Лекция 7.
Биосинтез белка

Information Pathways

The fourth and final part of this book considers biochemical questions raised by the genetic continuity and the evolution of living organisms. What is the molecular nature of the genetic material? How is genetic information transmitted with such fidelity? How is it ultimately translated into the amino acid sequence of protein molecules?

The fundamental unit of information in living systems is the **gene.** A gene is defined biochemically as that segment of DNA (or in a few cases RNA) that encodes the information required to produce a functional biological product. This product is most often a protein, and much of the material in the chapters to follow concerns genes that encode proteins. However, a gene product can also be one of several classes of RNA molecules. The storage and metabolism of these informational units now becomes the focal point of our discussion.

Modern biochemical research on gene structure and function has brought to biology a revolution comparable to that evoked over 100 years ago by Darwin's theory on the origin of species. An understanding of how information is stored and used in cells has brought penetrating new insights into some of the most fundamental problems concerning the structure and function of cells. Moreover, it has led to a more comprehensive conceptual framework for the science of biochemistry.

Today's knowledge of information pathways has arisen from the convergence of three different disciplines: genetics, physics, and biochemistry. The contributions of these three fields are epitomized by the discovery that opened the modern era of genetic biochemistry: the double-helical structure of DNA, as postulated by James Watson and Francis Crick in 1953 (see Fig. 12–15). Genetic theory contributed the concept of coding by genes. Physics made possible the determination of molecular structure by x-ray diffraction analysis. Biochemistry revealed the chemical composition of DNA. The great impact of the Watson–Crick hypothesis was largely due to its ability to account for a wide range of results derived from these varied sources.

A vastly improved understanding of DNA structure inevitably led to questions about its function. The structure itself suggested how DNA might be copied so that the information contained therein could be transmitted from one generation to the next. Understanding how the information in DNA was converted into functional proteins became possible through the discovery of messenger RNA and transfer RNA and the solution of the genetic code. These and other major advances led to the central dogma of molecular genetics, which defines three major processes in the cellular utilization of genetic information. The first is **replication,** the copying of parental DNA to form daughter

Facing page: The two β subunits of *E. coli* DNA polymerase III bound to DNA. The subunits, shown as gray ribbon structures, form a circle around the DNA, tethering the DNA polymerase III (which has at least 9 other subunits) to the DNA. This permits the enzyme to synthesize long stretches of DNA without dissociation. The complex set of operations by which macromolecules containing information are faithfully synthesized requires a great many enzymes, of which this is just part of one.

Replication   DNA

Transcription

RNA

Translation

Protein
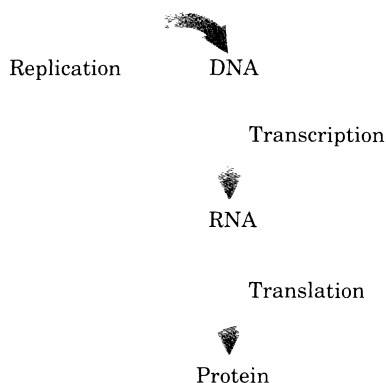
The central dogma of molecular genetics, showing the general pathways of information flow via the processes of replication, transcription, and translation. The term "dogma" is a misnomer here. It was introduced by Francis Crick at a time when little evidence supported these ideas. The "dogma" is now a well-established principle.

DNA molecules having identical nucleotide sequences. The second is **transcription,** the process by which parts of the coded genetic message in DNA are copied precisely in the form of RNA. The third is **translation,** in which the genetic message coded in messenger RNA is translated on the ribosomes into a protein with a specific sequence of amino acids.

Part IV is devoted to an explanation of these and related processes. First (Chapter 23) we will examine the structure, topology, and packaging of chromosomes and genes. The processes that make up the central dogma will be elaborated in Chapters 24 through 26. Then, as we have done for biosynthetic pathways, we will turn to regulation and examine how the expression of genetic information is controlled (Chapter 27).

A major theme running through these chapters is the added complexity encountered in the biosynthesis of a macromolecule when that macromolecule contains information. Assembling nucleic acids and proteins with the correct sequences of nucleotides and amino acids, respectively, represents nothing less than preserving the faithful expression of the template upon which life itself is based. The formation of phosphodiester bonds in DNA or peptide bonds in proteins might be expected to be a trivial feat for cells, given the arsenal of enzymatic and chemical tools described in Part III of this book. Nevertheless, the framework of patterns and rules established in the examination of metabolic pathways must be enlarged considerably when information is added to the equation. Forming *specific* bonds and preventing sequence errors in these polymers has an enormous impact on the thermodynamics, chemistry, and enzymology of the synthetic processes. For example, formation of a peptide bond should require an input of only about 21 kJ, and relatively simple enzymes that catalyze comparable reactions are known. To synthesize the correct peptide bond between two specific amino acids at a given point in a protein, however, the cell invests about 125 kJ in chemical energy and makes use of the combined activities of over 200 RNA molecules, enzymes, and specialized proteins. Information is expensive.

The dynamic interaction between nucleic acids and proteins is another central theme of Part IV. With the important exception of a few catalytic RNA molecules (discussed in Chapter 25), the processes that make up the pathways of cellular information flow are catalyzed and regulated by proteins. An understanding of these enzymes and proteins can have practical as well as intellectual rewards because they form the basis of the development of recombinant DNA technology. This technology is making possible the prenatal diagnosis of genetic disease; the production of a wide range of potent new pharmaceutical agents; the sequencing of the entire human genome; the introduction of new traits into bacteria, plants, and animals for industry and agriculture; human gene therapy; and many other advances. We finish our tour of the information pathways, and indeed the entire book, in Chapter 28 with a look at this technology and its implications for the future.

Protein Metabolism

Proteins are the end products of most information pathways. A typical cell requires thousands of different proteins at any given moment. These must be synthesized in response to the cell's current needs, transported (targeted) to the appropriate cellular location, and degraded when the need has passed. The protein synthesis pathway is much better understood than protein targeting or degradation, and coverage in this chapter reflects that fact.

Protein synthesis is the most complex of biosynthetic mechanisms, and understanding it has been one of the greatest challenges in the history of biochemistry. In eukaryotic cells, protein synthesis requires the participation of over 70 different ribosomal proteins; 20 or more enzymes to activate the amino acid precursors; a dozen or more auxiliary enzymes and other specific protein factors for the initiation, elongation, and termination of polypeptides; perhaps 100 additional enzymes for the final processing of different kinds of proteins; and 40 or more kinds of transfer and ribosomal RNAs. Thus almost 300 different macromolecules must cooperate to synthesize polypeptides. Many of these macromolecules are organized into the complex three-dimensional structure of the ribosome to carry out stepwise translocation of the mRNA as the polypeptide is assembled.

To appreciate the central importance of protein synthesis to every cell, it can be enlightening to consider the fraction of cellular resources that are devoted to this process. Protein synthesis can account for up to 90% of the chemical energy used by a cell for all biosynthetic reactions. In *E. coli*, the numbers of different types of proteins and RNA molecules involved in protein synthesis are similar to those in eukaryotic cells. Both prokaryotic and eukaryotic cells contain thousands of copies of each protein and RNA type per cell. When totaled, the 20,000 ribosomes, 100,000 related protein factors and enzymes, and 200,000 tRNAs present in a typical bacterial cell (with a volume of 100 nm^3) can account for more than 35% of the cell's dry weight.

Despite this great complexity, proteins are made at exceedingly high rates. A complete polypeptide chain of 100 residues is synthesized in an *E. coli* cell at 37 °C in about 5 s. The synthesis of the thousands of different proteins in each cell is tightly regulated so that only the required number of molecules of each is made under any given set of metabolic circumstances. To maintain the appropriate mix and concentration of proteins in a cell, the targeting and degradative processes must keep pace with synthesis. Research is gradually unraveling the
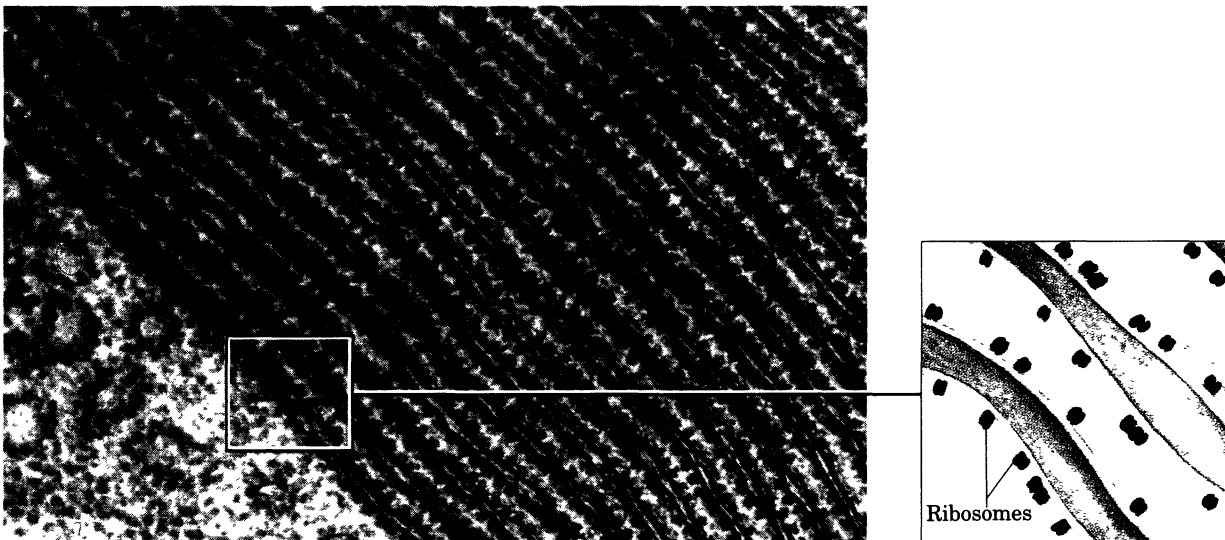
extraordinary set of biochemical processes that shepherd each protein to its proper location in the cell and selectively degrade proteins no longer required.

Protein Synthesis and the Genetic Code

Three major advances in the 1950s set the stage for our present knowledge of protein biosynthesis. In the early 1950s Paul Zamecnik and his colleagues designed a set of experiments to investigate the question: Where in the cell are proteins synthesized? They injected radioactive amino acids into rats, and at different time intervals after the injection the liver was removed, homogenized, and fractionated by centrifugation. The subcellular fractions were then examined for the presence of radioactive protein. When hours or days were allowed to elapse after injection of the labeled amino acids, *all* the subcellular fractions contained labeled proteins. However, when the liver was removed and fractionated only minutes after injection of the labeled amino acids, labeled protein was found only in a fraction containing small ribonucleoprotein particles. These particles, earlier discovered in animal tissues by electron microscopy, were thus identified as the site of protein synthesis from amino acids; later they were named ribosomes (Fig. 26–1).

Paul Zamecnik

Figure 26–1 Electron micrograph and schematic drawing of a portion of a pancreatic cell, showing ribosomes attached to the outer (cytosolic) face of the endoplasmic reticulum. The ribosomes are the numerous small dots bordering the parallel layers of membranes.

