardless of which reading frame is chosen.) (b) To determine which codon assignment shown in (a) is correct, a second mRNA consisting of AAC repeats was tested. This mRNA, which can be read in three frames, yielded the three types of polypeptides shown. Since only the ACA codon was common to both experiments, it must encode threonine; thus CAC must encode histidine in (a). The assignments AAC = asparagine (Asn) and CAA = glutamine (Gln) were derived from additional experiments. [See H. G. Korana, 1968, reprinted in *Nobel Lectures: Physiology or Medicine (1963–1970)*, Elsevier (1973), p. 341.]

The entire genetic code was finally worked out by a second type of experiment conducted by Marshall Nirenberg and his collaborators. In this approach, all the possible trinucleotides were tested for their ability to attract tRNAs attached to the 20 different amino acids found in natural proteins (Figure 4-24). In all, 61 of the 64 possible trinucleotides were found to code for a specific amino acid; the trinucleotides UAA, UGA, and UAG did not encode amino acids.

Although synthetic mRNAs were useful in deciphering the genetic code, in vitro protein synthesis from these mRNAs is very inefficient and yields polypeptides of variable size. Successful in vitro synthesis of a naturally occurring protein was achieved first when mRNA from bacteriophage F2 (a virus) was added to bacterial extracts, leading to formation of the coat, or capsid, protein (the "packaging" protein that covers the virus particle). Studies with such



▲ **FIGURE 4-24 Breaking the entire genetic code by use of chemically synthesized trinucleotides.** Marshall Nirenberg and his collaborators prepared 20 ribosome-free bacterial extracts containing all possible aminoacyl-tRNAs (tRNAs with an amino acid attached). In each sample, a different amino acid was radioactively labeled (green); the other 19 amino acids were bound to tRNAs but were unlabeled. Aminoacyl-tRNAs and trinucleotides passed through a nitrocellulose filter *(left)*, but ribosomes were retained by the filter *(center)* and would bind trinucleotides and their cognate tRNAs *(right)*. Each possible trinucleotide was tested separately for its ability to attract a specific tRNA by adding it with ribosomes to samples from each of the 20 aminoacyl-tRNA mixtures. The sample was then filtered. If the added trinucleotide caused the radiolabeled aminoacyl-tRNA to bind to the ribosome, then radioactivity would be detected on the filter (a positive test); otherwise, the label would pass through the filter (a negative test). By synthesizing and testing all possible trinucleotides, the researchers were able to match all 20 amino acids with one or more codons (e.g., phenylalanine with UUU as shown here). [See M. W. Nirenberg and P. Leder, 1964, *Science* **145**:1399.]

natural mRNAs established that AUG encodes methionine at the start of almost all proteins and is required for efficient initiation of protein synthesis, while the three trinucleotides (UAA, UGA, and UAG) that do not encode any amino acid act as stop codons, necessary for precise termination of synthesis.

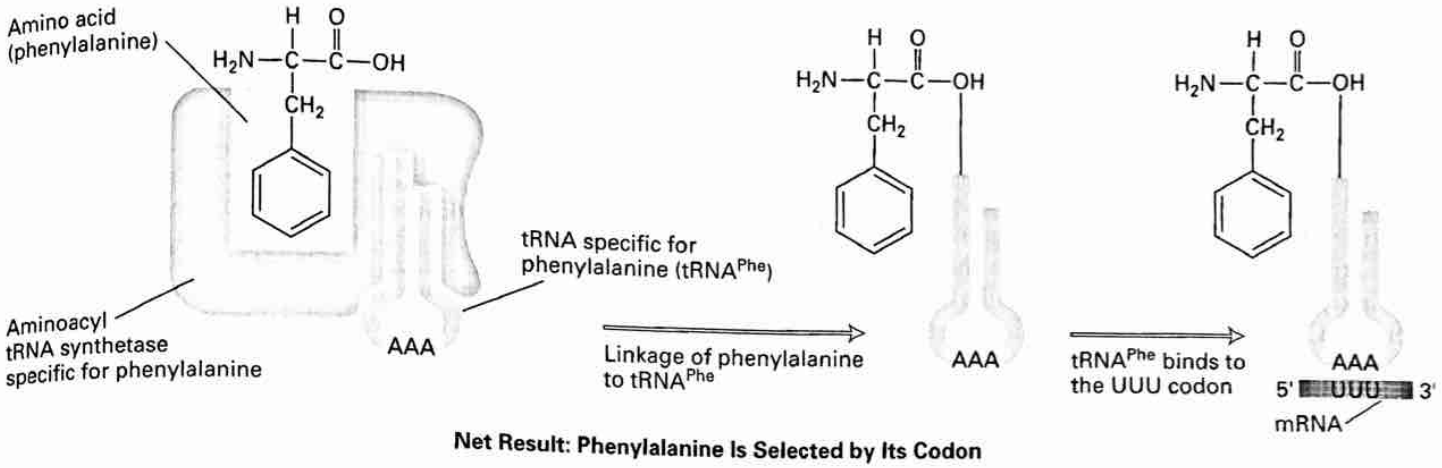
The Folded Structure of tRNA Promotes Its Decoding Functions

The next step in understanding the flow of genetic information from DNA to protein was to determine how the nucleotide sequence of mRNA is converted into the amino acid sequence of protein. This decoding process requires two types of adapter molecules: tRNAs and enzymes called *aminoacyl-tRNA synthetases*. First we describe the role of tRNAs in decoding mRNA codons, and then examine how synthetases recognize tRNAs.

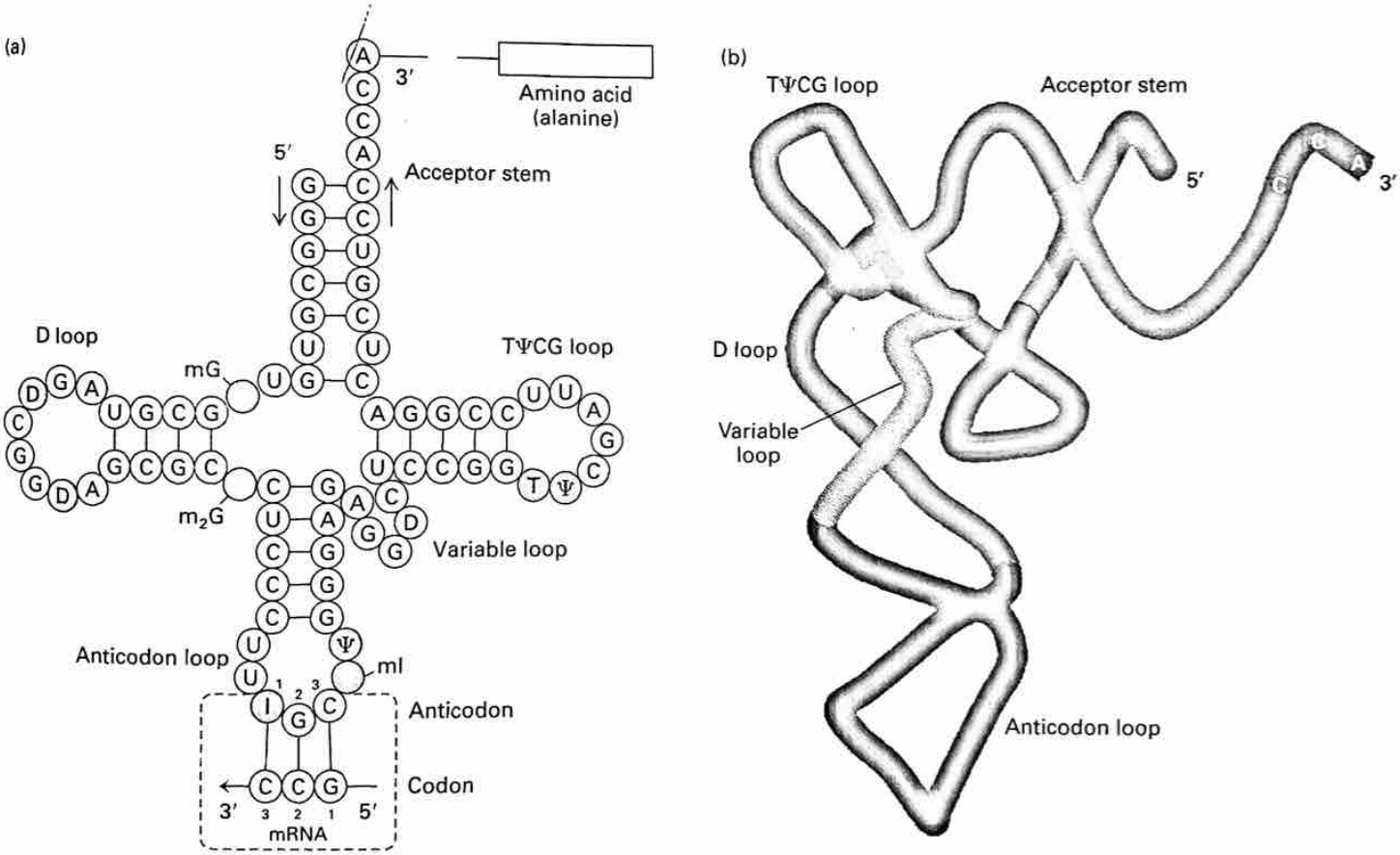
All tRNAs have two functions: to be chemically linked to a particular amino acid and to base-pair with a codon in mRNA so that the amino acid can be added to a growing peptide chain. Each tRNA molecule is recognized by one and only one of the 20 aminoacyl-tRNA synthetases. Likewise, each of these enzymes links one and only one of the 20 amino acids to a particular tRNA, forming an **aminoacyl-tRNA.** Once its correct amino acid is attached, a tRNA then recognizes a codon in mRNA, thereby delivering its amino acid to the growing polypeptide (Figure 4-25).

As studies on tRNA proceeded, 30–40 different tRNAs were identified in bacterial cells and as many as 50–100 in animal and plant cells. Thus the number of tRNAs in most cells is more than the number of amino acids found in proteins (20) and also differs from the number of codons in the genetic code (61). Consequently, many amino acids have more than one tRNA to which they can attach (explaining how there can be more tRNAs than amino acids); in addition, many tRNAs can attach to more than one codon (explaining how there can be more codons than tRNAs). As noted previously, most amino acids are encoded by more than one codon, requiring some tRNAs to recognize more than one codon.

The function of tRNA molecules, which are 70–80 nucleotides long, depends on their precise three-dimensional structures. In solution, all tRNA molecules fold into a similar stem-loop arrangement that resembles a cloverleaf when drawn in two dimensions (Figure 4-26a). The four stems are short double helices stabilized by Watson-Crick base pairing; three of the four stems have loops containing seven or



▲ FIGURE 4-25 Translation of nucleic acid sequences in mRNA into amino acid sequences in proteins requires a two-step decoding process. First, an aminoacyl-tRNA synthetase couples a specific amino acid to its corresponding tRNA. Second, a three-base sequence in the tRNA (the anticodon) base-pairs with a codon in the mRNA specifying the attached amino acid. If an error occurs in either step, the wrong amino acid may be incorporated into a polypeptide chain.



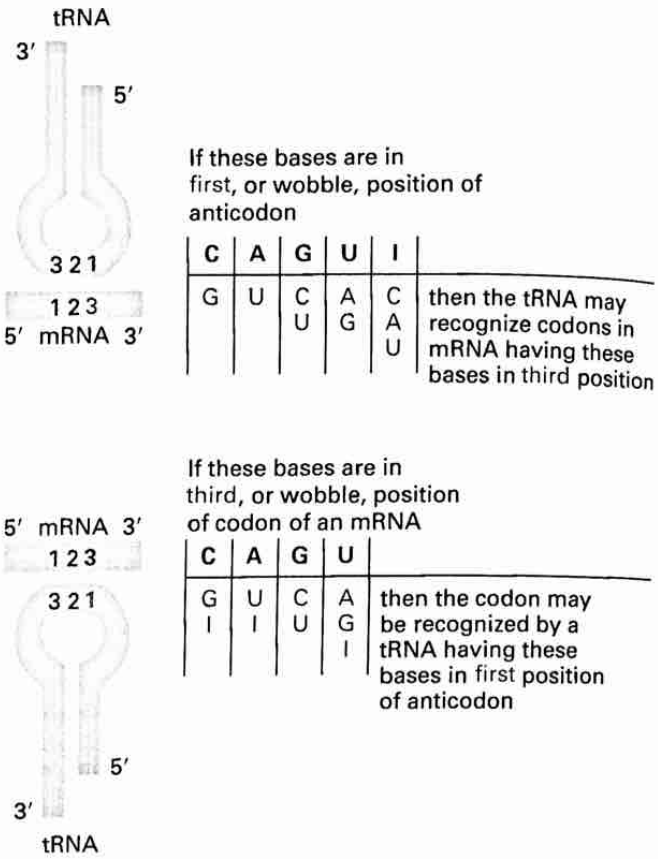
▲ FIGURE 4-26 Structure of tRNAs. (a) The primary structure of yeast alanine tRNA (tRNA^{Ala}), the first such sequence determined. This molecule is synthesized from the nucleotides A, C, G, and U, but some of the nucleotides, shown in red, are modified after synthesis: D = dihydrouridine, I = inosine, T = thymine, Ψ = pseudouridine, and m = methyl group. Although the exact sequence varies among tRNAs, they all fold into four base-paired stems and three loops. The partially unfolded molecule is commonly depicted as a cloverleaf. Dihydrouridine is nearly always present in the D loop; likewise, thymidylate, pseudouridylate, cytidylate, and guanylate are almost always present in the TΨCG loop. The triplet at the tip of the anticodon loop base-pairs with the corresponding codon in mRNA. Attachment of an amino acid to the acceptor arm yields an aminoacyl-tRNA. (b) Computer-generated three-dimensional model of the generalized backbone of all tRNAs. Note the L shape of the molecule. [Part (a) see R. W. Holly et al., 1965, *Science* **147**:1462; part (b) from J. G. Arnez and D. Moras, 1997, *Trends Biochem. Sci.* **22**:211.]

eight bases at their ends, while the remaining, unlooped stem contains the free 3′ and 5′ ends of the chain. Three nucleotides termed the **anticodon,** located at the center of one loop, can form base pairs with the three complementary nucleotides forming a codon in mRNA. As discussed later, specific aminoacyl-tRNA synthetases recognize the surface structure of each tRNA for a specific amino acid and covalently attach the proper amino acid to the unlooped *amino acid acceptor stem.* The 3′ end of all tRNAs has the sequence CCA, which in most cases is added after synthesis and processing of the tRNA are complete. Viewed in three dimensions, the folded tRNA molecule has an L shape with the anticodon loop and acceptor stem forming the ends of the two arms (Figure 4-26b).