The second advance was made by Mahlon Hoagland and Zamecnik; they found that when incubated with ATP and the cytosolic fraction of liver cells, amino acids became "activated." The amino acids were attached to a special form of heat-stable soluble RNA, later called transfer RNA (tRNA), to form **aminoacyl-tRNAs.** The enzymes catalyzing this process are the **aminoacyl-tRNA synthetases.**

The third major advance occurred when Francis Crick asked: How is the genetic information that is coded in the 4-letter language of nucleic acids translated into the 20-letter language of proteins? Crick reasoned that tRNA must serve the role of an adapter, one part of the

Figure 26–2 Crick's hypothesis of the adapter function of tRNA. Today we know that the amino acid is covalently bound at the 3′ end of the tRNA and that a specific nucleotide triplet elsewhere in the tRNA molecule interacts with a specific triplet codon in the mRNA through hydrogen bonding of complementary bases.
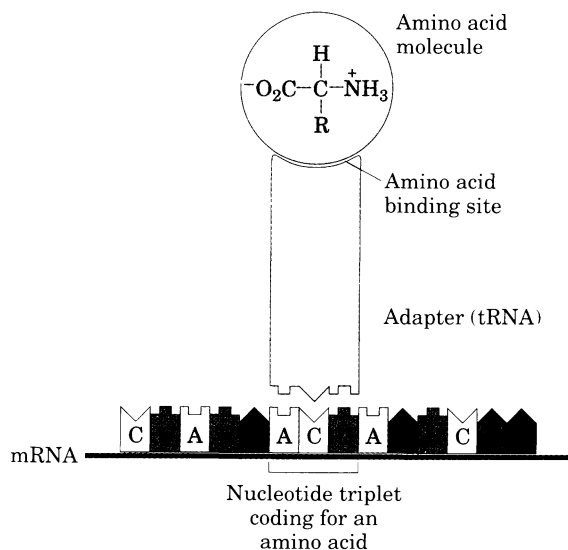
tRNA molecule binding a specific amino acid and some other part of the tRNA recognizing a short nucleotide sequence in the mRNA coding for that amino acid (Fig. 26–2). This idea was soon verified. The tRNA adapter "translates" the nucleotide sequence of an mRNA into the amino acid sequence of a polypeptide. The overall process of mRNA-guided protein synthesis is often referred to simply as **translation.**

These developments soon led to recognition of the major stages of protein synthesis and ultimately to the elucidation of the genetic code words for the amino acids. The nature of this code is the focus of the discussion that follows.
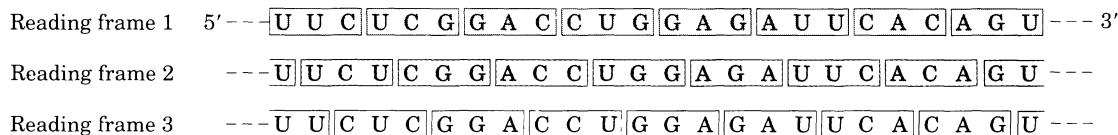
The Genetic Code Has Been Solved

By the 1960s it had long been apparent that at least three nucleotide residues of DNA are required to code for each amino acid. The four code letters of DNA (A, T, G, and C) in groups of two can yield only $4^2 = 16$ different combinations, not sufficient to code for 20 amino acids. But four bases in groups of three can yield $4^3 = 64$ different combinations. Early genetic experiments conclusively proved not only that the genetic code words or **codons** for amino acids are triplets of nucleotides but also that the codons do not overlap and there is no punctuation between codons for successive amino acid residues (Figs. 26–3, 26–4).

Figure 26–3 The triplet, nonoverlapping code. Evidence for the general nature of the genetic code came from many types of experiments, including genetic experiments on the effects of deletion and insertion mutations. Inserting or deleting one base pair (shown here in the mRNA transcript) alters the sequence of triplets in a nonoverlapping code, as shown, and all amino acids coded by the mRNA following the change are affected. Combining insertion and deletion mutations affects some amino acids but eventually restores the correct amino acid sequence. Adding or subtracting three nucleotides (not shown) leaves the remaining triplets intact, providing evidence that a codon has three, rather than four or five, nucleotides. The triplet codons shaded in gray are those transcribed from the original gene; codons shaded in blue are new codons resulting from the insertion or deletion mutations.
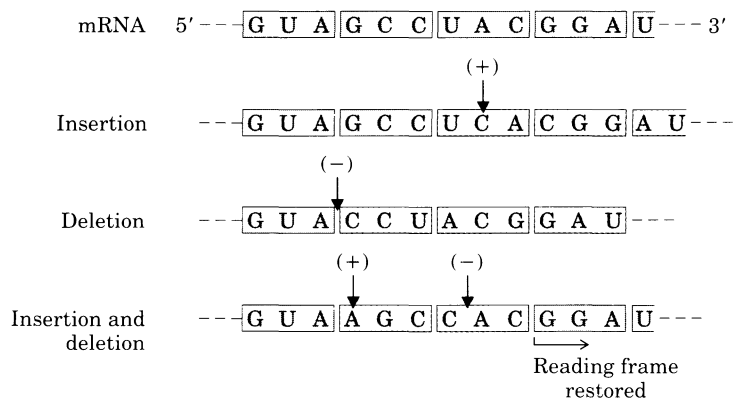
Nonoverlapping code

A U A C G A G U C
  1   2   3

Overlapping code

A U A C G A G U C
1

2

3

Figure 26–4 Overlapping versus nonoverlapping codes. In nonoverlapping codes, codons do not share nucleotides. In the example shown, the consecutive codons are numbered. In an overlapping code, some nucleotides in the mRNA are shared by different codons. A triplet code with maximum overlap, with consecutive codons defined by the numbered brackets, will have many nucleotides (such as the third nucleotide here) shared by three different codons. Note that in an overlapping code, the sequence of the first codon limits the possible sequences for the second codon. A nonoverlapping code provides much more flexibility in the sequence of neighboring codons and ultimately in the possible amino acid sequences designated by the code. The code used in all living systems is nonoverlapping.

The amino acid sequence of a protein is therefore defined by a linear sequence of contiguous triplet codons. The first codon in the sequence establishes a **reading frame,** in which a new codon begins every three nucleotide residues. In this scheme there are three possible reading frames for any given DNA sequence, and each will generally give a different sequence of codons (Fig. 26–5). Although it seemed clear that only one reading frame was likely to contain the information required for a given protein, the ultimate questions still loomed: What are the specific three-letter code words for the different amino acids? How could they be identified experimentally?

Reading frame 1 5′ - - - U U C U C G G A C C U G G A G A U U C A C A G U - - - 3′

Reading frame 2 - - - U U C U C G G A C C U G G A G A U U C A C A G U - - -

Reading frame 3 - - - U U C U C G G A C C U G G A G A U U C A C A G U - - -

Figure 26–5 In a triplet, nonoverlapping code, all mRNAs have three potential reading frames, shaded here in different colors. Note that the triplets, and hence the amino acids specified, are very different in each reading frame.

In 1961 Marshall Nirenberg and Heinrich Matthaei reported an observation that provided the first breakthrough. They incubated the synthetic polyribonucleotide polyuridylate (designated poly(U)) with an *E. coli* extract, GTP, and a mixture of the 20 amino acids in 20 different tubes. In each tube a different amino acid was radioactively labeled. Poly(U) can be regarded as an artificial mRNA containing many successive UUU triplets, and it should promote the synthesis of a polypeptide from only one of the 20 different amino acids—that coded by the triplet UUU. A radioactive polypeptide was formed in only one of the 20 tubes, that containing radioactive phenylalanine. Nirenberg and Matthaei therefore concluded that the triplet UUU codes for phenylalanine. The same approach revealed that the synthetic polyribonucleotide polycytidylate or poly(C) codes for formation of a polypeptide containing only proline (polyproline), and polyadenylate or poly(A) codes for polylysine. Thus the triplet CCC must code for proline and the triplet AAA for lysine.

The synthetic polynucleotides used in such experiments were made by the action of polynucleotide phosphorylase (p. 880), which catalyzes the formation of RNA polymers starting from ADP, UDP, CDP, and GDP. This enzyme requires no template and makes polymers with a base composition that directly reflects the relative concentrations of the nucleoside 5′-diphosphate precursors in the medium. If polynucleotide phosphorylase is presented with UDP, it makes only poly(U). If it is presented with a mixture of five parts of ADP and one of

Marshall Nirenberg

CDP, it will make a polymer in which about five-sixths of the residues are adenylate and one-sixth cytidylate. Such a random polymer is likely to have many triplets of the sequence AAA, lesser numbers of AAC, ACA, and CAA triplets, relatively few ACC, CCA, and CAC triplets, and very few CCC triplets (Table 26–1). With the use of different artificial mRNAs made by polynucleotide phosphorylase from different starting mixtures of ADP, GDP, UDP, and CDP, the base compositions of the triplets coding for almost all the amino acids were soon identified. However, these experiments could not reveal the *sequence* of the bases in each coding triplet.

H. Gobind Khorana

Table 26–1 Incorporation of amino acids into polypeptides in response to random polymers of RNA*

Amino acid	Observed frequency of incorporation (Lys = 100)	Tentative assignment for nucleotide composition† of corresponding codon	Expected frequency of incorporation based on assignment (Lys = 100)
Asparagine	24	(A)$_2$C	20
Glutamine	24	(A)$_2$C	20
Histidine	6	A(C)$_2$	4
Lysine	100	(A)$_3$	100
Proline	7	A(C)$_2$, (C)$_3$	4.8
Threonine	26	(A)$_2$C, A(C)$_2$	24

* Presented here is a summary of data from one of the early experiments designed to elucidate the genetic code. An RNA synthesized enzymatically, and containing only A and C residues in a 5:1 ratio, was used to direct polypeptide synthesis. Both the identity and quantity of amino acids incorporated were determined. Based upon the relative abundance of A and C residues in the synthetic RNA, and if the codon AAA (the most likely) is assigned a frequency of 100, there should be three different codons of composition (A)$_2$C, each at a relative frequency of 20; three codons of composition A(C)$_2$, each at a relative frequency of 4.0; and the codon CCC should occur at a relative frequency of 0.8. The CCC assignment here was based on information derived from prior studies with poly(C). Where two tentative codon assignments are made, both are proposed to code for the same amino acid.

† Note that these designations of nucleotide composition contain no information on nucleotide sequence.

In 1964 Nirenberg and Philip Leder achieved another breakthrough. They found that isolated *E. coli* ribosomes will bind a specific aminoacyl-tRNA if the corresponding synthetic polynucleotide messenger is present. For example, ribosomes incubated with poly(U) and phenylalanyl-tRNA^{Phe} (or Phe-tRNA^{Phe}) will bind both polymers, but if the ribosomes are incubated with poly(U) and some other aminoacyl-tRNA, the aminoacyl-tRNA will not be bound because it will not recognize the UUU triplets in poly(U) (Table 26–2). (Note that by convention, the identity of a tRNA is indicated by a superscript and an aminoacylated tRNA is indicated by a hyphenated name. For example, correctly aminoacylated tRNA^{Ala} is alanyl-tRNA^{Ala} or Ala-tRNA^{Ala}. If the tRNA is incorrectly aminoacylated, e.g., with valine, one would have Val-tRNA^{Ala}.) The shortest polynucleotide that could promote specific binding of Phe-tRNA^{Phe} was the trinucleotide UUU. By use of simple trinucleotides of known sequence it was possible to determine which aminoacyl-tRNA bound to each of about 50 of the 64 possible triplet codons. For some codons, either no aminoacyl-tRNAs would bind, or more than one were bound. Another method was needed to complete and confirm the entire genetic code.

Table 26–2 Experiment showing that trinucleotides are sufficient to induce specific binding of aminoacyl-tRNAs to ribosomes

| Trinucleotide | ^{14}C-Labeled aminoacyl-tRNA bound to ribosome* | | |
	Phe-tRNA^{Phe}	Lys-tRNA^{Lys}	Pro-tRNA^{Pro}
UUU	4.6	0	0
AAA	0	7.7	0
CCC	0	0	3.1

Source Modified from Nirenberg, M & Leder, P (1964) RNA code words and protein synthesis. *Science* **145**, 1399

* The numbers represent factors by which the amount of bound ^{14}C increased when the indicated trinucleotide was present, relative to controls in which no trinucleotide was added.

At about this time, a complementary approach was provided by H. Gobind Khorana, who developed methods to synthesize polyribonucleotides with defined, repeating sequences of two to four bases. The polypeptides produced using these RNAs as messengers had one or a few amino acids in repeating patterns. These patterns, when combined with information from the random polymers used by Nirenberg and colleagues, permitted unambiguous codon assignments. The copolymer $(AC)_n$, for example, has alternating ACA and CAC codons, regardless of the reading frame:

$$\lfloor A\ \ C\ \ A_{\rfloor\lfloor}C\ \ A\ \ C_{\rfloor\lfloor}A\ \ C\ \ A_{\rfloor\lfloor}C\ \ A\ \ C_{\rfloor\lfloor}A\ \ C\ \ A_{\rfloor}$$
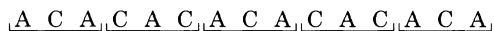
The polypeptide synthesized in response to this polymer was found to have equal amounts of threonine and histidine. Because the experiment described in Table 26–1 revealed a histidine codon with one A and two Cs, CAC must code for histidine and ACA for threonine.

Similarly, an RNA with three bases in a repeating pattern should yield three different types of polypeptide. Each polypeptide would be derived from a different reading frame and would contain a single kind of amino acid. An RNA with four bases in a repeating pattern should yield a single type of polypeptide with a repeating pattern of four amino acids (Table 26–3). Results from all of these experiments with polymers permitted the assignment of 61 of 64 possible codons. The other three were identified as termination codons, in part because they disrupted amino acid coding patterns when included in the sequence of a synthetic RNA polymer (Fig. 26–6; Table 26–3).

Table 26–3 Polypeptides produced in response to synthetic RNA polymers with repeating sequences of three and four bases

Polynucleotide	Polypeptide products
Trinucleotide repeats	
$(UUC)_n$	$(Phe)_n$, $(Ser)_n$, $(Leu)_n$
$(AAG)_n$	$(Lys)_n$, $(Arg)_n$, $(Glu)_n$
$(UUG)_n$	$(Leu)_n$, $(Cys)_n$, $(Val)_n$
$(CCA)_n$	$(Pro)_n$, $(His)_n$, $(Thr)_n$
$(GUA)_n$	$(Val)_n$, $(Ser)_n$, (chain terminator)*
$(UAC)_n$	$(Tyr)_n$, $(Thr)_n$, $(Leu)_n$
$(AUC)_n$	$(Ile)_n$, $(Ser)_n$, $(His)_n$
$(GAU)_n$	$(Asp)_n$, $(Met)_n$, (chain terminator)*
Tetranucleotide repeats	
$(UAUC)_n$	$(Tyr–Leu–Ser–Ile)_n$
$(UUAC)_n$	$(Leu–Leu–Thr–Tyr)_n$
$(GUAA)_n$	Di- and tripeptides*
$(AUAG)_n$	Di- and tripeptides*

* With these polynucleotides, the patterns of amino acid incorporation into polypeptides are affected by the presence of codons that are termination signals for protein biosynthesis In the repeating three-base sequences, one of the three reading frames includes only termination codons and thus only two homopolypeptides are observed (generated from the remaining two reading frames). In some of the repeating four-base sequences, every fourth codon is a termination codon in every reading frame, so that only short peptides are produced This is illustrated in Figure 26–6 for $(GUAA)_n$

Figure 26–6 The effect of a termination codon incorporated within a repeating tetranucleotide. Dipeptides or tripeptides will be synthesized, depending on where the ribosome initially binds. The three different reading frames are shown in different colors. Termination codons (indicated in red) are encountered every fourth codon in all three reading frames.

Reading frame 1   5′ --- G U A | A G U | A A G | U A A | G U A | A G U | A A --- 3′

Reading frame 2   --- G | U A A | G U A | A G U | A A G | U A A | G U A | A ---

Reading frame 3   --- G U | A A G | U A A | G U A | A G U | A A G | U A A ---

With these approaches the base sequences of all the triplet code words for each of the amino acids were established by 1966. Since then, these code words have been verified in many different ways. The complete codon "dictionary" for the amino acids is given in Figure 26–7. The cracking of the genetic code is regarded as the greatest scientific discovery of the 1960s.

Figure 26–7 The "dictionary" of amino acid code words as they occur in mRNAs. The codons are written in the 5'→3' direction. The third base of each codon, shown in bold type, plays a lesser role in specifying an amino acid than the first two. The three termination codons are shaded in red, and the initiation codon AUG is shaded in green. Note that all the amino acids except methionine and tryptophan have more than one codon. In most cases, codons that specify the same amino acid differ only in the third base.

Second letter of codon

First letter of codon (5' end)	U		C		A		G	
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
	UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop
	UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
	AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

The Genetic Code Has Several Important Characteristics

The key to the organization of the genetic information specifying a protein can be found in codons and in the array of codons that constitutes a reading frame. Keep in mind that no punctuation or signal is required to indicate the end of one codon and the beginning of the next. The reading frame must therefore be correctly set at the beginning of the readout of an mRNA molecule and then moved sequentially from one triplet to the next. If the initial reading frame is off by one or two bases, or if the ribosome accidentally skips a nucleotide in the mRNA, all the subsequent codons will be out of register and will lead to formation of a "missense" protein with a garbled amino acid sequence.

Several of the codons serve special functions. The **initiation codon,** AUG, signals the beginning of polypeptide chains. AUG not only is the initiation codon in both prokaryotes and eukaryotes but also codes for Met residues in internal positions of polypeptides. Of the 64 possible nucleotide triplets, three (UAA, UAG, and UGA) do not code for any known amino acids (Fig. 26–7); they are the **termination codons** (also called stop codons or nonsense codons), which normally signal the end of polypeptide chain synthesis. The three termination codons acquired the name "nonsense codons" because they were first found to result from single-base mutations in *E. coli* in which certain polypeptide chains are prematurely terminated. These **nonsense mutations,** arbitrarily named *amber, ochre,* and *opal,* respectively, helped make possible identification of UAA, UAG, and UGA as termination codons.

In a random sequence of nucleotides, one in every 20 codons in each reading frame, on average, will be a termination codon. Where a reading frame exists without a termination codon for 50 or more codons, the region is called an **open reading frame.** Long open reading frames usually correspond to genes that encode proteins. An uninterrupted gene coding for a typical protein with a molecular weight of 60,000 would require an open reading frame with 500 or more codons. See Box 26–1 (p. 900) for some interesting exceptions to this general pattern.

Perhaps the most striking feature of the genetic code is that it is **degenerate,** meaning that a given amino acid may be specified by more than one codon (Table 26–4). Only methionine and tryptophan have single codons. Degenerate does not mean imperfect; the genetic code is unambiguous because no codon specifies more than one amino acid. Note that the degeneracy of the code is not uniform. For example, leucine and serine have six codons, glycine and alanine have four, and glutamate, tyrosine, and histidine have two.

When an amino acid has multiple codons, the difference between the codons usually lies in the third base (at the 3′ end). For example, alanine is coded by the triplets GCU, GCC, GCA, and GCG. The codons for nearly all of the amino acids can be symbolized by XY^A_G or XY^U_C. The first two letters of each codon are therefore the primary determinants of specificity. This has some interesting consequences.

Table 26–4 Degeneracy of the genetic code

Amino acid	Number of codons	Amino acid	Number of codons
Ala	4	Leu	6
Arg	6	Lys	2
Asn	2	Met	1
Asp	2	Phe	2
Cys	2	Pro	4
Gln	2	Ser	6
Glu	2	Thr	4
Gly	4	Trp	1
His	2	Tyr	2
Ile	3	Val	4

Wobble Allows Some tRNAs to Recognize More than One Codon

Transfer RNAs recognize codons by base pairing between the mRNA codon and a three-base sequence on the tRNA called the **anticodon.** The two RNAs are paired antiparallel, the first base of the codon (always reading in the 5′→3′ direction) pairing with the third base of the anticodon (Fig. 26–8).

One might expect the anticodon triplet of a given tRNA to recognize only one codon triplet through Watson–Crick base pairing, so that there would be a different tRNA for each codon of an amino acid. However, the number of different tRNAs for each amino acid is *not* the same as the number of its codons. Moreover, some of the tRNAs contain the nucleotide inosinate (designated I), which contains the uncommon base hypoxanthine (see Fig. 12–5b). Molecular models show that inosinate can form hydrogen bonds with three different nucleotides, U, C, and A, but these pairings are rather weak compared with the strong hydrogen bonds between the Watson–Crick base pairs G≡C and A=U. In yeast, for example, one tRNA^Arg has the anticodon (5′)ICG, which can recognize three different arginine codons, (5′)CGA, (5′)CGU, and (5′)CGC. The first two bases of these codons are identical (CG) and form strong Watson–Crick base pairs (blue) with the corresponding bases of the anticodon:
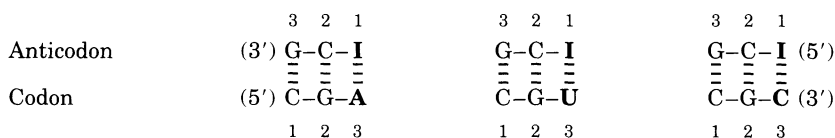
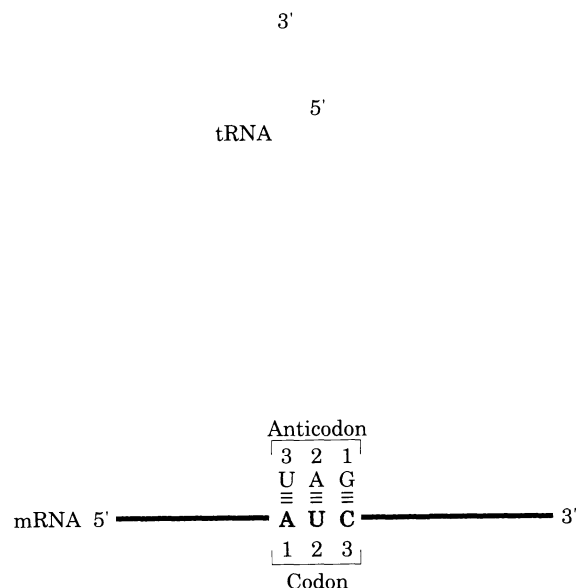|	3 2 1	3 2 1	3 2 1
Anticodon	(3′) G–C–**I**	G–C–**I**	G–C–**I** (5′)
Codon	(5′) C–G–**A**	C–G–**U**	C–G–**C** (3′)
	1 2 3	1 2 3	1 2 3

Figure 26–8 The pairing relationship of codon and anticodon. Alignment of the two RNAs is antiparallel. The tRNA is presented in the traditional cloverleaf configuration.

BOX 26-1

**Translational Frameshifting and RNA Editing:
mRNAs That Change Horses in Midstream**

Proteins are synthesized according to a pattern of contiguous triplet codons. Once the reading frame is set, codons are translated in order, without overlap or punctuation, until a termination codon is encountered. Usually, the other two possible reading frames within a gene contain no useful genetic information. However, a few genes are structured so that ribosomes "hiccup" at a certain point in the translation of the mRNA, leading to a change in the reading frame from that point on. In some cases this appears to be a mechanism used to produce two or more related proteins from a single transcript or to regulate the synthesis of a protein.

The best-documented example occurs in the translation of the mRNA for the *gag* and *pol* genes of the Rous sarcoma virus (see Fig. 25–31). The two genes overlap, with *pol* encoded by the reading frame in which each codon is offset to the left by one base pair (−1 reading frame) relative to *gag* (Fig. 1). The product of the *pol* gene (reverse transcriptase; p. 882) is translated initially as a larger *gag–pol* fusion protein using the same mRNA used for the *gag* protein alone. This fusion protein is later trimmed to the mature reverse transcriptase by proteolytic digestion. The large fusion protein is produced by a translational frameshift that occurs in the overlap region and allows the ribosome to bypass the UAG termination codon at the end of the *gag* gene (shown in red in Fig. 1). This frameshift occurs in about 5% of the translation events, so that the *gag–pol* fusion protein, and ultimately reverse transcriptase, is synthesized at the appropriate level for efficient replication of the viral genome—about 20-fold less than the *gag* protein. A similar mechanism is used to produce both the τ

and γ subunits of *E. coli* DNA polymerase III from *dna*X gene transcripts (see Table 24–2).

An example of the use of this mechanism for regulation occurs in the gene for *E. coli* release factor 2 (RF₂), a protein required for termination of protein synthesis at the termination codons UAA and UGA (described later in this chapter). The 26th codon of the gene for RF₂ is UGA, which would normally halt protein synthesis. The remainder of the gene is in the +1 reading frame (offset one base pair to the right) relative to this UGA codon. Low levels of RF₂ lead to a translational pause at this codon, because UGA is not recognized as a termination codon unless RF₂ binds to it. The absence of RF₂ prevents the termination of protein synthesis at this UGA and allows time for a frameshift so that UGA plus the C that follows it (UGAC) is read as GAC = Asp. Translation then proceeds in the new reading frame to complete synthesis of RF₂. In this way, RF₂ regulates its own synthesis in a feedback loop.