Besides addition of CCA at the 3′ terminus after a tRNA molecule is synthesized, several of its nucleic acid bases typically are modified. For example, most tRNAs are synthesized with a four-base sequence of UUCG near the middle of the molecule. The first uridylate is methylated to become a thymidylate; the second is rearranged into a pseudouridylate (abbreviated Ψ), in which the ribose is attached to carbon 5 instead of to nitrogen 1 of the uracil. These modifications produce a characteristic TΨCG loop in an unpaired region at approximately the same position in nearly all tRNAs (see Figure 4-26a).

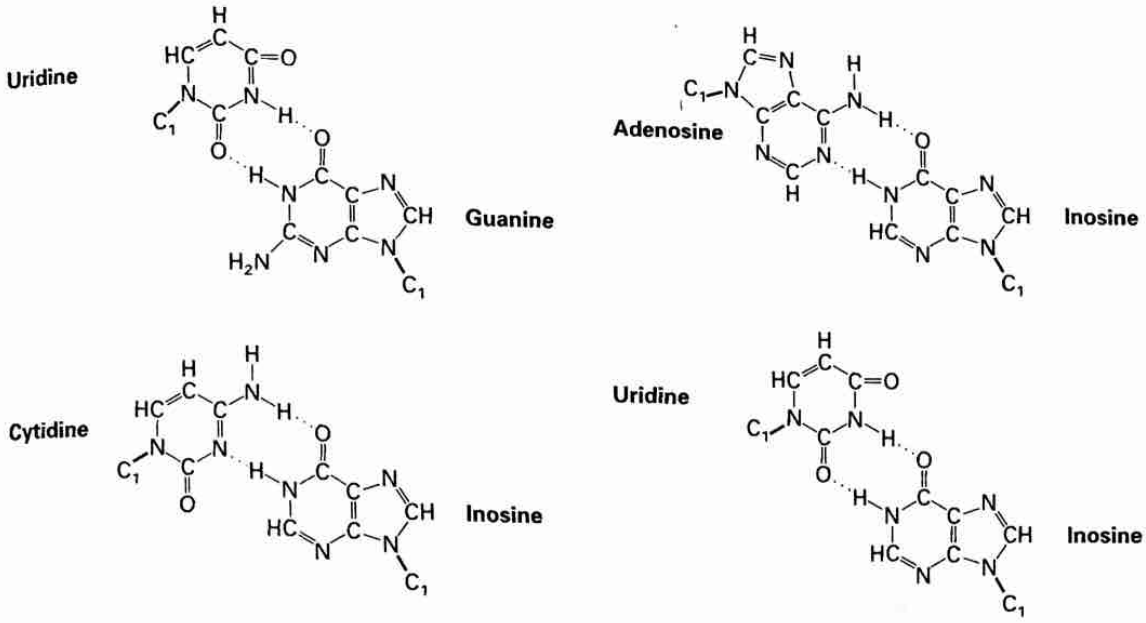
Nonstandard Base Pairing Often Occurs between Codons and Anticodons

If perfect Watson-Crick base pairing were demanded between codons and anticodons, cells would have to contain exactly 61 different tRNA species, one for each codon that specifies an amino acid. As noted above, however, many cells contain fewer than 61 tRNAs. The explanation for the smaller number lies in the capability of a single tRNA anticodon to recognize more than one, but not necessarily every, codon corresponding to a given amino acid. This broader recognition can occur because of nonstandard pairing between bases in the so-called "wobble" position: the third base in a mRNA codon and the corresponding first base in its tRNA anticodon. Although the first and second bases of a codon form standard Watson-Crick base pairs with the third and second bases of the corresponding anticodon, four nonstandard interactions can occur between bases in the wobble position. Particularly important is the G·U base pair, which structurally fits almost as well as the standard G·C pair. Thus, a given anticodon in tRNA with G in the first (wobble) position can base-pair with the two corresponding codons that have either pyrimidine (C or U) in the third position (Figure 4-27). For example, the phenylalanine codons UUU and UUC (5′ → 3′) are both recognized by the tRNA that has GAA (5′ → 3′) as the anticodon. In fact, any two codons of the type NNPyr (N = any base; Pyr = pyrimidine) encode a single amino acid and are decoded by a single tRNA with G in the first (wobble) position of the anticodon.



▲ FIGURE 4-27 The first and second bases in an mRNA codon form Watson-Crick base pairs with the third and second bases, respectively, of a tRNA anticodon. However, the base in the third (or wobble) position of an mRNA codon often forms a nonstandard base pair with the base in the first (or wobble) position of a tRNA anticodon. Wobble pairing allows a tRNA to recognize more than one mRNA codon *(top)*; conversely, it allows a codon to be recognized by more than one kind of tRNA *(bottom),* although each tRNA will bear the same amino acid. Note that a tRNA with I (inosine) in the wobble position can "read" (become paired with) three different codons (see Figure 4-28), and a tRNA with G or U in the wobble position can read two codons. Although A is theoretically possible in the wobble position of the anticodon, it is almost never found in nature.

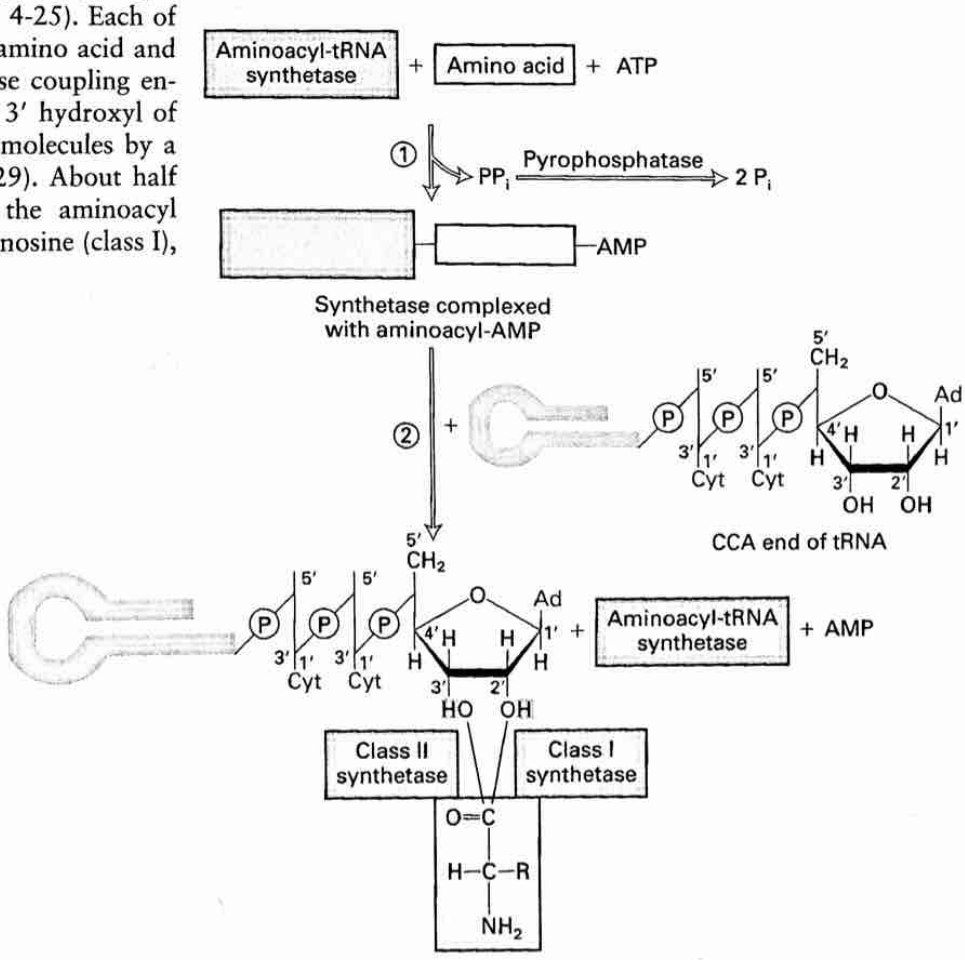
Although adenine rarely is found in the anticodon wobble position, many tRNAs in plants and animals contain *inosine* (I), a deaminated product of adenine, at this position. Inosine can form nonstandard base pairs with A, C, and U (Figure 4-28). A tRNA with inosine in the wobble position thus can recognize the corresponding mRNA codons with A, C, or U in the third (wobble) position (see Figure 4-27). For this reason, inosine-containing tRNAs are heavily employed in translation of the synonymous codons that specify a single amino acid. For example, four of the six codons for leucine have a 3′ A, C, or U (see Table 4-2); these four codons are all recognized by the same tRNA (3′-GAI-5′), which has inosine in the wobble position of the anticodon (and thus recognizes CUA, CUC, and CUU), and uses a G·U pair in position 1 to recognize the UUA codon.



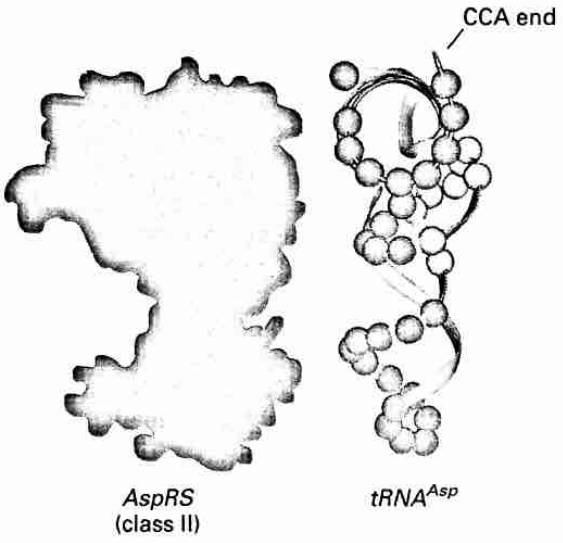
◀ **FIGURE 4-28 The nonstandard, wobble base pairs U · G, C · I, A · I, and U · I.** The heavy black lines indicate the bonds by which the nitrogenous bases attach to the 1′ carbon of ribose (C_1).

Aminoacyl-tRNA Synthetases Activate Amino Acids by Linking Them to tRNAs

Recognition of the codon or codons specifying a given amino acid by a particular tRNA is actually the second step in decoding the genetic message. The first step, attachment of the appropriate amino acid to a tRNA, is catalyzed by a specific aminoacyl-tRNA synthetase (see Figure 4-25). Each of the 20 different synthetases recognizes *one* amino acid and *all* its compatible, or *cognate*, tRNAs. These coupling enzymes link an amino acid to the free 2′ or 3′ hydroxyl of the adenosine at the 3′ terminus of tRNA molecules by a two-step ATP-requiring reaction (Figure 4-29). About half the aminoacyl-tRNA synthetases transfer the aminoacyl group to the 2′ hydroxyl of the terminal adenosine (class I), and about half to the 3′ hydroxyl (class II). In this reaction, the amino acid is linked to the tRNA by a high-energy bond and thus is said to be *activated*. The energy of this bond subsequently drives the formation of peptide bonds between adjacent amino acids in a growing polypeptide chain. The



▶ **FIGURE 4-29 Aminoacylation of tRNA. Amino acids are covalently linked to tRNAs by aminoacyl-tRNA synthetases.** Each of these enzymes recognizes one kind of amino acid and all the cognate tRNAs that recognize codons for that amino acid. The two-step aminoacylation reaction requires energy from the hydrolysis of ATP. The equilibrium of the overall reaction favors the indicated products because the pyrophosphate (PP_i) released in step ① is converted to inorganic phosphate (P_i) by a pyrophosphatase. The 3′ end of all tRNAs, to which the amino acid attaches, has the sequence CCA. Class I synthetases (purple) attach the amino acid to the 2′ hydroxyl of the terminal adenylate in tRNA; class II synthetases (green) attach the amino acid to the 3′ hydroxyl. (Ad = adenine; Cyt = cytosine.)



▲ FIGURE 4-30 Recognition of a tRNA by aminoacyl synthetases. Aspartyl-tRNA synthetase (AspRS) is a class II enzyme, and arginyl-tRNA synthetase (ArgRS) is a class I enzyme. Shown here are the outlines of the three-dimensional structures of the two synthetases. The tRNA shown between them as a ribbon diagram will bind to either and is a slightly modified version of $tRNA^{Asp}$. It is used as an illustration of common surface interactions between tRNA and class I and II enzymes. Sites on the opposite sides of this modified $tRNA^{Asp}$ make contacts with the two enzymes: the blue balls show contacts with the class II AspRS; those that make contact with class I ArgRS are indicated by yellow balls. The synthetases are shown positioned away from the tRNA for clarity, but the fit of the surfaces at close range is obvious. The ability of ArgRS to interact with the noncognate $tRNA^{Asp}$ is lost when residue G37 in $tRNA^{Asp}$ is methylated, a normal modification that occurs in vivo. However, the shape and binding sites of this modified tRNA are characteristic of class I and class II interactions with tRNAs. This molecular graphic picture was produced using the DRAWNA program. [Adapted from M. Sissler et al., 1997, *Nucl. Acids Res.* **25**:4899; courtesy of R. Giegé.]

equilibrium of the aminoacylation reaction is driven further toward activation of the amino acid by hydrolysis of the high-energy phosphoanhydride bond in pyrophosphate. The overall reaction is

$$\text{Amino acid} + \text{ATP} + \text{tRNA} \xrightarrow{\text{enzyme}} \text{aminoacyl-tRNA} + \text{AMP} + 2\,P_i$$

The amino acid sequences of the aminoacyl-tRNA synthetases (ARSs) from many organisms are now known, and the three-dimensional structures of over a dozen enzymes of both classes have been solved. Each of these enzymes has a rather precise binding site for ATP (GTP is not admitted and CTP and UTP are too small) and binding pockets for its specific amino acid. Class I and class II enzymes bind to opposite faces of the incoming tRNAs. The binding surfaces of class I enzymes tend to be somewhat complementary to those of class II enzymes. These different binding surfaces and the consequent alignment of bound tRNAs probably account in part for the difference in the hydroxyl group to which the aminoacyl group is transferred (Figure 4-30). Because some amino acids are so similar structurally, aminoacyl-tRNA synthetases sometimes make mistakes. These are corrected, however, by the enzymes themselves, which check the fit in the binding pockets and facilitate deacylation of any misacylated tRNAs. This crucial function helps guarantee that a tRNA delivers the correct amino acid to the protein-synthesizing machinery.

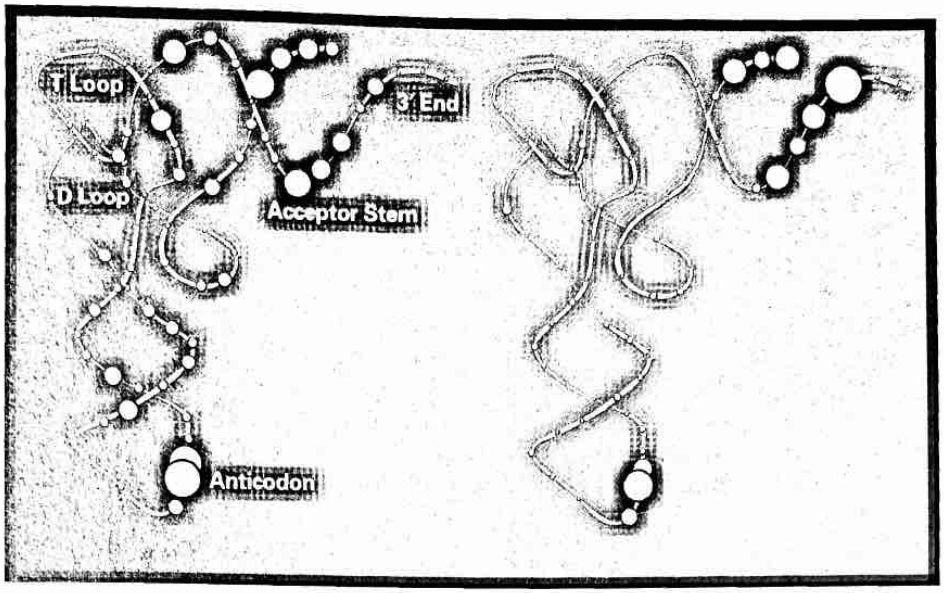
Each tRNA Molecule Is Recognized by a Specific Aminoacyl-tRNA Synthetase

The ability of aminoacyl-tRNA synthetases to recognize their correct cognate tRNAs is just as important to the accurate translation of the genetic code as codon-anticodon pairing. Once a tRNA is loaded with an amino acid, codon-anticodon pairing directs the tRNA into the proper ribosome site; if the wrong amino acid is attached to the tRNA, an error in protein synthesis results.

As noted already, each aminoacyl-tRNA synthetase can aminoacylate all the different tRNAs whose anticodons correspond to the same amino acid. Therefore, all these cognate tRNAs must have a similar binding site, or "identity element," that is recognized by the synthetase. One approach for studying the identity elements in tRNAs that are recognized by aminoacyl-tRNA synthetases is to produce synthetic genes that encode tRNAs with normal and various mutant sequences by techniques discussed in Chapter 7. The normal and mutant tRNAs produced from such synthetic genes then can be tested for their ability to bind purified synthetases.

Very probably no single structure or sequence completely determines a specific tRNA identity. However, some important structural features of several *E. coli* tRNAs that allow their cognate synthetases to recognize them are known. Perhaps the most logical identity element in a tRNA molecule is the anticodon itself. Experiments in which the anticodons of methionine tRNA ($tRNA^{Met}$) and valine tRNA ($tRNA^{Val}$) were interchanged showed that the anticodon is of major importance in determining the identity of these two tRNAs. In addition, x-ray crystallographic analysis of the complex between glutamine aminoacyl-tRNA synthetase (GlnRS) and glutamine tRNA ($tRNA^{Gln}$) showed that each of the anticodon bases neatly fits into a separate, specific "pocket" in the three-dimensional structure of GlnRS. Thus this synthetase specifically recognizes the correct anticodon.

However, the anticodon may not be the principal *identity element* in other tRNAs (see Figure 4-30). Figure 4-31 shows the extent of base sequence conservation in *E. coli* tRNAs that become linked to the same amino acid. Identity elements are found in several regions, particularly the end of the acceptor arm. A simple case is presented by $tRNA^{Ala}$:



▲ **FIGURE 4-31 Identity elements in tRNA involved in recognition by aminoacyl-tRNA synthetases, as demonstrated by both conservation and experimentation.** The 67 known tRNA sequences in *E. coli* were compared by computer analysis. The conserved nucleotides in different tRNAs that recognize the same amino acid are shown as yellow circles in the left drawing, with the tRNA chain in blue. Increasing size indicates increasing conservation of a base at a given position. It is clear that nucleotides in the anticodon loop and in the acceptor stem are most often similar when a particular amino acid must be recognized. This appreciation is heightened by results shown in the right drawing. Here, nucleotides that have been experimentally demonstrated to have a role in identity (selection of an amino acid by an ARS-tRNA complex) are shown as yellow circles. In this case, the circle size indicates the relative frequency that a given position acts as an identity element. [From W. H. McClain, 1993, *J. Mol. Biol.* **234**:257; also see L. D. H. Schulman and J. Abelson, 1988, *Science* **240**:1590.]

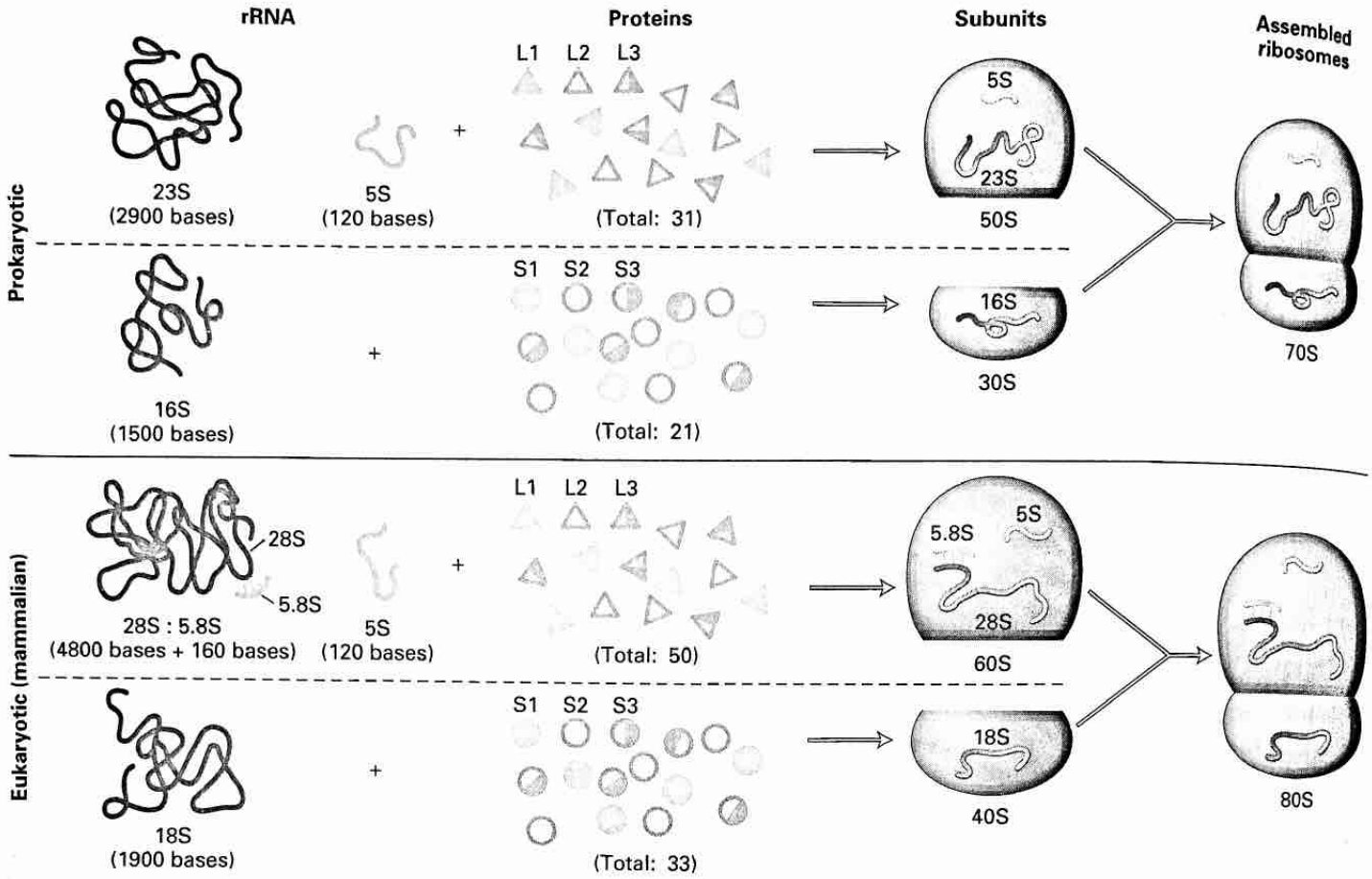
a single G·U base pair (G3·U70) in the acceptor stem is necessary and sufficient for recognition of this tRNA by its cognate aminoacyl-tRNA synthetase. Solution of the three-dimensional structure of additional complexes between aminoacyl-tRNA synthetases and their cognate tRNAs should provide a clear understanding of the rules governing the recognition of tRNAs by specific synthetases.

Ribosomes Are Protein-Synthesizing Machines

If the many components that participate in translating mRNA had to interact in free solution, the likelihood of simultaneous collisions occurring would be so low that the rate of amino acid polymerization would be very slow. The efficiency of translation is greatly increased by the binding of the mRNA and the individual aminoacyl-tRNAs to the most abundant RNA-protein complex in the cell—the ribosome. This two-part machine directs the elongation of a polypeptide at a rate of three to five amino acids added per second. Small proteins of 100–200 amino acids are therefore made in a minute or less. On the other hand, it takes 2 to 3 hours to make the largest known protein, titin, which is found in muscle and contains 30,000 amino acid residues. The machine that accomplishes this task must be precise and persistent.

With the aid of the electron microscope, ribosomes were first discovered as discrete, rounded structures prominent in animal tissues secreting large amounts of protein; initially, however, they were not known to play a role in protein synthesis. Once reasonably pure ribosome preparations were obtained, radiolabeling experiments showed that radioactive amino acids first were incorporated into growing polypeptide chains associated with ribosomes before appearing in finished chains.