An especially unusual frameshifting mechanism occurs through the editing of mRNAs prior to translation. The genes in mitochondrial DNA that encode the cytochrome oxidase subunit II in some protists do not have open reading frames that correspond precisely to the protein product. Instead, the codons specifying the amino terminus of the protein are in a different reading frame from the codons specifying the carboxyl terminus. The problem is corrected not on the ribosome, but by a posttranscriptional editing process in which four uridines are added to create three new codons and shift the reading frame so that the entire gene can be translated directly, as shown in Figure 2a; the
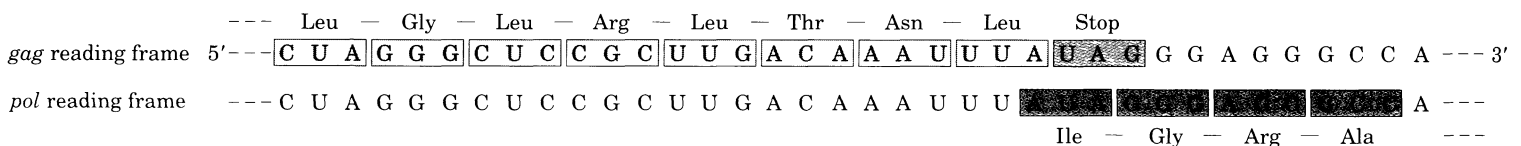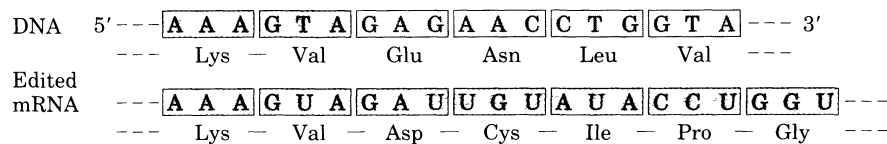
```
                 ---   Leu  —  Gly  —  Leu  —  Arg  —  Leu  —  Thr  —  Asn  —  Leu      Stop
gag reading frame  5'---|C U A||G G G||C U C||C G C||U U G||A C A||A A U||U U A||U A G| G G A G G G C C A --- 3'

pol reading frame  --- C U A G G G C U C C G C U U G A C A A A U U U ▮▮▮▮ ▮▮▮▮ ▮▮▮▮ ▮▮▮▮ A ---
                                                                    Ile  —  Gly  —  Arg  —  Ala       ---
```

Figure 1 The *gag–pol* overlap region in Rous sarcoma virus.

DNA   5'----|A A A||G T A||G A G||A A C||C T G||G T A|--- 3'
      ---   Lys — Val   Glu   Asn   Leu   Val   ---
Edited
mRNA ---|A A A||G U A||G A U||U G U||A U A||C C U||G G U|---
      ---   Lys — Val — Asp — Cys — Ile — Pro — Gly   ---
                                   └────→
                            (a)

mRNA 5'--- A A A G U A G A U U G U A U A C C U G G U--- 3'
                  •   •   •
              U U A U A U C U A A U A U A U G G A U A U

Guide RNA

      3'   5'
                            (b)

Figure 2 RNA editing of the transcript of the cytochrome oxidase subunit II gene from mitochondria of *Trypanosoma brucei*.
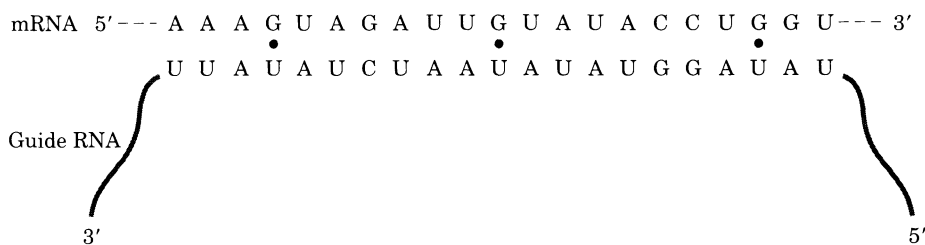
added uridine residues are shown in red. Only a small part of the gene (the region affected by editing) is shown. Neither the function nor mechanism of this editing process is understood. A special class of RNA molecules encoded by these mitochondria have been detected that have sequences complementary to the final, edited mRNAs. These appear to act as templates for the editing process and are referred to as guide RNAs (Fig. 2b). Note that the base pairing involves a number of G≡U base pairs (symbolized by blue dots), which are common in RNA molecules.

A distinct form of RNA editing occurs in the gene for the apolipoprotein B component of low-density lipoprotein in vertebrates. One form of apo-lipoprotein B, called apoB-100 (M_r 513,000), is synthesized in the liver. A second form, apoB-48 (M_r 250,000), is synthesized in the intestine. Both are synthesized from an mRNA produced from the gene for apoB-100. A cytosine deaminase enzyme found only in the intestine binds to the mRNA at codon 2,153 (CAA = Gln) and converts the C to a U to introduce the termination codon UAA at this position. The apoB-48 produced in the intestine from the modified mRNA is simply an abbreviated form (corresponding to the amino-terminal half) of apoB-100 (Fig. 3). This reaction permits the synthesis of two different proteins from one gene in a tissue-specific manner.
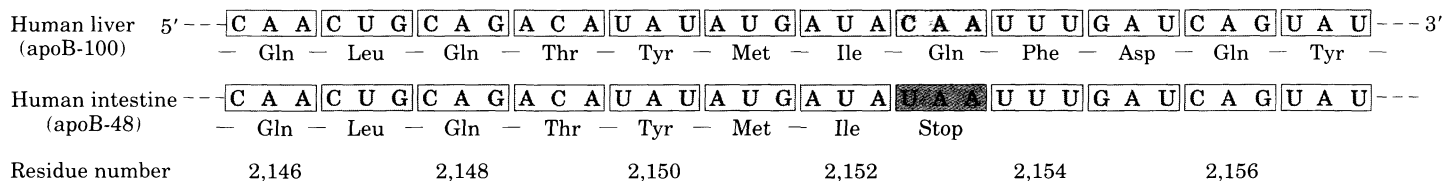
Human liver   5'---|C A A||C U G||C A G||A C A||U A U||A U G||A U A||C A A||U U U||G A U||C A G||U A U|---3'
(apoB-100)   — Gln — Leu — Gln — Thr — Tyr — Met — Ile — Gln — Phe — Asp — Gln — Tyr —

Human intestine ---|C A A||C U G||C A G||A C A||U A U||A U G||A U A||U A A||U U U||G A U||C A G||U A U|---
(apoB-48)   — Gln — Leu — Gln — Thr — Tyr — Met — Ile   Stop

Residue number   2,146   2,148   2,150   2,152   2,154   2,156

Figure 3 RNA editing of the transcript of the gene for the apolipoprotein B-100 component of low-density lipoprotein.

The third bases of the arginine codons (A, U, and C) form rather weak hydrogen bonds with the I residue at the first position of the anticodon. Examination of these and other codon–anticodon pairings led Crick to conclude that the third base of most codons pairs rather loosely with the corresponding base of its anticodons; to use his picturesque word, the third bases of such codons "wobble." Crick proposed a set of four relationships called the **wobble hypothesis:**

1. The first two bases of a codon in mRNA always form strong Watson–Crick base pairs with the corresponding bases of the anticodon in tRNA and confer most of the coding specificity.

2. The first base of some anticodons (reading in the 5'→3' direction; remember that this is paired with the third base of the codon) determines the number of codons read by a given tRNA. When the first base of the anticodon is C or A, binding is specific and only one codon is read by that tRNA. However, when the first base is U or G, binding is less specific and two different codons may be read. When inosinate (I) is the first, or wobble, nucleotide of an anticodon, three different codons can be read by that tRNA. This is the maximum number of codons that can be recognized by a tRNA. These relationships are summarized in Table 26–5.

3. When an amino acid is specified by several different codons, those codons that differ in either of the first two bases require different tRNAs.

4. A minimum of 32 tRNAs are required to translate all 61 codons.

Table 26–5 The wobble base of the anticodon determines how many codons of a given amino acid a tRNA can recognize

---

In the following, X and Y denote complementary bases capable of strong Watson–Crick base pairing with each other. The bases in the wobble or 3' position of the codons and 5' position of the anticodons are shaded in red.

1. One codon recognized:

Anticodon	(3') X–Y–**C** (5')	(3') X–Y–**A** (5')
Codon	(5') Y–X–**G** (3')	(5') Y–X–**U** (3')

2. Two codons recognized:

Anticodon	(3') X–Y–**U** (5')	(3') X–Y–**G** (5')
Codon	(5') Y–X–$^{\mathbf{A}}_{\mathbf{G}}$ (3')	(5') Y–X–$^{\mathbf{C}}_{\mathbf{U}}$ (3')

3. Three codons recognized:

Anticodon	(3') X–Y–**I** (5')
Codon	(5') Y–X–$^{\mathbf{A}}_{\mathbf{U}}{}_{\mathbf{C}}$ (3')

What can the reason be for this unexpected complexity of codon–anticodon interactions? In brief, the first two bases of a codon confer most of the codon–anticodon specificity. The wobble (or third) base of

the codon contributes to specificity, but because it pairs only loosely with its corresponding base in the anticodon, it permits rapid dissociation of the tRNA from its codon during protein synthesis. If all three bases of mRNA codons engaged in strong Watson–Crick pairing with the three bases of the tRNA anticodons, tRNAs would dissociate too slowly and severely limit the rate of protein synthesis. Codon–anticodon interactions optimize both accuracy *and* speed.

Overlapping Genes in Different Reading Frames Are Found in Some Viral DNAs

Although a given nucleotide sequence can, in principle, be read in any of its three reading frames, most DNA sequences encode a protein product in only one reading frame. In the coding frame there must be no termination codons, and each codon must correspond to the appropriate amino acid. As illustrated in Figure 26–9, the genetic code imposes strict limits on the numbers of amino acids that can be encoded by the codons of reading frame 2 without changing the amino acids specified by reading frame 1. Sometimes one amino acid (and its corresponding codon) may be substituted for another in reading frame 1 and still retain the function of the encoded protein, making it more likely that reading frames 2 or 3 might also encode a useful protein; but even taking these factors into account, the flexibility in other reading frames is very limited.