A ribosome is composed of several different ribosomal RNA (rRNA) molecules and more than 50 proteins, organized into a large subunit and a small subunit. The proteins in the two subunits differ, as do the molecules of rRNA. The small ribosomal subunit contains a single rRNA molecule, referred to as *small rRNA;* the large subunit contains a molecule of *large rRNA* and one molecule each of two much smaller rRNAs in eukaryotes (Figure 4-32). The ribosomal subunits and the rRNA molecules are commonly designated in svedbergs (S), a measure of the sedimentation rate of suspended particles centrifuged under standard conditions (Chapter 3). The lengths of the rRNA molecules, the quantity of proteins in each subunit, and consequently the sizes of the subunits differ in prokaryotic and eukaryotic cells. (The small and large rRNAs are about 1500 and 3000 nucleotides long in bacteria and about 1800 and 5000 nucleotides long in humans.) Perhaps of more

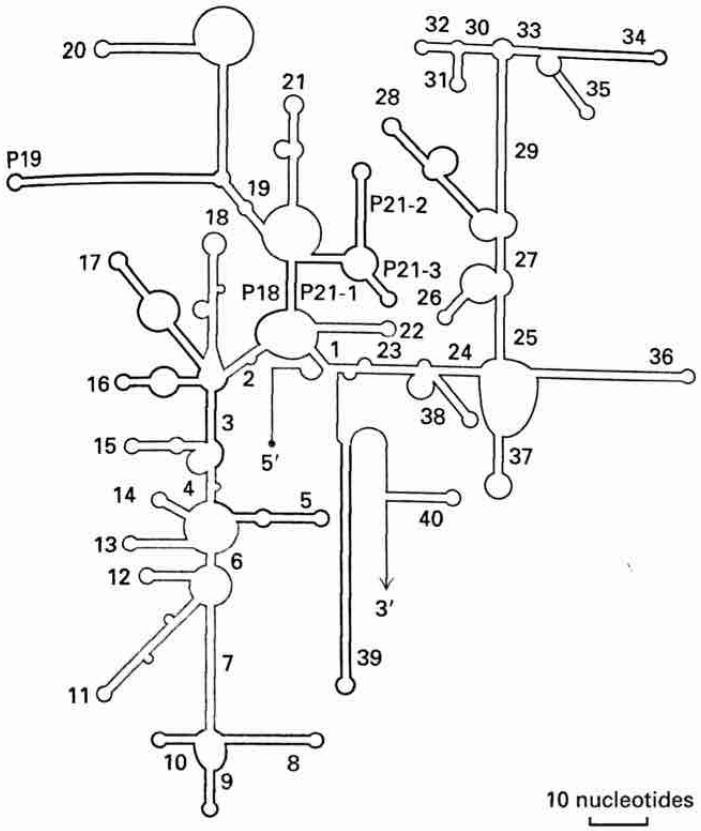


▲ **FIGURE 4-32 The general structure of ribosomes in prokaryotes and eukaryotes.** In all cells, each ribosome consists of a large and a small subunit. The two subunits contain rRNAs of different lengths, as well as a different set of proteins. All ribosomes contain two major rRNA molecules (dark red)—23S and 16S rRNA in bacteria, 28S and 18S rRNA in eukaryotes—and one or two small RNAs (light red). The proteins are named L1, L2, etc., and S1, S2, etc., depending on whether they are found in the large or the small subunit.

interest than these differences are the great structural and functional similarities among ribosomes from all species. This consistency is another reflection of the common evolutionary origin of the most basic constituents of living cells.

The sequences of the small and large rRNAs from several thousand organisms are now known. Although the primary nucleotide sequences of these rRNAs vary considerably, the same parts of each type of rRNA theoretically can form base-paired stem-loops, generating a similar three-dimensional structure for each rRNA in all organisms. Evidence that such stem-loops occur in rRNA was obtained by treating rRNA with chemical agents that cross-link paired bases; the samples then were digested with enzymes that destroy single-stranded rRNA, but not any cross-linked (base-paired) regions. Finally, the intact, cross-linked rRNA that remained was collected and sequenced, thus identifying the stem-loops in the original rRNA. Experiments of this type have located about 45 stem-loops at similar positions in small rRNAs from many different prokaryotes and eukaryotes (Figure 4-33). An even larger number of regularly positioned stem-loops have been demonstrated in large rRNAs. All the ribosomal proteins have been identified and their sequences determined, and many have been shown to bind specific regions of rRNA. It seems clear that the fundamental protein-synthesizing machinery in all present-day cells arose only once and has been modified about a common plan during evolution.

During protein synthesis, a ribosome moves along an mRNA chain, interacting with various protein factors and tRNA and very likely even undergoes shape changes. Despite the complexity of the ribosome, great progress has been made in determining both the overall structure of bacterial ribosomes and in identifying reactive sites that bind specific proteins, mRNA, and tRNA and that participate in important steps in protein synthesis. Quite detailed models of the large and small ribosomal subunits from *E. coli* have been constructed based on cryoelectron microscope and neutron-scattering studies (Figure 4-34). These studies not only have determined the dimensions and

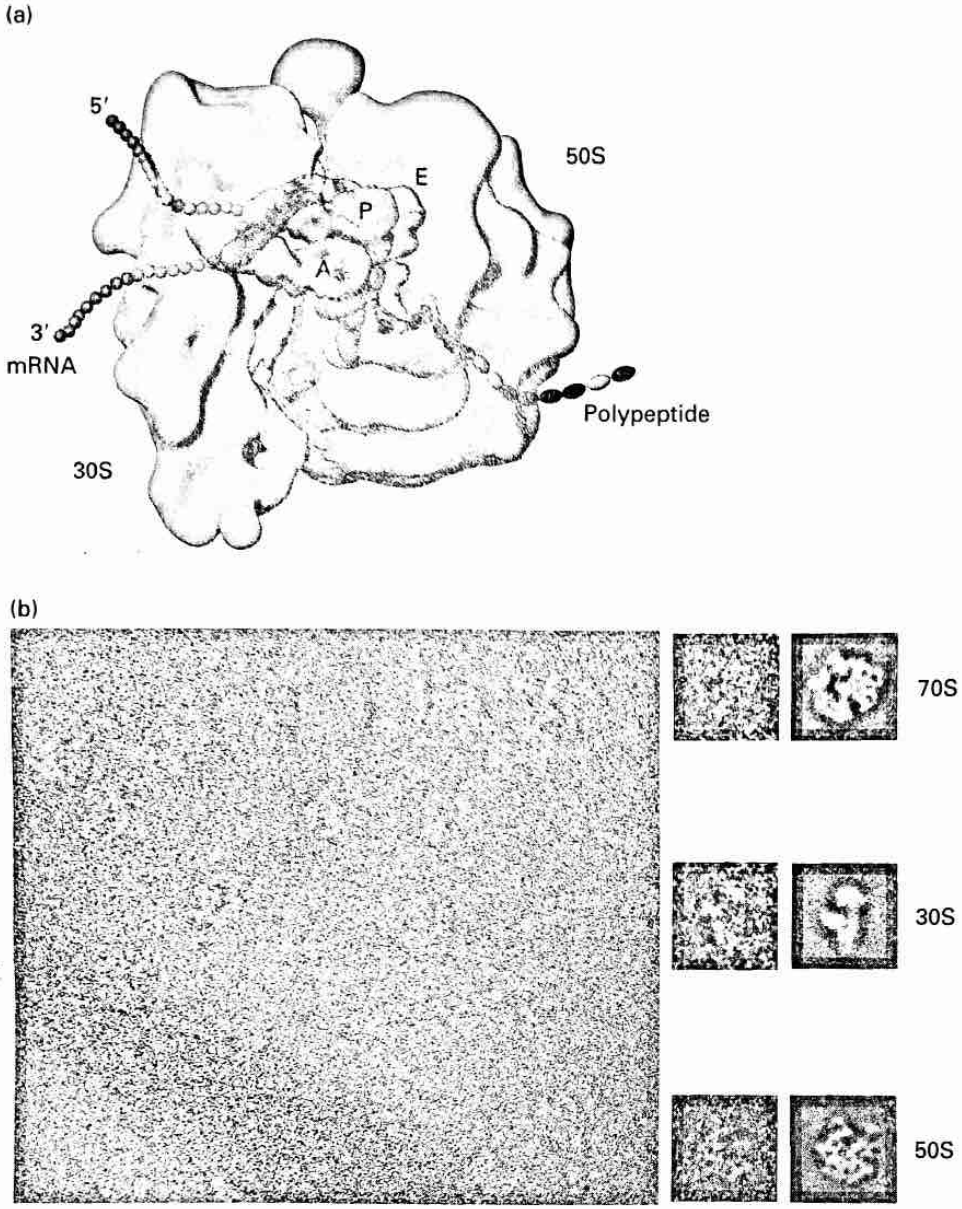


▲ FIGURE 4-33 Two-dimensional map of the secondary structure of the small (16S) rRNA from bacteria, showing the location of base-paired stems and loops. In general, the length and position of the stem-loops are very similar in all species, although the exact sequence varies from species to species. The most highly conserved regions are represented as red lines, and the numbered stem-loops unique to prokaryotes are preceded by a P. Eukaryotic small (18S) rRNAs exhibit a generally similar pattern of stem-loops, although, as with prokaryotes, a few are unique. [Adapted from E. Huysmans and R. DeWachter, 1987, *Nucl. Acids Res.* **14**:73].

overall shape of the ribosomal subunits, but also have localized the positions of tRNAs bound to the ribosome during protein chain elongation. Powerful chemical experiments have also helped unravel the complex interactions between proteins and RNAs. In a technique called **footprinting**, for example, ribosomes are treated with chemical reagents that modify single-stranded RNA unprotected by binding either to protein or to another RNA. If the total sequence of the RNA is known, then the location of the modified nucleotides can be located within the molecule. (This technique, which is also useful for locating protein-binding sites in DNA, is described in Chapter 10.) Thus the overall structure and function of ribosomes during protein synthesis is finally, after 40 years, yielding to successful experiments. How these results aid in understanding the specific steps in protein synthesis is described in the next section.

SUMMARY The Three Roles of RNA in Protein Synthesis

- Genetic information is copied into mRNA in the form of a commaless, overlapping, degenerate triplet code. Each amino acid is encoded by one or more three-base sequences, or codons, in mRNA. Each codon specifies one amino acid, but most amino acids are encoded by multiple codons (see Table 4-2).
- The AUG codon for methionine is the most common start codon, specifying the amino acid at the NH_2-terminus of a protein chain. Three codons function as stop codons and specify no amino acids.
- A reading frame, the uninterrupted sequence of codons in mRNA from a specific start codon to a stop codon, is translated into the linear sequence of amino acids in a protein.
- Decoding of the nucleotide sequence in mRNA into the amino acid sequence of proteins depends on transfer RNAs and amino-acyl tRNA synthetases (see Figure 4-25).
- All tRNAs have a similar three-dimensional structure that includes an acceptor arm for attachment of a specific amino acid and a stem-loop with a three-base anticodon sequence at its ends (see Figure 4-26). The anticodon can base-pair with its corresponding codon or codons in mRNA.
- Because of nonstandard interactions, a tRNA may base-pair with more than one mRNA codon, and conversely, a particular codon may base-pair with multiple tRNAs.
- Each of the 20 aminoacyl-tRNA synthetases recognizes a single amino acid and covalently links it to a cognate tRNA, forming an aminoacyl-tRNA (see Figure 4-29). This reaction activates the amino acid, so it can participate in peptide-bond formation.
- The composition of ribosomes—the large ribonucleoprotein complexes on which proteins are synthesized—is quite similar in all organisms (see Figure 4-32). All ribosomes are composed of a small and a large subunit. Each contains numerous different proteins and one rRNA (small or large). The large subunit also contains one accessory RNA (5S).
- Analogous rRNAs from many different species fold into quite similar three-dimensional structures containing numerous stem-loops and binding sites for proteins, mRNA, and tRNAs. As a ribosome moves along an mRNA, a region of the large rRNA molecule in each ribosome sequentially binds the aminoacylated ends of incoming tRNAs and probably catalyzes peptide-bond formation (see Figure 4-34).



◀ FIGURE 4-34 Overall structure of the *E. coli* ribosome at 25-Å resolution inferred from cryoelectron microscopy and three-dimensional reconstruction based on the analysis of 4300 individual projections. (a) This model shows the shapes of the large (blue) and small (yellow) subunits of the ribosome with three aminoacyl-tRNAs (pink, green, yellow) superimposed at the A, P, and E sites. The roles of these tRNA-binding sites during protein synthesis are discussed later. Chemical cross-linking experiments have demonstrated that the mRNA (orange beads) passes close to the anticodon loops of the tRNAs and that the nascent polypeptide chain is buried in the tunnel in the large ribosomal subunit that begins within 10–15 Å of the 3′ aminoacylated end of the tRNAs. The tunnel termination site on the ribosome surface has also been accurately mapped. (b) Large panel shows a field of 70S ribosomes. Small panels (left) show cryoelectron microscopy images of a single 70S ribosome, small (30S) subunit, and large (50S) subunit. Small panels (right) show computer-derived averages of many dozens of images in the same orientation. Cryoelectron microscopy is carried out on unstained samples of ribosomes or subunits flash frozen as "vitreous ices" (without ice crystals) in a very thin layer of water (Chapter 5). Individual images are analyzed by computer projections. [See R. K. Agrawal et al., *Cell*, in press; J. Frank, 1995, *Nature* **356**:441; J. Frank et al., 1995, *Biochem. Cell Biol.* **73**:757. Courtesy of J. Frank.]

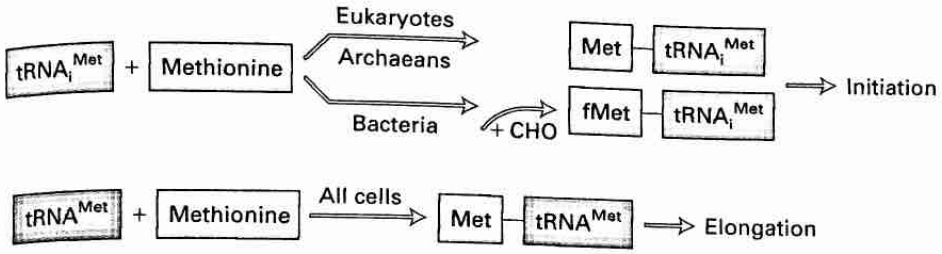
4.5 Stepwise Formation of Proteins on Ribosomes

The previous sections have introduced the major participants in protein synthesis—mRNA, aminoacylated tRNAs, and ribosomes containing large and small rRNAs. Here, we take a detailed look at how these components are brought together to carry out the biochemical events leading to formation of proteins on ribosomes. The complex process of translating mRNA into protein can be divided into three stages—initiation, elongation, and termination—which we consider in order.

The AUG Start Codon Is Recognized by Methionyl-tRNA$_i^{Met}$

As noted earlier, the AUG codon for methionine functions as the start codon in the vast majority of mRNAs. A critical aspect of initiation is to begin protein synthesis at the start codon, thereby setting the stage for the correct in-frame translation of the entire mRNA. Both prokaryotes and eukaryotes contain two different methionine tRNAs: tRNA$_i^{Met}$ can initiate protein synthesis, and tRNA^{Met} can incorporate methionine only into a growing protein chain (Figure 4-35). The same aminoacyl-tRNA synthetase (MetRS) charges both tRNAs with methionine. But *only* Met-tRNA$_i^{Met}$ (i.e., activated methionine attached to tRNA$_i^{Met}$) can bind at the appropriate site on the small ribosomal subunit (the P site; see discussion below) to begin synthesis of a protein chain. (The regular tRNA^{Met} binds only to another ribosomal site, the A site, as described later.) In bacteria, but not in archaeans or eukaryotes, the amino group of the methionine of Met-tRNA$_i^{Met}$ is modified by addition of a formyl group and is sometimes designated fMet-tRNA$_i^{Met}$. However, Met-tRNA$_i^{Met}$ is commonly used to designate the initiator tRNA in all cells.

During the initial stage of protein synthesis in all cells, a ribosome assembles, complexed with a mRNA and an



◀ **FIGURE 4-35 Two types of methionine tRNA are found in all cells.** One, designated $tRNA_i^{Met}$, is used exclusively to start protein chains, and the other, designated $tRNA^{Met}$, delivers methionine to internal sites in a growing protein chain. In bacteria, a formyl group (CHO) is added to methionyl-$tRNA_i^{Met}$, forming fMet-$tRNA_i^{Met}$.

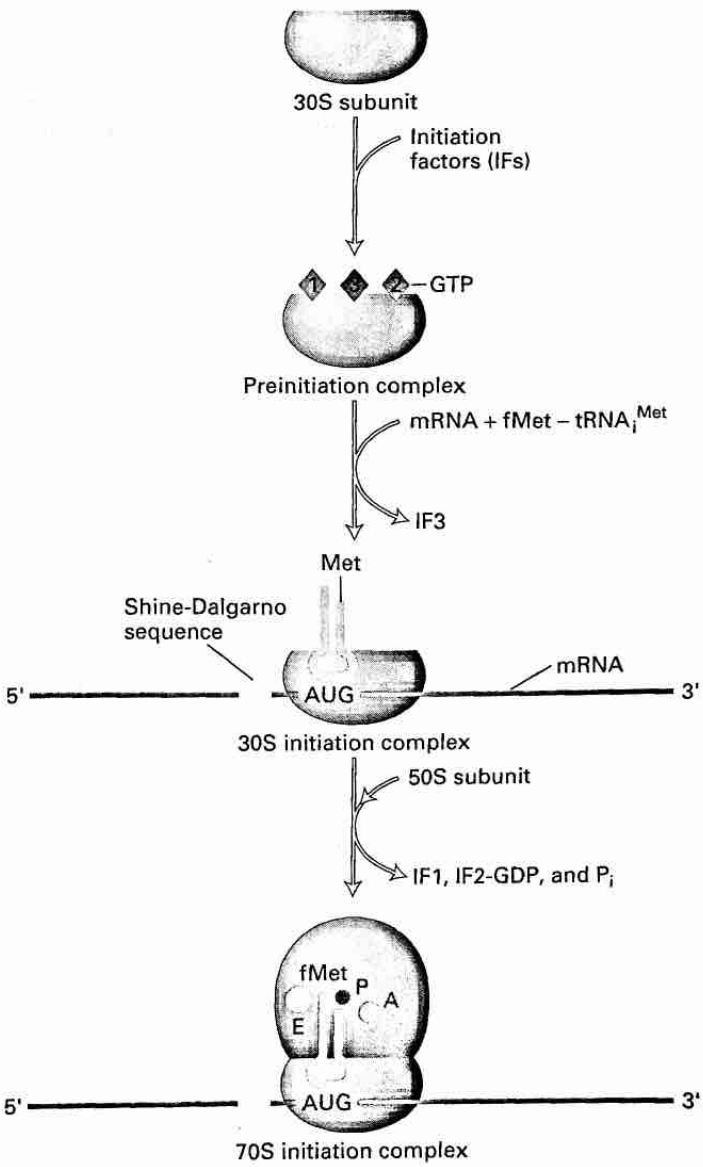
activated initiator tRNA, which is correctly positioned at the start codon. Because the details of initiation and the mechanism for locating the translation start site differ in bacteria and eukaryotes, we discuss the two systems separately.

Bacterial Initiation of Protein Synthesis Begins Near a Shine-Dalgarno Sequence in mRNA

To begin the assembly of a bacterial translation complex, sequential interactions occur between specific proteins referred to as **initiation factors** (IFs) and the small (30S) ribosomal subunit. The resulting preinitiation complex together with fMet-$tRNA_i^{Met}$ then binds to the mRNA at a specific site usually located quite near the AUG initiation codon. This binding, assisted by IF1 and IF3, yields the *30S initiation complex* (Figure 4-36).

In most bacteria, the small ribosomal subunit identifies start codons through interaction of the small (16S) rRNA with an eight-nucleotide sequence in mRNA called the *Shine-Dalgarno sequence* after its discoverers. This sequence, located near the AUG start codon, base-pairs to a sequence at or very near the 3′ end of 16S rRNA, thereby binding the mRNA and small ribosomal subunit to each other. Although the Shine-Dalgarno sequence varies some at different initiation sites, on average six out of eight nucleotides are complementary to the 3′-end sequence of 16S rRNA. Thus bacterial rRNA plays a direct role in recruiting the small ribosomal subunit to a translation start site on the mRNA. The long mRNAs produced from bacterial operons contain multiple *internal* Shine-Dalgarno sequences located near the start sites for each of the encoded proteins (see Figure 4-17a). Because a small ribosomal subunit can bind to each Shine-Dalgarno sequence, synthesis of the different proteins can occur independently and simultaneously from the multiple start sites on these long mRNAs.

Assembly of a complete *70S initiation complex* is achieved by the recruitment of the large ribosomal subunit,



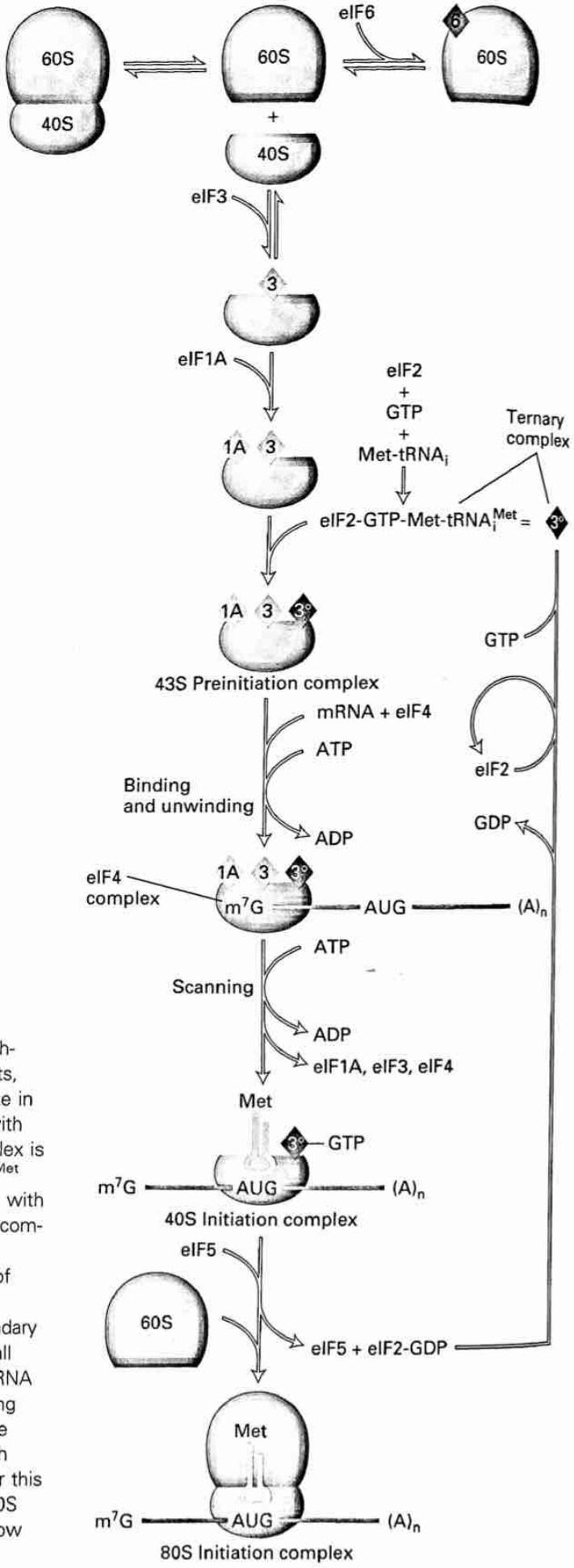
▶ **FIGURE 4-36 Bacterial initiation of protein synthesis.** The three initiation factors (IF1, IF2, and IF3) bind to the small subunit of the ribosome. The 30S preinitiation complex, together with the fMet-$tRNA_i^{Met}$, binds the mRNA, guided by the Shine-Dalgarno sequence, which is complementary to the 3′ terminus of the 16S rRNA. The AUG sequence is then located, from which protein initiation will begin. Addition of the large (50S) ribosomal subunit is accompanied by release of the protein factors and hydrolysis of IF2-GTP to yield the complete 70S initiation complex. The charged initiator tRNA (fMet-$tRNA_i^{Met}$) is positioned in the P site. The ribosome now is ready for entry of the next aminoacyl-tRNA at the A site, which will lead to the formation of the first peptide bond. The initiation process is similar in archaeans, except there is no formylation of the Met-$tRNA_i^{Met}$.

an energy-requiring step that is powered by hydrolysis of GTP bound to IF2; in the process IF1, IF2-GDP, and P_i are discharged. In the completed initiation complex, the charged initiator tRNA is positioned at the *P site* on the ribosome. As discussed later, during elongation the growing peptide chain remains attached to the tRNA located at this site.

Eukaryotic Initiation of Protein Synthesis Occurs at the 5′ End and Internal Sites in mRNA

Initiation of protein synthesis in eukaryotic cells, as in bacteria, begins with formation of a preinitiation complex prior to mRNA binding. However, the process differs significantly from initiation in bacteria. Two eukaryotic factors, eIF3 (a large multimeric protein with about eight subunits) and eIF6, serve to keep the ribosomal subunits apart after they have previously finished synthesizing a protein chain so that they can participate again in starting a new chain. To assemble a eukaryotic preinitiation complex, an active "ternary complex" of eIF2 bound to a GTP molecule and Met-tRNA$_i^{Met}$ associates with a small (40S) ribosomal subunit complexed with two other factors, eIF3 and eIF1A, which stabilize binding of the ternary complex (Figure 4-37). Cells can regulate protein synthesis by phosphorylating a serine residue on the eIF2 bound to GDP; this complex is then unable to bind Met-tRNA$_i^{Met}$, thus inhibiting protein synthesis.

Recognition of the translation start site by the small ribosomal subunit in the eukaryotic preinitiation complex requires different types of factors than in prokaryotic cells. Most eukaryotic mRNAs have a single start site near the 5′ capped end of the mRNA (see Figure 4-19). Interaction of the small ribosomal subunit with the methylated 5′-cap structure



▶ **FIGURE 4-37 Eukaryotic initiation of protein synthesis.** Establishment of the initiation complex must follow a precise sequence of events, and the large and small ribosomal subunits must be kept apart until late in the process. Two eukaryotic factors, eIF3 (a large multimeric protein with about eight subunits) and eIF6 fulfill this role. A 43S preinitiation complex is formed when a ternary complex of eIF2 bound to GTP and Met-tRNA$_i^{Met}$ associates with the small (40S) ribosomal subunit, which is complexed with two other factors, eIF3 and eIF1A, that stabilize binding of the ternary complex. (The inactive eIF2-GDP fails to bind Met-tRNA$_i^{Met}$; it also can be phosphorylated, thereby inhibiting protein initiation.) The 5′ cap (m^7G) of the mRNA to be translated is guided to the preinitiation complex by a subunit of the multiprotein eIF4 complex, which also unwinds any secondary structure at the 5′ end of the mRNA. Subsequent scanning by the small ribosomal subunit, assisted by eIF3 and eIF4G, positions the initiator tRNA at the AUG start codon, yielding the 40S initiation complex and releasing eIF1A, eIF3, and eIF4. With the Met-tRNA$_i^{Met}$ properly positioned at the start codon, another factor, eIF5, assists union of the 40S complex with the 60S subunit. Hydrolysis of GTP in eIF2-GTP provides the energy for this step. Factors eIF5 and eIF2-GDP are then released, yielding the final 80S initiation complex, with Met-tRNA$_i^{Met}$ at the P site. The complex can now accept the second aminoacyl-tRNA (see Figure 4-40).