Figure 26–9 An amino acid sequence specified by one reading frame severely limits the potential amino acids encoded by any other reading frame. **(a)** The codons that can exist in reading frame 1 to produce the indicated amino acid sequence. Most of the permitted nucleotide changes (red) are in the third (wobble) position of each codon. **(b)** At the top are shown the codons that can exist in reading frame 2 without changing the amino acid sequence encoded by reading frame 1. Below are shown the alternative codons that correspond to the alternative mRNA sequences listed in **(a)**. The possible amino acids that can be encoded by reading frame 2 without changing the amino acid sequence encoded by reading frame 1 are in parentheses.
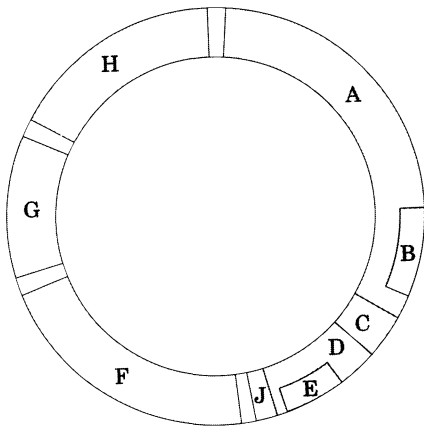

(a)

(b)

Figure 26–10 Genes within genes. The circular DNA of φX174 contains nine genes (A to J). Gene B lies within the sequence of gene A but uses a different reading frame. Similarly, gene E lies within gene D and also uses a different reading frame (see Fig. 26–11). The unshaded segments are untranslated spacer regions.
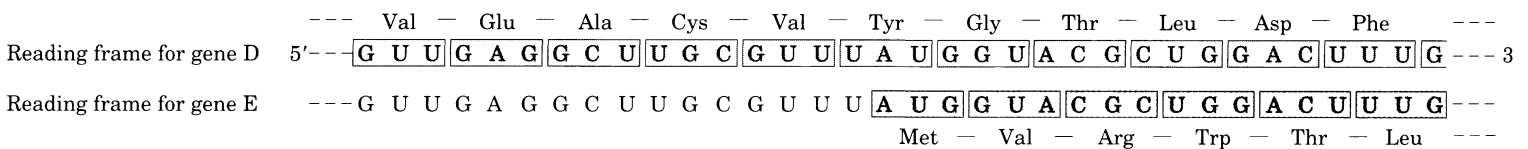
Although only one reading frame is generally used to encode a protein and genes do not overlap, there are a few interesting exceptions. In several viruses the same DNA base sequence codes for two different proteins by employing two different reading frames. The discovery of such "genes within genes" arose from the observation that the DNA of bacteriophage φX174, which contains 5,386 nucleotide residues, is not long enough to code for the nine different proteins that are known to be the products of the φX174 DNA genome, unless the genes overlap. The entire nucleotide sequence of the φX174 chromosome was compared with the amino acid sequences of the proteins encoded by the φX174 genes; this indicated several overlapping gene sequences. Figure 26–10 shows that genes B and E are nested within A and D, respectively. There are also five cases (not shown) in which the initiation codon of one gene overlaps the termination codon of the other gene. Figure 26–11 shows how genes D and E share a segment of DNA but use different reading frames; a similar situation exists for genes A and B. The sum of all the nested and overlapping sequences accounts completely for the surprisingly small size of the φX174 genome compared with the number of amino acid residues in the nine proteins for which it codes.

Figure 26–11 Portion of the nucleotide sequence of the mRNA transcript of gene D of φX174 DNA, showing how gene E, which is nested within gene D, is coded by a different reading frame from that used by gene D.

This discovery was quickly followed by similar observations in other viral DNAs, including those of phage λ, the cancer-causing simian virus 40 (SV40), RNA phages such as Qβ and Q17, and phage G4, a close relative of φX174. Phage G4 is remarkable in that at least one codon is shared by *three* different genes. It has been suggested that overlapping genes or genes within genes may be found only in viruses because the fixed, small size of the viral capsid requires economical use of a limited amount of DNA to code for the variety of proteins needed to infect a host cell and replicate within it. Also, because viruses reproduce (and therefore evolve) faster than their host cells, they may represent the ultimate in biological streamlining.

The genetic code is nearly universal. With the intriguing exception of a few minor variations that have been found in mitochondria, some bacteria, and some single-celled eukaryotes (Box 26–2, p. 906), amino acid codons are identical in all species that have been examined. Human beings, *E. coli,* tobacco plants, amphibians, and viruses share the same genetic code. Thus it would appear that all life forms had a common evolutionary ancestor with a single genetic code that has been very well preserved throughout the course of biological evolution.

The genetic code tells us how protein sequence information is stored in nucleic acids and provides some clues about how that information is translated into protein. We now turn to the molecular mechanisms of the translation process.

Protein Synthesis

As we have seen for DNA and RNA, the synthesis of polymeric biomole-cules can be separated into initiation, elongation, and termination stages. Protein synthesis is no exception. The activation of amino acid precursors prior to their incorporation into polypeptides and the post-translational processing of the completed polypeptide constitute two important and especially complex additional stages in the synthesis of proteins, and therefore require separate discussion. The cellular com-ponents required for each of the five stages in *E. coli* and other bacteria are listed in Table 26–6. The requirements in eukaryotic cells are quite similar. An overview of these stages will provide a useful outline for the discussion that follows.

Table 26–6 Components required for the five major stages in protein synthesis in *E. coli*

Stage	Necessary components
1. Activation of amino acids	20 amino acids 20 aminoacyl-tRNA synthetases 20 or more tRNAs ATP Mg^{2+}
2. Initiation	mRNA *N*-Formylmethionyl-tRNA Initiation codon in mRNA (AUG) 30S ribosomal subunit 50S ribosomal subunit Initiation factors (IF-1, IF-2, IF-3) GTP Mg^{2+}
3. Elongation	Functional 70S ribosome (initiation complex) Aminoacyl-tRNAs specified by codons Elongation factors (EF-Tu, EF-Ts, EF-G) Peptidyl transferase GTP Mg^{2+}
4. Termination and release	Termination codon in mRNA Polypeptide release factors (RF_1, RF_2, RF_3) ATP
5. Folding and processing	Specific enzymes and cofactors for removal of initiating residues and signal sequences, additional proteolytic processing, modification of terminal residues, attachment of phosphate, methyl, carboxyl, carbohydrate, or prosthetic groups

Stage 1: Activation of Amino Acids During this stage, which takes place in the cytosol, not on the ribosomes, each of the 20 amino acids is covalently attached to a specific tRNA at the expense of ATP energy. These reactions are catalyzed by a group of Mg^{2+}-dependent activating enzymes called aminoacyl-tRNA synthetases, each specific for one amino acid and its corresponding tRNAs. Where two or more tRNAs exist for a given amino acid, one aminoacyl-tRNA synthetase generally aminoacylates all of them. Aminoacylated tRNAs are commonly re-ferred to as being "charged."

BOX 26–2   **Natural Variations in the Genetic Code**

In biochemistry, as in other disciplines, exceptions to general rules can be problematic for educators and frustrating for students. At the same time they teach us that life is complex and inspire us to search for more surprises. Understanding the exceptions can even reinforce the original rule in surprising ways.

It would seem that there is little room for variation in the genetic code. Recall from Chapters 6 and 7 that even a single amino acid substitution can have profoundly deleterious effects on the structure of a protein. Suppose that somewhere there was a bacterial cell in which one of the codons specifying alanine suddenly began specifying arginine; the resulting substitution of arginine for alanine at multiple positions in scores of proteins would unquestionably be lethal. Variations in the code occur in some organisms nonetheless, and they are both interesting and instructive. The very rarity of these variations and the types of variations that occur together provide powerful evidence for a common evolutionary origin of all living things.

The mechanism for altering the code is straightforward: changes must occur in one or more tRNAs, with the obvious target for alterations being the anticodon. This will lead to the systematic insertion of an amino acid at a codon that does not specify that amino acid in the normal code (Fig. 26–7). The genetic code, in effect, is defined by the anticodons on tRNAs (which determine where an amino acid is placed in a growing polypeptide) and by the specificity of the enzymes—aminoacyl-tRNA synthetases—that charge the tRNAs (which determine the identity of the amino acid attached to a given tRNA).

Because of the catastrophic effects most sudden code changes would have on cellular proteins, one might predict that code alterations would occur only in cases where relatively few proteins would be affected. This could happen in small genomes encoding only a few proteins. The biological consequences of a code change could also be limited by restricting changes to the three termination codons, because these do not generally occur within genes (see Box 26–1 for exceptions to *this* rule). A change that converts a termination codon to a codon specifying an amino acid will affect termination in the products of only a subset of genes, and sometimes the effects in those genes will be minor because some genes have multiple (redundant) termination codons. This pattern is in fact observed.

Changes in the genetic code are very rare. Most of the characterized code variations occur in mitochondria, whose genomes encode only 10 to 20 proteins. Mitochondria have their own tRNAs, and the code variations do not affect the much larger cellular genomes. The most common changes in mitochondria, and the only changes observed in cellular genomes, involve termination codons.

In mitochondria, the changes can be viewed as a kind of genomic streamlining. Vertebrate mDNAs have genes that encode 13 proteins, 2 rRNAs, and 22 tRNAs (see Fig. 18–29). An unusual set of wobble rules allows the 22 tRNAs to decode all 64 possible codon triplets, rather than the 32 tRNAs required for the normal code. Four codon families (where the amino acid is determined entirely by the first two nucleotides) are decoded by a single tRNA with a U in the first (or wobble) position in the anticodon. Either the U pairs somehow with all four bases in the third position of the codon, or a "two out of three" mechanism is used in these cases (i.e., no pairing occurs at the third position of the codon). Other tRNAs recognize codons with either A or G in the third position, and yet others recognize U or C, so that virtually all the tRNAs recognize either two or four codons.

In the normal code, only two amino acids are specified by single codons, methionine and tryptophan (Table 26–4). If all mitochondrial tRNAs recognize two codons, then additional codons for Met and Trp might be expected in mitochondria. Hence, the single most common code variation observed is the UGA specification, from "termination" to Trp. A single tRNA^{Trp} can be used to recognize and insert a Trp residue at the codon UGA and the normal Trp codon UGG. Converting AUA from an Ile codon to a Met codon has a similar effect; the normal Met codon is AUG, and a single tRNA can be used for both codons. This turns out to be the second most common mitochondrial code variation. The known coding variations in mitochondria are summarized in Table 1.

Turning to the much rarer changes in the codes for cellular (as distinct from mitochondrial) genomes, we find that the only known variation in a prokaryote is again the use of UGA to encode Trp residues in the simplest free-living cell, *Mycoplasma capricolum*. In eukaryotes, the only known extramitochondrial coding changes occur in a few species of ciliated protists, where the termination codons UAA and UAG both specify glutamine.

Changes in the code need not be absolute—a codon need not always encode the same amino acid. In *E. coli* there are two examples of amino

Table 1 Known variant codon assignments in mitochondria

	Codons*				
	UGA	AUA	AGA AGG	CUN	CGG
Normal code assignment	Stop	Ile	Arg	Leu	Arg
Animals					
Vertebrates	Trp	Met	Stop	+	+
Drosophila	Trp	Met	Ser	+	+
Yeasts					
Saccharomyces cerevisiae	Trp	Met	+	Thr	+
Torulopsis glabrata	Trp	Met	+	Thr	?
Schizosaccharomyces pombe	Trp	+	+	+	+
Filamentous fungi	Trp	+	+	+	+
Trypanosomes	Trp	+	+	+	+
Higher plants	+	+	+	+	Trp
Chlamydomonas reinhardtii	?	+	+	+	?

~ ? Indicates that the codon has not been observed in the indicated mitochondrial genome, N, any nucleotide, +, the codon has the same meaning as in the normal code

acids being inserted at positions not specified in the general code. The first is the occasional use of the codon GUG (Val) as an initiating codon. This occurs only for those genes in which the GUG is properly located relative to special translation initiating signals in the mRNA (as discussed later in this chapter) that override the normal coding pattern. Thus, GUG has an altered coding specification only when it is positioned within a certain "context" of other sequences.

The use of contextual signals to alter coding patterns also applies to the second *E. coli* example. A few proteins in all cells (e.g., formate dehydrogenase in bacteria and glutathione peroxidase in mammals) require the element selenium for their activity. It is generally present in the form of the modified amino acid selenocysteine (Fig. 1). Modified amino acids are generally produced in post-translational reactions (described later in this chapter), but in *E. coli*, selenocysteine is introduced into formate dehydrogenase during translation in response to an in-frame UGA codon. A specialized type of serine tRNA, present at lower levels than other serine tRNAs, recognizes UGA and no other codons. This tRNA is charged with serine, and the serine is then enzymatically converted to selenocysteine prior to its use on the ribosome. The charged tRNA will not recognize just any UGA codon; instead some contextual signal in the mRNA, still to be identified, permits the tRNA to recognize only those few UGA codons that specify selenocysteine within certain genes. In effect, there are 21 standard amino acids in *E. coli*, and UGA doubles as a codon for termination and (sometimes) for selenocysteine.