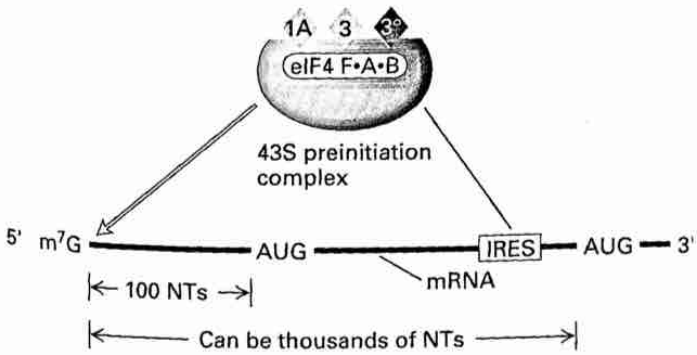
present on all eukaryotic mRNAs requires a set of proteins collectively called *eIF4*. After the methylated cap structure is recognized by eIF4F, any secondary structure at the 5′ end is removed by an associated helicase activity. The bound preinitiation complex, which contains the eIF3 group of proteins, then probably slides along the mRNA, most often stopping at the first AUG. However, selection of the initiating AUG is facilitated by specific surrounding sequences called *Kozak sequences* (for Marilyn Kozak, who defined them):

mRNA 5′–ACCAUGG–

The A preceding the AUG seems to be the most important nucleotide affecting initiation efficiency. During the selection of the starting AUG, there is a proven interaction between the eIF3 complex, the 40S ribosomal subunit, and the eIF4G component of the eIF4 complex. Scanning of the mRNA by the preinitiation complex eventually yields a *40S initiation complex* in which Met-tRNA$_i^{Met}$ is correctly positioned at the translation start site.

Once the small ribosomal subunit with its bound Met-tRNA$_i^{Met}$ is correctly positioned at the start codon, union with the large (60S) ribosomal subunit completes formation of the *80S initiation complex* (see Figure 4-37).

Initiation of translation of most mRNAs by the eukaryotic protein-synthesizing machinery begins near the 5′ capped end as just described. However, some viral mRNAs, which are translated by the host-cell machinery in infected eukaryotic cells, lack a 5′ cap; translation of these mRNAs is initiated at *internal ribosome entry sites* (IRESs). It is now known that some cellular mRNAs also can be translated beginning at IRESs (Figure 4-38). Many of the same protein factors that assist in ribosome scanning from a 5′ cap are required for locating an internal AUG start codon, but exactly how an IRES is recognized is less clear.



▲ **FIGURE 4-38 Eukaryotic initiation of protein synthesis can occur near the 5′ capped end and at internal ribosome entry sites (IRESs) of a mRNA.** In the first case, the 43S preinitiation complex binds to the 5′-cap structure (m^{7}G) and then slides along the mRNA until it reaches an acceptable AUG start codon, usually within about 100 nucleotides (see Figure 4-39). Much less frequently, the preinitiation complex binds to an IRES within the mRNA sequence, far downstream of the 5′ end, and then scans downstream for an AUG start codon. (Nts = nucleotides.)

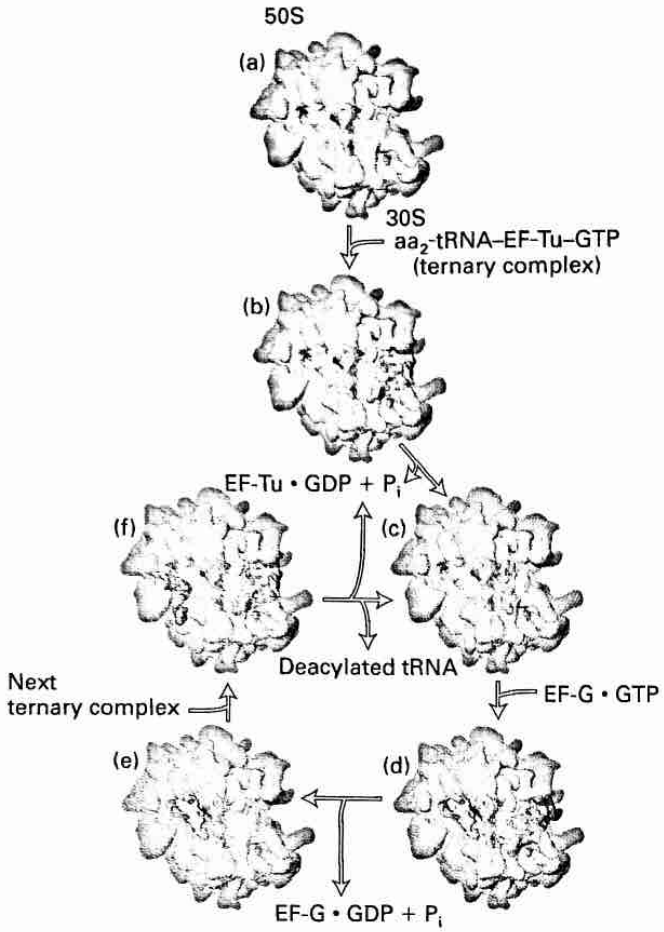
During Chain Elongation Each Incoming Aminoacyl-tRNA Moves through Three Ribosomal Sites

The correctly positioned bacterial 70S or eukaryotic 80S ribosome–Met-tRNA$_i^{Met}$ complex is now ready to begin the task of stepwise addition of amino acids by the in-frame translation of the mRNA. As is the case with initiation, a set of special proteins, termed **elongation factors** (EFs), are required to carry out the process. The key steps in elongation are entry of each succeeding aminoacyl-tRNA, formation of a peptide bond, and the movement, or *translocation,* of the ribosome with respect to the mRNA. For example, the second aminoacyl-tRNA is brought into the ribosome as a ternary complex in association with an EF1 α-GTP in eukaryotes (EF-Tu–GTP in bacteria) and becomes bound to the A site on the ribosome (Figure 4-39). Recall that the initiating Met-tRNA$_i^{Met}$ is bound at the P site and base-paired with the AUG start codon. If the anticodon of the incoming (second) aminoacyl-tRNA correctly matches the second codon of the mRNA, a tight binding ensues at the A site. If this second codon does not match the incoming aminoacyl-tRNA, it diffuses away. The choice of the correct aminoacyl-tRNA and its tight binding at the A site requires energy that is supplied by hydrolysis of the EF1 α-GTP (or EF-Tu–GTP) complex.

With the initiating Met-tRNA$_i^{Met}$ at the P site and the second aminoacyl-tRNA tightly bound at the A site, the α amino group of the second amino acid reacts with the "activated" (aminoacylated) methionine on the initiator tRNA, forming a peptide bond. A revolutionary piece of biochemistry has provided insight about this critical *peptidyltransferase reaction,* which actually effects the transfer of the growing peptide chain on the peptidyl-tRNA at the P site to the activated amino acid on the incoming aminoacyl-tRNA. In bacterial protein synthesis, the 23S rRNA in the large ribosomal subunit itself may carry out the peptidyltransferase function. Supporting this catalytic role for 23S rRNA is the finding that when the vast majority of the protein is carefully removed from the large ribosomal subunit and the remaining rRNA, or even a pure segment of 23S rRNA, is mixed with analogs of aminoacylated-tRNA and peptidyl-tRNA, the peptidyltransferase reaction still occurs.

During the process of peptide synthesis and tRNA site changes, the ribosome is moved along the mRNA a distance equal to one codon with the addition of each amino acid. This translocation step is catalyzed by eukaryotic EF2-GTP (bacterial EF-G–GTP), which is hydrolyzed to provide the required energy. Referring again to Figure 4-39, we can see that after peptide linkage tRNA$_i^{Met}$, now without its activated methionine, is moved to an exit (E) site on the ribosome and is soon discharged. Concurrently, another ternary complex, carrying the next amino acid to be added, enters the ribosome, and the cycle continues.

MEDIA CONNECTIONS
Protein Synthesis

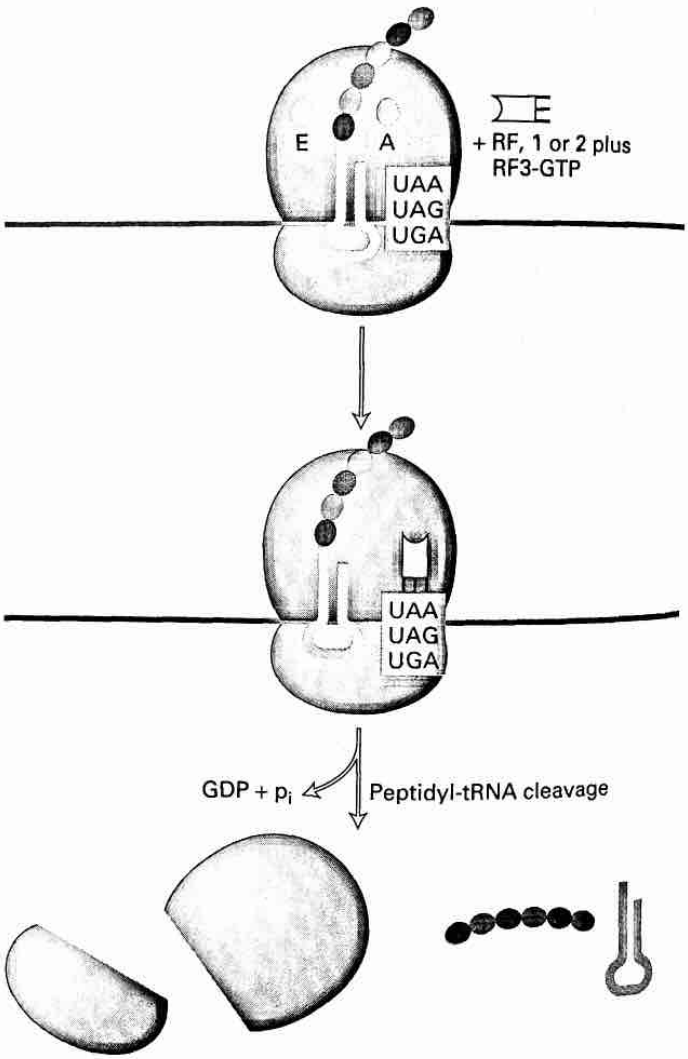


▲ **FIGURE 4-39 The elongation cycle in protein synthesis visualized for *E. coli* ribosomes.** The models of the *E. coli* ribosome associated with tRNAs in various states are based on cryoelectron microscopy studies and cross-linking experiments (see Figure 4-34). (a) Elongation begins on the 70S initiation complex with Met-$tRNA_i^{Met}$ (green) in the P site, assembled as depicted for the eukaryotic system in Figure 4-37. Both ribosomal subunits are shown in blue. (b) A ternary complex consisting of elongation factor EF-Tu bound to GTP (orange and red) plus the correct aminoacyl-tRNA (purple) to decode the second codon then enters the ribosome. (c) Hydrolysis of EF-Tu–GTP to EF-Tu–GDP and P_i brings aa_2-$tRNA_2^{aa}$ to the A site (purple). (d, e) After binding of EF-G–GTP (dark blue), the peptidyl synthesis and translocation steps occur: Met-$tRNA_i^{Met}$ (now yellow and moving out of the P site) donates its methionine to the aa_2-$tRNA_2^{aa}$ (now green and in the P site), producing aa_2-$tRNA_2^{aa}$–Met. Hydrolysis of the bound EF-G–GTP furnishes energy for translocation, and the products are released. (f) $tRNA_i^{Met}$ (now brown) moves completely to the E site and is ejected; a new ternary complex containing aa_3-$tRNA_3^{aa}$ arrives and the cycle can continue (c → d → e → f). [Adapted from R. K. Agrawal et al., *Cell,* in press.]

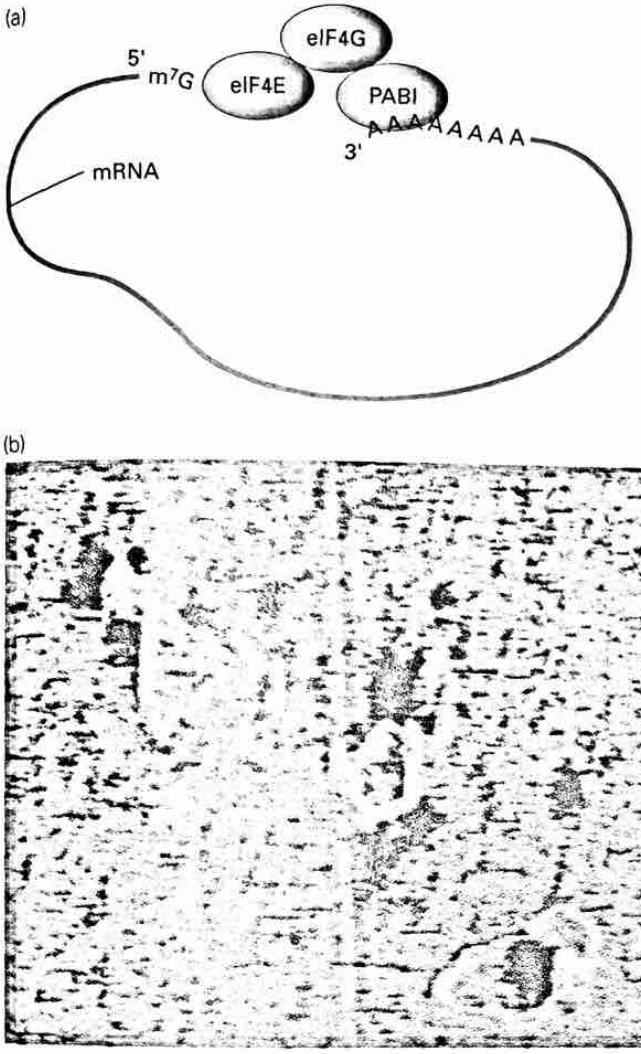
Protein Synthesis Is Terminated by Release Factors When a Stop Codon Is Reached

The final phase of protein synthesis, like initiation and elongation, requires highly specific molecular signals that decide the fate of the mRNA-ribosome-tRNA-peptidyl complex. Recent research has uncovered two types of specific **termination** (release) **factors** in bacteria and their eukaryotic counterparts. Bacterial RF1 and RF2 (and eukaryotic eRF1), whose shape is thought to be similar to that of tRNAs ("molecular mimicry"), apparently act by recognizing the stop codons themselves. RF1 recognizes UAG, and RF2 recognizes UGA; both these factors recognize UAA. Like some of the initiation factors and the elongation factors discussed previously, the third release factor, RF3 (and its eukaryotic counterpart), is a GTP-binding protein. RF3 acts in concert with the codon-recognizing factors to promote cleavage of the peptidyl-tRNA, thus releasing the completed protein chain (Figure 4-40).

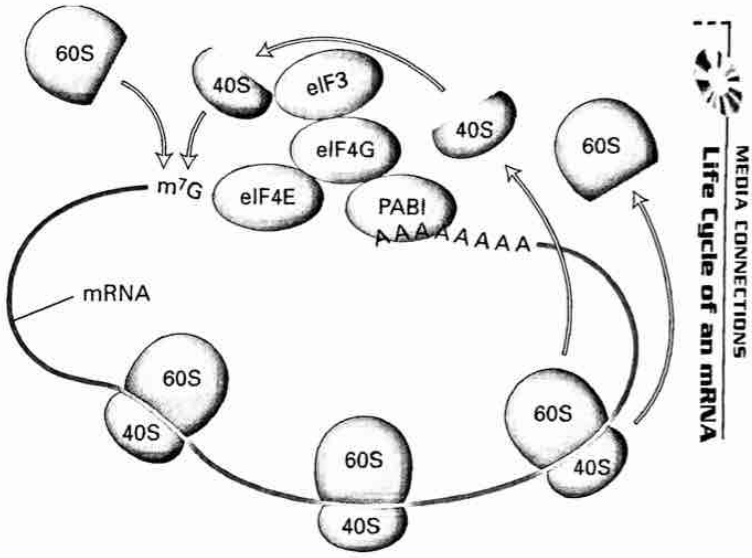
Folding of a newly released protein into its native three-dimensional conformation is facilitated by other proteins



▲ **FIGURE 4-40 Termination of translation.** When a ribosome bearing a nascent protein chain reaches a stop codon (UAA, UGA, UAG), release factors (RFs) enter the ribosomal complex, probably at or near the A site. In bacteria, RF1 or RF2 first recognizes a stop codon. RF3-GTP then catalyzes cleavage of the peptide chain from the tRNA and release of the two ribosomal subunits, reactions that require hydrolysis of GTP.



▲ **FIGURE 4-41 Formation of circular eukaryotic mRNA by protein-protein interactions bridging the 5′ and 3′ ends.** (a) Poly(A)-binding protein I (PABI) binds both to the 3′ poly(A) tail on a eukaryotic mRNA and to subunits of the initiation factor eIF4, which are bound to the m^7G cap at the 5′ end of the mRNA. (b) Force-field electron micrographs of mRNA circles held together by interactions between the three purified proteins described in part (a). [Part (a) adapted from A. Sachs et al., 1997, *Cell* **89**:841; part (b) courtesy of A. Sachs.]



▲ **FIGURE 4-42 Model of protein synthesis on circular polysomes and recycling of ribosomal subunits.** Multiple individual ribosomes can simultaneously translate a eukaryotic mRNA, shown here in circular form stabilized by interactions between proteins bound at the 3′ and 5′ ends. When a ribosome completes translation and dissociates from the 3′ end, the separated subunits can readily find the nearby 5′ cap (m^7G) and initiate another round of synthesis. This rapid recycling of subunits on the same mRNA contributes to the formation of polysomes and increases the efficiency of translation.

called *chaperones* (Chapter 3). Additional release factors then promote dissociation of the ribosome, freeing the subunits, mRNA, and terminal tRNA for another round of protein synthesis.

Simultaneous Translation by Multiple Ribosomes and Their Rapid Recycling Increase the Efficiency of Protein Synthesis

As noted earlier, translation of a single eukaryotic mRNA molecule to yield a typical-sized protein takes 30–60 seconds. Two phenomena significantly increase the overall rate at which cells can synthesize a protein: the simultaneous translation of a single mRNA molecule by multiple ribosomes and rapid recycling of ribosomal subunits after they disengage from the 3′ end of a mRNA. Simultaneous translation of an mRNA by multiple ribosomes is readily observable in electron micrographs and by sedimentation analysis, revealing mRNA attached to multiple ribosomes bearing nascent growing peptide chains. These structures are referred to as *polyribosomes* or *polysomes*. However, the second phenomenon—rapid recycling of ribosomal subunits—was not discovered so easily.

Polysomes in some tissues were seen in electron micrographs to be circular. It was not until recently that studies with yeast cells explained the circular shape of polyribosomes and suggested the mode by which ribosomes recycle efficiently.

Researchers initially found that the poly(A) tail at the 3′ end of eukaryotic mRNA stimulates in vitro translation by yeast protein-synthesizing components. This observation at first seems puzzling, since translation is begun upstream (usually near the 5′ end) of the mRNA, at a great distance from the poly(A) tail. A cytosolic protein found in all eukaryotic cells, *poly(A)-binding protein I* (PABI), proved literally to be a missing link in explaining the relationship of the poly(A) tail to translation. In earlier studies, yeast cells that had mutations in the gene encoding PABI were found to translate mRNA poorly. The role of PABI in improving efficiency of protein synthesis was demonstrated by the recent finding that PABI and two components of yeast eIF4

(4G and 4E) can interact on mRNA to circularize the molecule (Figure 4-41). By linking the poly(A) tail to the 5′ end, PABI positions recently disengaged ribosomal subunits near the start site for protein synthesis, thus enhancing their reentry into the process and improving efficiency. Thus in many eukaryotic cells, ribosomal subunits may follow the circular pathway depicted in Figure 4-42, providing another example of the merger of form and function.

SUMMARY Stepwise Formation of Proteins on Ribosomes

- Of the two methionine tRNAs found in all cells, only one ($tRNA_i^{Met}$) functions in initiation of protein synthesis.
- During initiation, the large and small ribosomal subunits assemble at the translation start site in an mRNA molecule, and the tRNA carrying the amino-terminal methionine (Met-$tRNA_i^{Met}$) base-pairs with the start codon, forming a 70S (bacterial) or 80S (eukaryotic) initiation complex (see Figures 4-36 and 4-37).
- The mechanism by which the small ribosomal subunit locates the start codon differs somewhat in bacteria and eukaryotes. In all cells, however, it depends on specific initiation factors and requires energy supplied by hydrolysis of GTP.
- Elongation of a protein chain entails three steps, which occur over and over (see Figure 4-39). First, an aminoacyl-tRNA binds tightly to the A site on the ribosome and base-pairs with its corresponding codon in the mRNA. Second, a peptide bond is formed between the incoming amino acid and the growing chain at the P site, transferring the peptidyl chain to the incoming tRNA. Peptide-bond formation is accompanied by movement of the incoming tRNA to the P site and of the now unloaded tRNA to the E site. Finally, the ribosome translocates to the next codon, and the old tRNA is discharged from the ribosome. As with initiation, elongation requires specific protein factors and GTP hydrolysis.
- The large rRNA in a ribosome most likely catalyzes the peptidyltransferase reaction in which peptide bonds between adjacent amino acids are formed.
- Termination is carried out by two types of termination factors: those that recognize stop codons and those that promote hydrolysis of peptidyl-tRNA (see Figure 4-40).
- In eukaryotic cells, multiple ribosomes commonly are bound to a single mRNA, forming a circular polysome; all the bound ribosomes translate the mRNA concurrently. As each ribosome completes translation and is released from the 3′ end of the mRNA, the subunits probably reassemble quickly at the 5′ end (see Figure 4-42). This circular ribosome pathway greatly increases the efficiency of protein synthesis.

PERSPECTIVES for the Future

In this chapter we first considered the basic chemical makeup of DNA and RNA and some elementary principles about their two- and three-dimensional structures. We then described the elementary properties of RNA and DNA polymerases, which catalyze the synthesis of nucleic acids in cells. The large-scale protein machines that make RNA and DNA according to regulated plans are examined in detail in later chapters. Finally, we discussed the genetic code and the participation of tRNA and the protein-synthesizing machine, the ribosome, in decoding the information in mRNA to allow accurate assembly of protein chains.

These facts form the bedrock principles of molecular cell biology and are not likely to change. Nevertheless, important extensions of these principles are now being made, and these will form a new generation of fundamental principles. For example, sequencing of the entire genomes of many different bacteria and archaea has been completed, and dozens more are under way. The genomic sequence of baker's yeast, a simple eukaryote, also has been determined, as has that of the nematode worm *C. elegans*. The genome of *Drosophila melanogaster* may well be finished by the time this book is released, and early in the next century, the human genome will yield its complete sequence to researchers. As discussed in Chapter 7, once the genomic sequence of an organism is known, all the possible encoded proteins can be deduced with the aid of computer programs.

Paralleling the work on genomic analysis, biochemists and structural biologists are collaborating to determine through structural studies all the possible folds that protein domains assume in order to expand our knowledge of the interactions between the cell's principal macromolecules, namely, protein-protein, protein-RNA, and protein-DNA interactions. The fundamental principles of interaction and a catalogue of typical macromolecular interactions will be in the hands of scientists early in the twenty-first century.

Much of current biological research, and that in the next few years, is focused on discovering how these molecular interactions endow cells with decision-making capacity and their special properties. For this reason several of the following chapters describe current knowledge about how such interactions regulate RNA and DNA synthesis and how such regulation endows cells with the capacity to become specialized and grow into complicated organs. Other chapters deal with how protein-protein interactions underlie the construction of specialized organelles in cells, how cells and cell organelles move, and how these properties of mobility coupled with regulated growth lead to construction of body plans of correct size and shape.

One frontier, neurobiology, will benefit from all these molecular and cellular advances but will require a new level of integrated research for ultimate solutions. Future advances in the molecular biology of the specialized cell types that compose the brain will, no doubt, contribute to understanding brain function. It is clear, however, that neurobiology poses a new

challenge in understanding how numerous functions within and between brain cells are integrated. Such integration, which exists at a vastly more complicated level in the brain than elsewhere, is critical to brain functioning.

PERSPECTIVES in the Literature

Macromolecular synthesis is carried out by multiprotein machines. Compare the machines for making RNA, DNA, and protein. What are the similarities and differences among them that seem most cogent to their individual tasks? As we acquire ever deeper knowledge of these machines, this will be a worthwhile project for every student, and indeed for every practicing biologist. Start your exploration of the literature in this broad topic with the following references:

Ban, N., et al. 1998. A 9Å resolution X-ray crystallographic map of the large ribosomal subunit. *Cell* 93:1105–1115.

Holstege, F. C. P., and R. A. Young. 1999. Transcriptional regulation: contending with complexity. *Proc. Nat'l Acad. Sci. USA* 96:2–4.

Holstege, F. C. P., and R. A. Young. 1998. Review Issue, Macromolecular Machines. *Cell* 92:291–423.

Testing Yourself on the Concepts

1. What are the minimal components needed for information flow from DNA to protein?

2. Investigators use the PCR (polymerase chain reaction) to amplify regions of DNA contained between two single-stranded oligonucleotide primers that anneal to opposite strands of a DNA template. In designing primers, it is important to consider sequence information, T_m, and directionality of the primers. Why?

3. Mutations that alter DNA can alter the function of an expressed protein. A mutation is characterized in a gene from a model organism and found to contain a single base change relative to the wild-type gene, yet the encoded protein still functions in this organism. How is this possible?

MCAT/GRE-Style Questions

Key Concept Please read the sections titled "Messenger RNA Carries Information from DNA in a Three-Letter Genetic Code" (p. 117) and "Experiments with Synthetic mRNAs and Trinucleotides Broke the Genetic Code" (p. 119) and answer the following questions. Where relevant, refer to the following RNA sequences, which are coding-strand sequences derived from genes within the nuclear mammalian genome:

I. GUGCUCUUGCAAGACA

II. AGUAUUGUUGCAGGAUU

III. UAGUUCUUUUACAGGAC

1. Using Table 4-2, determine which of the above sequences has the potential to encode the following protein sequence: Val-Leu-Leu-Gln-Asp.

   a. I only.
   b. I and II only.
   c. I and III only.
   d. I, II, and III.