These variations tell us that the code is not quite as universal as once believed, but they also tell us that flexibility in the code is severely constrained. It is clear that the variations are derivatives of the general code; no example of a completely different code has ever been found. The variants do not provide evidence for new forms of life, nor do they undermine the concepts of evolution or universality of the genetic code. The limited scope of code variants strengthens the principle that all life on this planet evolved on the basis of a single (very slightly flexible) genetic code.

$$COO^-$$
$$H_3\overset{+}{N}-CH$$
$$CH_2$$
$$Se$$
$$H$$
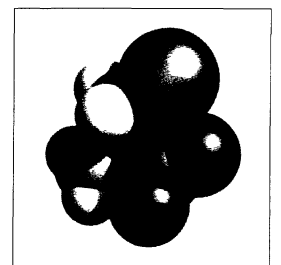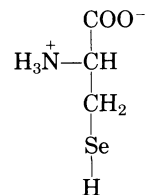
Figure 1 Selenocysteine.

Stage 2: Initiation Next, the mRNA bearing the code for the polypeptide to be made binds to the smaller of two major ribosomal subunits; this is followed by the binding of the initiating aminoacyl-tRNA and the large ribosomal subunit to form an initiation complex. The initiating aminoacyl-tRNA base-pairs with the mRNA codon AUG that signals the beginning of the polypeptide chain. This process, which requires GTP, is promoted by specific cytosolic proteins called initiation factors.

Stage 3: Elongation The polypeptide chain is now lengthened by covalent attachment of successive amino acid units, each carried to the ribosome and correctly positioned by its tRNA, which base-pairs to its corresponding codon in the mRNA. Elongation is promoted by cytosolic proteins called elongation factors. The binding of each incoming aminoacyl-tRNA and the movement of the ribosome along the mRNA are facilitated by the hydrolysis of two molecules of GTP for each residue added to the growing polypeptide.

Stage 4: Termination and Release The completion of the polypeptide chain is signaled by a termination codon in the mRNA. The polypeptide chain is then released from the ribosome, aided by proteins called release factors.

Stage 5: Folding and Processing In order to achieve its biologically active form the polypeptide must fold into its proper three-dimensional conformation. Before or after folding, the new polypeptide may undergo enzymatic processing to remove one or more amino acids from the amino terminus; to add acetyl, phosphate, methyl, carboxyl, or other groups to certain amino acid residues; to cleave the protein proteolytically; or to attach oligosaccharides or prosthetic groups.

In our expanded discussion of these stages a particular emphasis will be placed on stage 1. The reason is evident on considering the overall goal of the process: to synthesize a polypeptide chain with a defined sequence. To accomplish this task, two fundamental chemical requirements must be met: (1) the carboxyl group of each amino acid must be activated to facilitate formation of a peptide bond (see Fig. 5–15), and (2) a link must be maintained between each new amino acid and the information that encodes it in the mRNA. As we will see, both of these requirements are met by attaching the amino acid to a tRNA, and attaching the right amino acid to the right tRNA is therefore critical to the overall process of protein biosynthesis.

Before examining each stage in detail, we must introduce two key components in protein biosynthesis: the ribosome and tRNAs.

The Ribosome Is a Complex Molecular Machine

Each *E. coli* cell contains 15,000 or more ribosomes, which make up almost a quarter of the dry weight of the cell. Bacterial ribosomes contain about 65% rRNA and about 35% protein. They have a diameter of about 18 nm and a sedimentation coefficient of 70S.

Bacterial ribosomes consist of two subunits of unequal size (Fig. 26–12), the larger having a sedimentation coefficient of 50S and the smaller of 30S. The 50S subunit contains one molecule of 5S rRNA, one molecule of 23S rRNA, and 34 proteins. The 30S subunit contains one molecule of 16S rRNA and 21 proteins. The proteins are designated
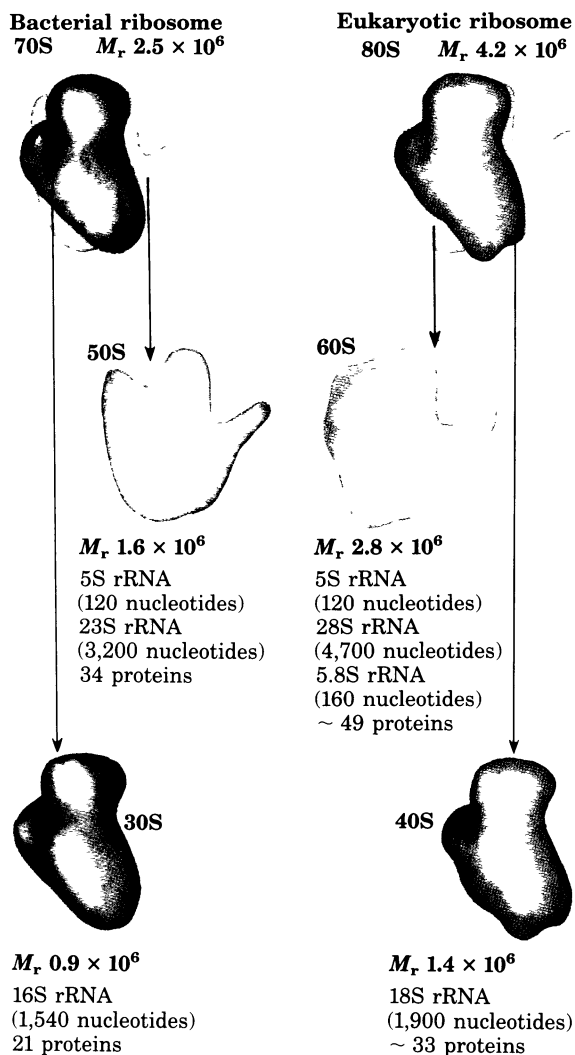

Bacterial ribosome
70S M_r 2.5 × 10⁶

50S

M_r 1.6 × 10⁶
5S rRNA
(120 nucleotides)
23S rRNA
(3,200 nucleotides)
34 proteins

30S

M_r 0.9 × 10⁶
16S rRNA
(1,540 nucleotides)
21 proteins

Eukaryotic ribosome
80S M_r 4.2 × 10⁶

60S

M_r 2.8 × 10⁶
5S rRNA
(120 nucleotides)
28S rRNA
(4,700 nucleotides)
5.8S rRNA
(160 nucleotides)
~ 49 proteins

40S

M_r 1.4 × 10⁶
18S rRNA
(1,900 nucleotides)
~ 33 proteins

Figure 26–12 Components of bacterial and eukaryotic ribosomes. The designation S (Svedberg units) refers to rates of sedimentation in the centrifuge. The S values (sedimentation coefficients) are not necessarily additive when subunits are combined.

Figure 26–13 Predicted folding patterns in *E. coli* 16S and 5S rRNAs, based on maximizing the potential intrastrand base pairing.

16S rRNA

5S rRNA

by numbers. Those in the large 50S subunit are numbered L1 to L34 (L for large) and those in the smaller subunit S1 to S21 (S for small). All the ribosomal proteins of *E. coli* have been isolated and many have been sequenced. Their variety is enormous, with molecular weights from about 6,000 to 75,000.

The sequences of nucleotides in the rRNAs of many organisms have been determined. Each of the three single-stranded rRNAs of *E. coli* has a specific three-dimensional conformation conferred by intrachain base pairing. Figure 26–13 shows a postulated representation of the 16S and 5S rRNAs in a maximally base-paired conformation. The rRNAs appear to serve as a framework to which the ribosomal proteins are bound.

In a method pioneered by Masayasu Nomura, the ribosome can be broken down into its RNA and protein components, then reconstituted in vitro. When the 21 different proteins and the 16S rRNA of the 30S subunit are isolated from *E. coli* and then mixed under appropriate experimental conditions, they spontaneously reassemble to form 30S subunits identical in structure and activity to native 30S subunits. Similarly, the 50S subunit can assemble itself from its 34 proteins and its 5S and 23S rRNAs, providing the 30S subunit is also present. Each of the 55 proteins in the bacterial ribosome is believed to play a role in the synthesis of polypeptides, either as an enzyme or as a structural component in the overall process. However, the detailed function of only a few of the ribosomal proteins is known.

The two ribosomal subunits have irregular shapes. The three-dimensional structures of the 30S and 50S subunits of *E. coli* ribosomes (Fig. 26–12) have been deduced from x-ray diffraction, electron microscopy, and other structural methods. The two oddly shaped subunits fit together in such a way that a cleft is formed through which the mRNA passes as the ribosome moves along it during the translation process and from which the newly formed polypeptide chain emerges (Fig. 26–14).

The ribosomes of eukaryotic cells (other than mitochondrial and chloroplast ribosomes) are substantially larger and more complex than bacterial ribosomes (Fig. 26–12). They have a diameter of about 23 nm and a sedimentation coefficient of about 80S. They also have two subunits, which vary in size between species but on average are 60S and 40S. The rRNAs and most of the proteins of eukaryotic ribosomes have also been isolated. The small subunit contains an 18S rRNA, and the large subunit contains 5S, 5.8S, and 28S rRNAs. Altogether, eukaryotic ribosomes contain over 80 different proteins. (In contrast, the ribosomes of mitochondria and chloroplasts are somewhat smaller and simpler than bacterial ribosomes.)

Masayasu Nomura

Figure 26–14 Two different views of models of an *E. coli* ribosome, showing the relationship between the 30S and 50S subunits. The arrow indicates the cleft between the subunits.
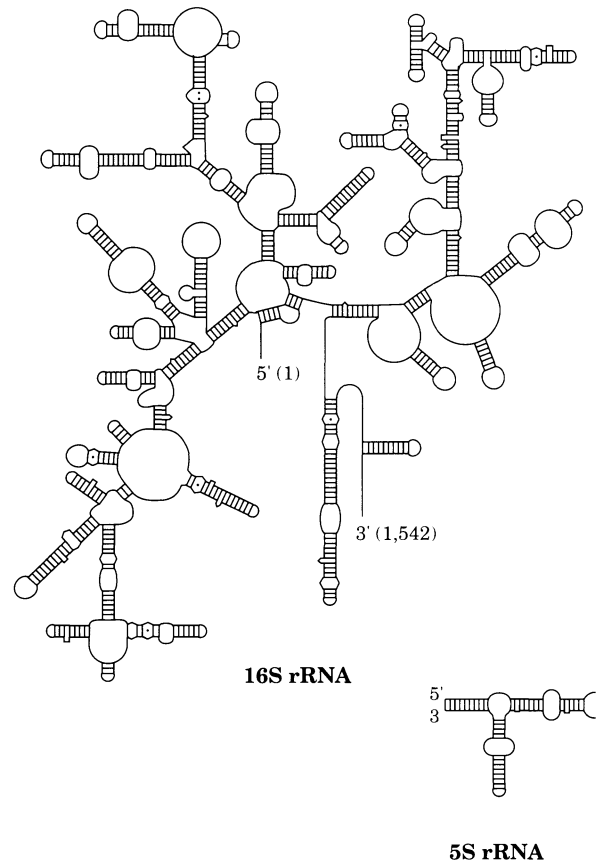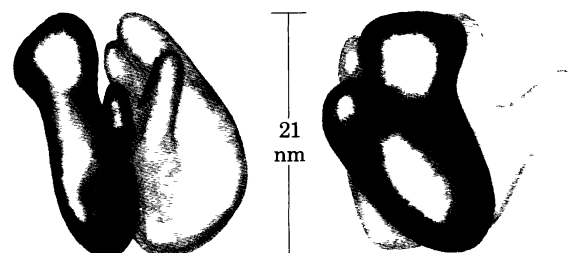
Robert W. Holley
1922–1993

Transfer RNAs Have Characteristic Structural Features

To understand how tRNAs can serve as adapters in translating the language of nucleic acids into the language of proteins, we must first examine their structure in more detail. As shown in Chapter 12, tRNAs are relatively small and consist of a single strand of RNA folded into a precise three-dimensional structure (see Fig. 12–27a). In bacteria and in the cytosol of eukaryotes, tRNAs have between 73 and 93 nucleotide residues, corresponding to molecular weights between 24,000 and 31,000. (Mitochondria contain distinctive tRNAs that are somewhat smaller.) As we have noted earlier in this chapter, there is at least one kind of tRNA for each amino acid; for some amino acids there are two or more specific tRNAs. At least 32 tRNAs are required to recognize all the amino acid codons (some recognize more than one codon), but some cells have many more than 32.