2. Which of the following occurs when the eukaryotic translational machinery encounters the UAG codon:

   a. The bound preinitiation complex stops scanning and positions the Met-tRNA$_i^{Met}$ at this site.
   b. The termination factors recognize this codon and translation ends.
   c. This codon is recognized by the corresponding anticodon of an empty tRNA molecule that is not linked to an amino acid.
   d. This codon is not recognized by any factors that ultimately cause the translational machinery to stop.

3. In which of the following would translation of sequence II yield a novel polypeptide compared with mammals:

   a. Rice.
   b. Acetabularia.
   c. Yeast.
   d. None of the above.

4. A naturally occurring coding strand composed of alternating C and U residues would result in the formation of

   a. A polypeptide containing alternating Leu and Ser residues.
   b. A polypeptide containing either Leu or Ser residues.
   b. A polypeptide containing only Leu residues.
   c. A polypeptide containing only Ser residues.

Key Experiment Please read the section titled "DNA Can Undergo Reversible Strand Separation" (p. 105) and use Figure 4-9 to answer the following questions:

5. An investigator would be able to distinguish a solution containing RNA from one containing DNA by

   a. Heating the solutions to 82.5 °C and measuring the absorption of light at 260 nm.
   b. Comparing the T_m of each solution.
   c. Monitoring the change in absorption of light at 260 nm while elevating the temperature.
   d. Measuring the absorption of light at 260 nm.

6. The percentage of G·C base pairs in a DNA molecule is related to the T_m because

a. The stability of G·C and A·T base pairs is intrinsically different.
b. A·T base pairs require a higher temperature for denaturation.
c. The triple bonds of G·C base pairs are less stable than the double bonds of A·T base pairs.
d. The G·C content equals the A·T content.

7. A circular DNA molecule will give rise to data as in Figure 4-9 only
a. If one of the strands is cut first.
b. If protein is bound to the DNA.
c. If replication is taking place.
d. If renaturation is allowed to take place.

Key Application Plant ribosome-inactivating proteins (RIPs) are toxins that inhibit eukaryotic translation. RIPs can be used to control viruses in both humans and plants, if they can be targeted specifically to virally infected cells. Catalytically, the RIPs act to remove an adenine residue from a conserved stem loop of the 60S rRNA, thereby preventing the binding of the eEF2-GTP complex during protein elongation. Please refer to the section titled "Eukaryotic Initiation of Protein Synthesis Occurs at the 5′ End and Internal Sites in mRNA" (p. 130) and answer the following questions:

8. Which of the following is not present in cells treated with an RIP:
a. The 43S preinitiation complex.
b. The ternary complex.
c. The 80S initiation complex.
d. The 40S initiation complex.

9. A mutant RIP, when added to eukaryotic cells, can be judged to be inactive by
a. Lack of depurination of rRNA.
b. Lack of union of the 40S and 60S ribosomal subunits.
c. Lack of viral growth.
d. Cytotoxicity.

10. RIPs do not inhibit prokaryotic translation, because
a. They do not recognize the 30S initiation complex.
b. They do not depurinate the 16S rRNA.
c. They do not recognize the Shine-Dalgarno sequence.
d. They do not recognize the large ribosomal subunit of prokaryotes.

11. Recognition of viral mRNAs by the eukaryotic machinery always involves
a. Initiation at the 5′ capped end of the mRNA.
b. Initiation at an internal ribosome entry site.
c. Protein factors that assist in ribosome scanning.
d. Initial binding of the 60S rRNA.

Key Terms

aminoacyl-tRNA *120*
anticodon *122*
codon *117*
complementary *103*
deoxyribonucleic acid (DNA) *100*
DNA polymerases *112*
double helix *103*
elongation factors *131*
exons *108*
genes *100*
genetic code *117*
growing fork *113*
initiation factors *129*
intron *108*
messenger RNA (mRNA) *100*
nucleosides *101*
nucleotides *101*
Okazaki fragments *113*
ribonucleic acid (RNA) *100*
ribosomal RNA (rRNA) *100*
ribosomes *116*
RNA polymerases *112*
RNA processing *115*
splicing *116*
termination factors *132*
transcription *100*
transfer RNA (tRNA) *100*
translation *100*

References

General References

Lehninger, A. L., D. L. Nelson, and M. M. Cox. 1993. *Principles of Biochemistry,* 2d ed. Worth.

The new age of RNA. 1993. *FASEB J.*, vol. 7.

Stryer, L. 1995. *Biochemistry,* 4th ed. W. H. Freeman and Company.

Nucleic Acids: Structure and General Properties

Brenner, S., F. Jacob, and M. Meselson. 1961. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* 190:576–581.

Cate, J. H., et al. 1996. Crystal structure of a Group I ribozyme domain: principles of RNA packing. *Science* 273:1678–1685.

Cech, T. R., and B. L. Bass. 1986. Biological catalysis by RNA. *Ann. Rev. Biochem.* 55:599–629.

Crothers, D. M., T. E. Haran, and J. G. Nadeau. 1990. Intrinsically bent DNA. *J. Biol. Chem.* 265:7093–7099.

Dickerson, R. E. 1983. The DNA helix and how it is read. *Sci. Am.* 249(6):94–111.

Felsenfeld, G. 1985. DNA. *Sci. Am.* 253(4):58–66.

Kornberg, A., and T. A. Baker. 1992. *DNA Replication,* 2d ed. W. H. Freeman and Company, chap. 1. A good summary of the principles of DNA structure.

Vasquez, K. M., and J. H. Wilson. 1998. Triplex-directed modification of genes and gene activity. *Trends Biochem. Sci.* 23:4–9.

Wang, J. C. 1980. Superhelical DNA. *Trends Biochem. Sci.* 5:219–221.

Zhang, B., and T. T. Cech. 1997. Peptide bond formation by in vitro selected ribozymes. *Nature* 390:96–100.

Nucleic Acid Synthesis

Doolittle, W. F. 1997. Archaea and the origins of DNA replication proteins. *Cell* 87:995–998.

Hampsey. M. 1998. Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol. Mol. Biol. Rev.* 62:465–503.

Mattaj, I. W., D. Tollervey, and B. Seraphin. 1993. Small nuclear RNAs in messenger RNA and ribosomal RNA processing. *FASEB J.* 7:47–53.

McClure, W. R. 1985. Mechanism and control of transcription initiation in prokaryotes. *Ann. Rev. Biochem.* 54:171–204.

Steitz, J. A. 1988. "Snurps." *Sci. Am.* 258(6):56–65.

Wahle, E., and W. Keller. 1992. The biochemistry of 3′ -end cleavage and polyadenylation of messenger RNA precursors. *Ann. Rev. Biochem.* 61:419–440.

Young, R. A. 1991. RNA polymerase II. *Ann. Rev. Biochem.* 60:689–715.

The Genetic Code

Khorana, H. G. 1968. Nucleic acid synthesis in the study of the genetic code. Reprinted in *Nobel Lectures: Physiology or Medicine* (1963–1970). Elsevier (1973).

Nirenberg, M. W., and J. H. Matthei. 1961. The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Nat'l. Acad. Sci. USA* 47:1588–1602.

Nirenberg, M. W., and P. Leder. 1964. RNA codewords and protein synthesis. *Science* 145:1399–1407.

Initiation

Jackson, R. J. 1991. Initiation without an end. *Nature* 353:14–15.

Kozak, M. 1991. An analysis of vertebrate mRNA sequences: intimations of translational control. *J. Cell Biol.* 115:887–903.

Le, S.-Y., and J. B. Maizel. 1997. A common RNA structural motif involved in the internal initiation of translation of cellular RNAs. *Nucl. Acids Res.* 35:173–369.

Sonenberg, N. 1988. Cap-binding proteins of eukaryotic messenger RNA: functions in initiation and control of translation. *Prog. Nucl. Acid Res. Mol. Biol.* 35:173–206.

Sonenberg, N. 1991. Picornavirus RNA translation continues to surprise. *Trends Genet.* 7:105–106.

Transfer RNA and Amino Acids

Arnez, J. G., and D. Moras. 1997. Structural and functional considerations of the aminoacylation reaction. *Trends Biochem. Sci.* 22:211–216.

Bjork, G. R., et al. 1987. Transfer RNA modification. *Ann. Rev. Biochem.* 56:263–287.

Cavarelli, J., and D. Moras. 1993. Recognition of tRNAs by aminoacyl-tRNA synthetases. *FASEB J.* 7:79–86.

Grosjean, H., J. Edqvist, K. B. Stråby, and R. Giegé. 1996. Enzymatic formation of modified nucleosides in tRNA: dependence on tRNA architecture. *J. Mol. Biol.* 255:67–85.

Hoagland, M. B., et al. 1958. A soluble ribonucleic acid intermediate in protein synthesis. *J. Biol. Chem.* 231:241–257.

Holley, R. W., et al. 1965. Structure of a ribonucleic acid. *Science* 147:1462–1465.

McClain, W. H. 1993. Transfer RNA identity. *FASEB J.* 7:72–78.

Rich, A., and S.-H. Kim. 1978. The three-dimensional structure of transfer RNA. *Sci. Am.* 240(1):52–62 (offprint 1377).

Rould, M. A., J. J. Perona, and T. A. Steitz. 1991. Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase. *Nature* 352:213–218.

Sissler, M., G. Eriani, F. Martin, R. Giegé, and C. Florentz. 1997. Mirror image alternative interaction patterns of the same tRNA with either class I arginyl-tRNA synthetase or class II aspartyl-tRNA synthetase. *Nucl. Acids Res.* 25:4899–4906.

Ribosomes

De Rijk, P., Y. Van de Peer, and R. De Wachter. 1997. Database on the structure of large ribosomal subunit RNA. *Nucl. Acids Res.* 25:117–122.

Frank, J. 1998. How the ribosome works. *Am. Scientist* 86:428–439.

Frank, J., et al. 1995. A model of protein synthesis based on cryo-electron microscopy of the *E. coli* ribosome. *Nature* 376:441–444.

Moazed, D., and H. F. Noller. 1991. Sites of interaction of the CCA end of peptidyl-tRNA with 23S rRNA. *Proc. Nat'l. Acad. Sci. USA* 88:3725–3728.

Nitta, I., et al. 1998 Reconstitution of peptide bond formation with *Escherichia coli* 23S ribosomal RNA domains. *Science* 281:666–669.

Noller, H. F. 1993. tRNA-rRNA interactions and peptidyl transferase. *FASEB J.* 7:87–89.

Olsen, G. J., and C. R. Woese. 1993. Ribosomal RNA: a key to phylogeny. *FASEB J.* 7:113–123.

Van de Peer, Y., J. Jansen, P. De Rijk, and R. De Wachter. 1997. Database on the structure of small ribosomal subunit. *Nucl. Acids Res.* 25:111–116.

Yonath, A., K. R. Leonard, and H. G. Wittman. 1987. A tunnel in the large ribosomal subunit revealed by three-dimensional reconstitution. *Science* 236:813–816.

The Steps in Protein Synthesis

Agrawal, R. K., et al. 1998. Visualization of elongation factor EF-G on the *E. coli* ribosome: the mechanism of translocation. *Proc. Nat'l Acad. Sci. USA* 95:6134–6138.

Freistroffer, D. V., et al. 1997. Release factor RF3 in *E. coli* accelerates the dissociation of release factors RF1 and RF2 from the ribosome in a GTP-dependent manner. *EMBO J.* 16:4126–4133.

Heurgué-Hamard, R., et al. 1998. Ribosome release factor RF4 and termination factor RF3 are involved in dissociation of peptidyl-tRNA from the ribosome. *EMBO J.* 17:808–816.

Jacob, W. F., M. Santer, and A. E. Dahlberg. 1987. A single base change in the Shine-Dalgarno region of 16S rRNA of *E. coli* affects translation of many proteins. *Proc. Nat'l. Acad. Sci. USA* 84:4757–4761.

Kozak, M. 1997. Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 or +6. *EMBO J.* 16:2482–2492.

Merrick, W. C., and J. W. B. Hershey. 1995. The pathway and mechanism of eukaryotic protein synthesis. In *Translational Control,* Cold Spring Harbor Laboratory Press.

Moazed, D., and F. Noller. 1989. Interaction of tRNA with 23S rRNA in the ribosomal A, P, and E sites. *Cell* 57:585–587.

Nakamura, Y., K. Ito, and L. A. Isaksson. 1996. Emerging understanding of translation termination. *Cell* 87:147–150.

Noller, H. F., V. Hoffarth, and L. Zimniak. 1992. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* 256:1416–1419.

Pavlov, M. Y., et al. 1997. Fast recycling of *Escherichia coli* ribosomes requires both ribosome recycling factor (RRF) and release factor RF3. *EMBO J.* 16:4134–4141.

Sachs, A. B., P. Sarnow, and M. W. Hentze. 1997. Starting at the beginning, middle, and end: translation initiation in eukaryotes. *Cell* 89:831–838.

Saks, M. E., J. R. Sampson, and N. N. Abelson. 1994. The transfer RNA identity problem: a search for rules. *Science* 263: 191–197.

Wells, S. E., et al. 1998. Circularization of mRNA by eukaryotic translation initiation factors. *Mol. Cell* 2:135–140.