Many tRNAs have been isolated in homogeneous form. In 1965, after several years of work, Robert W. Holley and his colleagues worked out the complete nucleotide sequence of alanine tRNA (tRNA^{Ala}) from yeast. This, the very first nucleic acid to be sequenced in its entirety, was found to contain 76 nucleotide residues, ten of which have modified bases. Its complete base sequence is shown in Figure 26–15.

Since Holley's pioneering studies, the base sequences of many other tRNAs from various species have been worked out and have revealed many common denominators of structure. Eight or more of the nucleotide residues of all tRNAs have unusual modified bases, many of which are methylated derivatives of the principal bases. Most tRNAs have a guanylate (pG) residue at the 5' end, and all have the trinucleotide sequence CCA(3') at the 3' end. All tRNAs, if written in a form in which there is maximum intrachain base pairing through the allowed
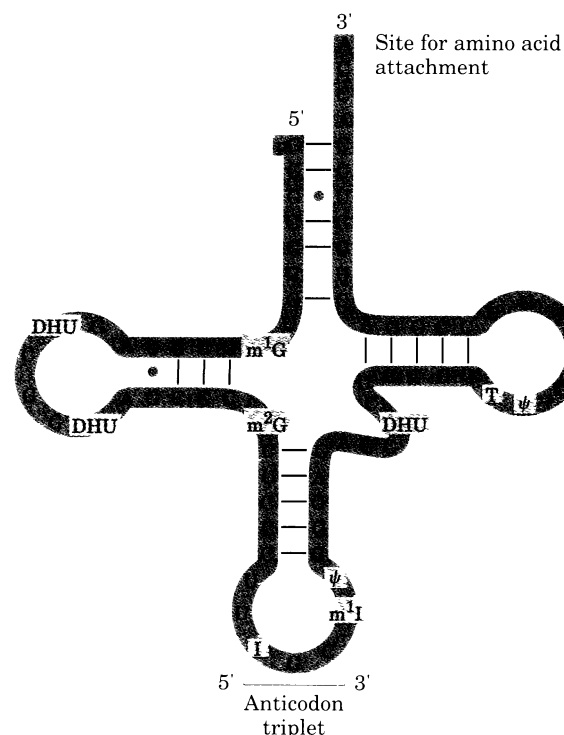
Figure 26–15 The nucleotide sequence of yeast tRNA^{Ala} as deduced by Holley and his colleagues. The cloverleaf conformation shown here is that in which intrastrand base-pairing is maximal. In addition to A, G, U, and C, the following symbols are used for the modified nucleotides: ψ, pseudouridine; I, inosine; T, ribothymidine; DHU, 5,6-dihydrouridine; m^{1}I, 1-methylinosine; m^{1}G, 1-methylguanosine; m^{2}G, N^2-dimethylguanosine. The modified bases are shaded in red, and most are illustrated in Fig. 25–25. The blue lines between the parallel sections indicate base pairs. The anticodon is capable of recognizing three codons for alanine (GCA, GCU, and GCC). Other features of tRNA structure are shown in Fig. 26–16. Note the presence of G=U base pairs in both the amino acid arm (top) and the DHU arm (left), signified by a blue dot to indicate a non–Watson–Crick pairing. In RNAs guanosine is often found base-paired with uridine, although the G=U pair is not as stable as the Watson–Crick G≡C pair (Chapter 12)

3'

Site for amino acid attachment

5'

DHU

m^{1}G

DHU

m^{2}G

DHU

T ψ

ψ

m^{1}I

I

5' ———— 3'
Anticodon
triplet

Figure 26–16 General structure of all tRNAs When drawn with maximum intrachain base pairing, all tRNAs show the cloverleaf structure. The large dots on the backbone represent nucleotide residues, and the blue lines represent base pairings. Characteristic and/or invariant residues common to all tRNAs are shaded in red. Transfer RNAs differ in length, from 73 to 93 nucleotides. Extra nucleotides occur in the extra arm or in the DHU arm. At the end of the anticodon arm is the anticodon loop, which always contains seven unpaired nucleotides. The DHU arm contains up to three DHU residues, depending on the tRNA. In some tRNAs the DHU arm has only three hydrogen-bonded base pairs In addition to the symbols explained in Fig. 26–15: Pu, purine nucleotide; Py, pyrimidine nucleotide; G$^\times$, guanylate or 2'-O-methylguanylate.

pairs A=U, G≡C, and G=U (see Fig. 12–26), form a cloverleaflike structure with four arms; the longer tRNAs have a short fifth or extra arm (Fig. 26–16; also evident in Fig. 26–15). The actual three-dimensional structure of a tRNA looks more like a twisted L than a cloverleaf (Fig. 26–17).

Two of the arms of a tRNA are critical for the adapter function. The **amino acid** or **AA arm** carries a specific amino acid esterified by its carboxyl group to the 2'- or 3'-hydroxyl group of the adenosine residue at the 3' end of the tRNA. The **anticodon arm** contains the anticodon. The other major arms are the **DHU** or **dihydrouridine arm,** which contains the unusual nucleotide dihydrouridine, and the **TψC arm,** which contains ribothymidine (T), not usually present in RNAs, and pseudouridine (ψ), which has an unusual carbon–carbon bond between the base and pentose (see Fig. 25–25). The functions of the DHU and TψC arms have not yet been determined.
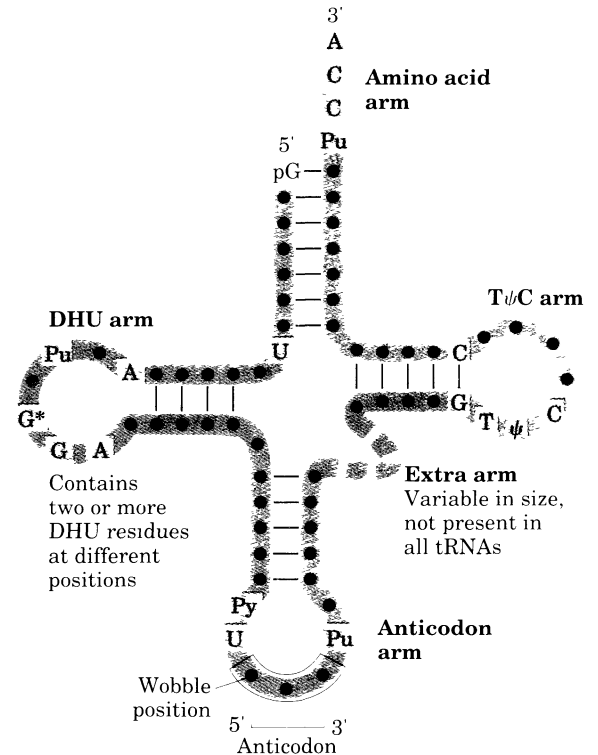

Figure 26–17 The three-dimensional structure of yeast tRNA^{Phe} deduced from x-ray diffraction analysis. It resembles a twisted L. **(a)** A schematic with the various arms identified in Fig. 26–16 shaded in different colors **(b)** A space-filling model. Color coding is the same in both representations. The three bases of the anticodon are shown in red and the CCA sequence at the 3' end (the attachment point for amino acids) is shown in orange. The TψC and DHU arms are blue and yellow, respectively
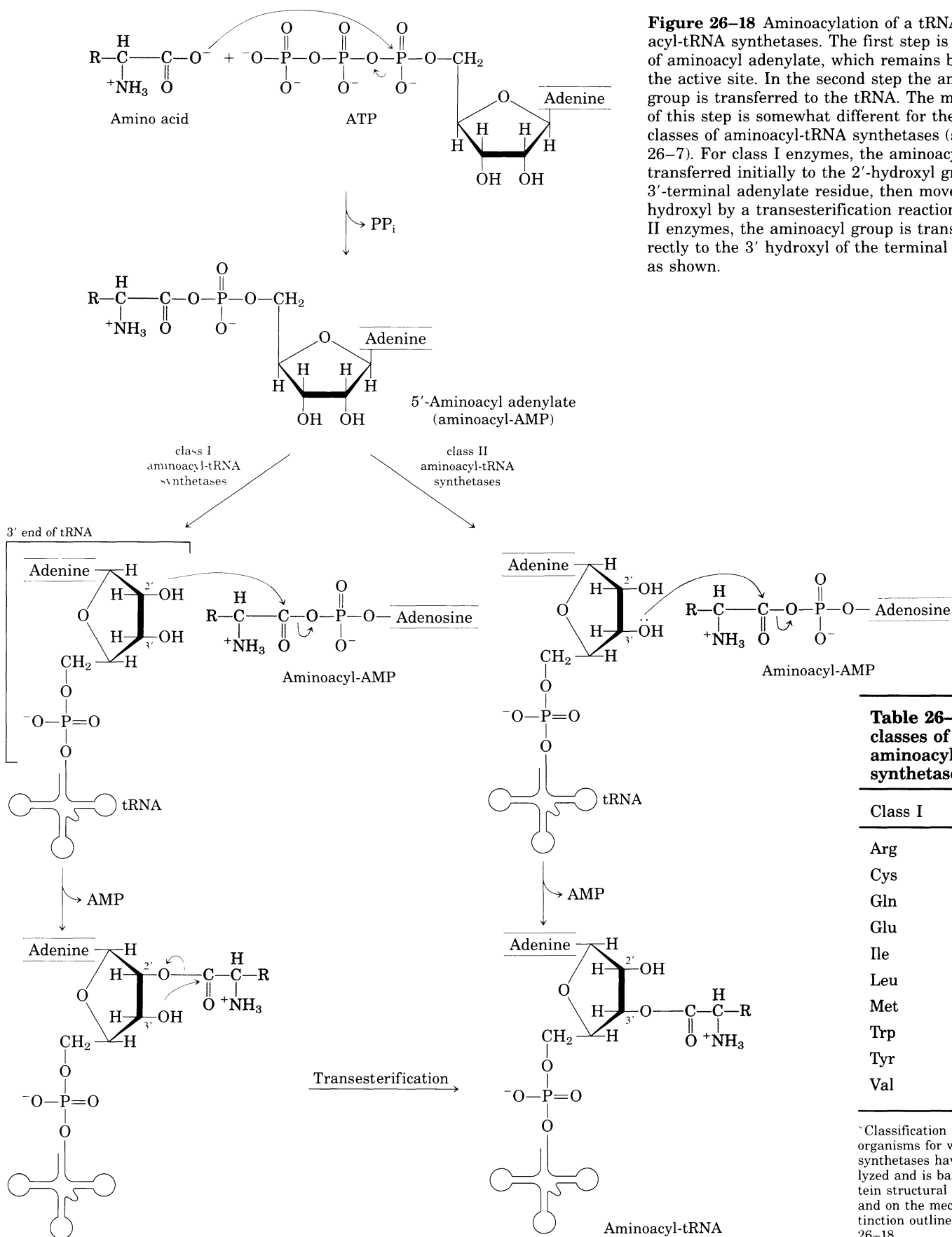

(a)

(b)
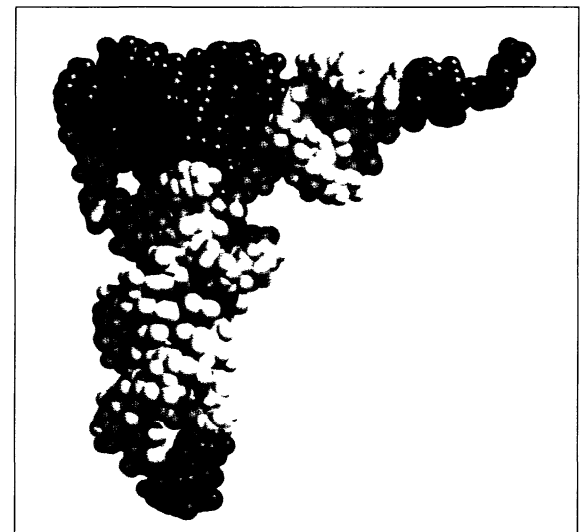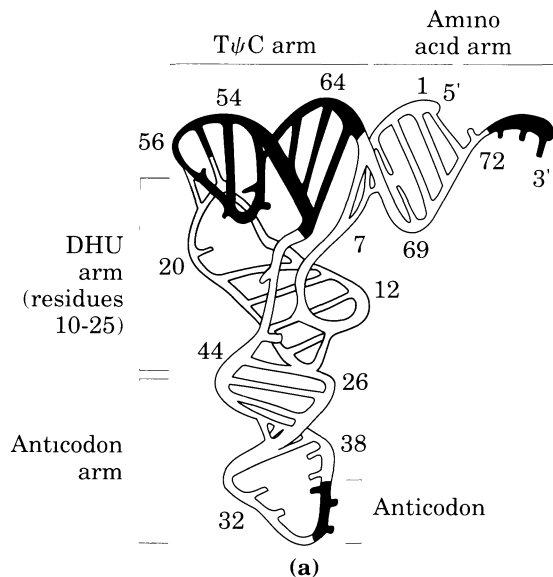
Figure 26–18 Aminoacylation of a tRNA by amino-acyl-tRNA synthetases. The first step is formation of aminoacyl adenylate, which remains bound to the active site. In the second step the aminoacyl group is transferred to the tRNA. The mechanism of this step is somewhat different for the two classes of aminoacyl-tRNA synthetases (see Table 26–7). For class I enzymes, the aminoacyl group is transferred initially to the 2'-hydroxyl group of the 3'-terminal adenylate residue, then moved to the 3' hydroxyl by a transesterification reaction. For class II enzymes, the aminoacyl group is transferred directly to the 3' hydroxyl of the terminal adenylate, as shown.

Table 26–7 Two classes of aminoacyl-tRNA synthetases*

Class I	Class II
Arg	Ala
Cys	Asn
Gln	Asp
Glu	Gly
Ile	His
Leu	Lys
Met	Phe
Trp	Pro
Tyr	Ser
Val	Thr

*Classification applies to all organisms for which tRNA synthetases have been analyzed and is based on protein structural distinctions and on the mechanistic distinction outlined in Figure 26–18
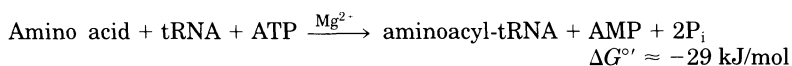
Aminoacyl-tRNA Synthetases Attach the Correct Amino Acids to Their tRNAs

In the first stage of protein synthesis, which takes place in the cytosol, the 20 different amino acids are esterified to their corresponding tRNAs by aminoacyl-tRNA synthetases, each of which is specific for one amino acid and one or more corresponding tRNA. In most organisms there is generally one aminoacyl-tRNA synthetase for each amino acid. As noted earlier, for amino acids that have two or more corresponding tRNAs the same aminoacyl-tRNA synthetase usually aminoacylates all of them. In *E. coli,* the only exception to this rule is lysine, for which there are two aminoacyl-tRNA synthetases. There is only one tRNA^Lys in *E. coli,* and the biological rationale for the presence of two Lys-tRNA synthetases is unclear. Nearly all the aminoacyl-tRNA synthetases of *E. coli* have been isolated; all have been sequenced (either the protein itself or its gene), and a number have been crystallized. They have been divided into two classes (Table 26–7) based on distinctions in primary and tertiary structure and on differences in reaction mechanism, as detailed below. The overall reaction catalyzed by these enzymes is

$$\text{Amino acid} + \text{tRNA} + \text{ATP} \xrightarrow{\text{Mg}^{2-}} \text{aminoacyl-tRNA} + \text{AMP} + \text{PP}_i$$
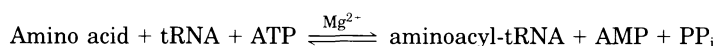
The activation reaction occurs in two separate steps in the enzyme active site. In the first step, an enzyme-bound intermediate, aminoacyl adenylate (aminoacyl-AMP) is formed by reaction of ATP and the amino acid at the active site (Fig. 26–18). In this reaction, the carboxyl group of the amino acid is bound in anhydride linkage with the 5'-phosphate group of the AMP, with displacement of pyrophosphate.

In the second step the aminoacyl group is transferred from enzyme-bound aminoacyl-AMP to its corresponding specific tRNA. As shown in Figure 26–18, the course of this second step depends upon the class to which the enzyme belongs (Table 26–7). The reason for the mechanistic distinction between the two enzyme classes is unknown. The resulting ester linkage between the amino acid and the tRNA (Fig. 26–19) has a high standard free energy of hydrolysis ($\Delta G^{\circ\prime} =$ -29 kJ/mol). The pyrophosphate formed in the activation reaction undergoes hydrolysis to phosphate by inorganic pyrophosphatase. Thus *two* high-energy phosphate bonds are ultimately expended for each amino acid molecule activated, rendering the overall reaction for amino acid activation essentially irreversible:

$$\text{Amino acid} + \text{tRNA} + \text{ATP} \xrightarrow{\text{Mg}^{2-}} \text{aminoacyl-tRNA} + \text{AMP} + 2\text{P}_i$$
$$\Delta G^{\circ\prime} \approx -29 \text{ kJ/mol}$$

Figure 26–19 General structure of aminoacyl-tRNAs. The aminoacyl group is esterified to the 3' position of the terminal adenylate residue. The ester linkage that both activates the amino acid and joins it to the tRNA is shaded red.

Some Aminoacyl-tRNA Synthetases Are Capable of Proofreading

The aminoacylation of tRNA accomplishes two things: the activation of an amino acid for peptide bond formation and attachment of the amino acid to an adapter tRNA that directs its placement within a growing polypeptide. As we will see, the identity of the amino acid attached to a tRNA is not checked on the ribosome. Attaching the correct amino acid to each tRNA is therefore essential to the fidelity of protein synthesis as a whole.
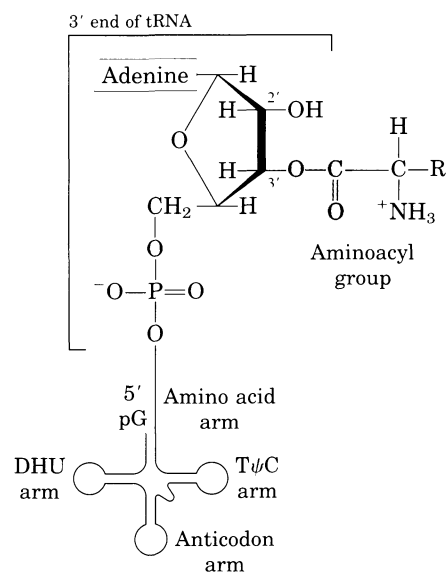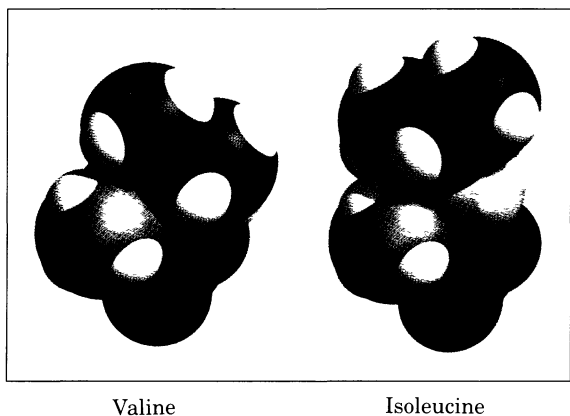
Valine   Isoleucine

The potential for any enzyme to discriminate between two different substrates is limited by the available binding energy that can be derived from enzyme–substrate interactions (Chapter 8). Discrimination between two similar amino acid substrates has been studied in detail in the case of Ile-tRNA^{Ile} synthetase, which faces the molecular problem that valine differs from isoleucine only by one methylene (CH_2) group. For this enzyme, activation of isoleucine (to form Ile-AMP) is favored over valine by a factor of 200, in the range expected given the potential contribution of binding energy from a methylene group. However, valine is incorporated into proteins in positions normally occupied by isoleucine at a frequency of only about 1 in 3,000.

The difference is brought about by a separate proofreading function of Ile-tRNA synthetase; this function is also present in some other aminoacyl-tRNA synthetases. All aminoacyl-AMPs produced by Ile-tRNA synthetase are checked in a second active site on the same enzyme, and incorrect ones are hydrolyzed. This proofreading activity reflects a general principle already seen in the discussion of proofreading by DNA polymerases (p. 822). If available binding interactions involving different groups on two substrates do not provide for a sufficient discrimination between the two on the enzyme, then this available binding energy must be used *twice* (or more) in separate steps requiring discrimination. Forcing the system through two successive "filters" rather than one increases the potential fidelity by a power of 2. In the case of Ile-tRNA synthetase, the first filter is the initial amino acid binding and activation to aminoacyl-AMP. The second filter is the separate active site, which catalyzes deacylation of incorrect aminoacyl-AMPs. The aminoacyl-AMP intermediates remain bound to the enzyme. When tRNA^{Ile} binds to the enzyme, the presence of Ile-AMP leads to aminoacylation of the tRNA. If Val-AMP is present on the enzyme instead, it is hydrolyzed to valine and AMP and the tRNA is not aminoacylated. Because the R group of valine is slightly smaller than that of isoleucine, the Val-AMP fits the hydrolytic (proofreading) site of the Ile-tRNA synthetase, but Ile-AMP does not.

In addition to proofreading after formation of the aminoacyl-AMP intermediate, most aminoacyl-tRNA synthetases are also capable of hydrolyzing the ester linkage between amino acids and tRNAs in aminoacyl-tRNAs. This hydrolysis is greatly accelerated for incorrectly charged tRNAs, providing yet a third filter to enhance the fidelity of the overall process. In contrast, in a few aminoacyl-tRNA synthetases that activate amino acids that have no close structural relatives, little or no proofreading occurs; in these cases the active site can sufficiently discriminate between the proper substrate amino acid and incorrect amino acids.

The overall error rate of protein synthesis ($\sim$1 mistake per 10^4 amino acids incorporated) is not nearly as low as for DNA replication, perhaps because a mistake in a protein is erased by destroying the protein and is not passed on to future generations. This degree of fidelity is sufficient to ensure that most proteins contain no mistakes and that the large amount of energy required to synthesize a protein is rarely wasted.

The Interaction between Aminoacyl-tRNA Synthetase and tRNA Constitutes a "Second Genetic Code"

An individual aminoacyl-tRNA synthetase must be specific not only for a single amino acid but for a certain tRNA as well. Discriminating

among several dozen tRNAs is just as important for the overall fidelity of protein biosynthesis as is distinguishing among amino acids. The interaction between aminoacyl-tRNA synthetases and tRNAs has been referred to as the "second genetic code," to reflect its critical role in maintaining the accuracy of protein synthesis. The "coding" rules are apparently more complex than those in the "first" code.

Figure 26–20 summarizes what is known about the nucleotides involved in recognition by some or all aminoacyl-tRNA synthetases. Some nucleotides are conserved in all tRNAs and therefore cannot be used for discrimination. Nucleotide positions that are involved in discrimination by the aminoacyl-tRNA synthetases have been identified by the fact that changes at those nucleotides alter the enzyme's substrate specificity. These interactions seem to be concentrated in the amino acid arm and the anticodon arm, but are also located in many other parts of the molecule. The conformation of the tRNA (as opposed to its sequence) can also be important in recognition.

Some aminoacyl-tRNA synthetases recognize the tRNA anticodon itself. Changing the anticodon of one tRNA^{Val} from UAC to CAU makes this tRNA an excellent substrate for Met-tRNA synthetase. The Val-tRNA synthetase will similarly recognize a modified tRNA^{Met} in which the anticodon has been changed to UAC. Recognition by aminoacyl-tRNA synthetases of other tRNAs (about half of them, including those for alanine and serine) is affected little or not at all by changes at the anticodon. In some cases ten or more specific nucleotides are involved in recognition of a tRNA by its specific aminoacyl-tRNA synthetase (Fig. 26–20). In contrast, across a range of organisms from bacteria to humans the primary determinant for tRNA recognition by the Ala-tRNA synthetases is a single G=U base pair in the amino acid arm of tRNA^{Ala} (Fig. 26–21a). A short RNA with as few as seven base pairs arranged in a simple hairpin minihelix is efficiently aminoacylated by the Ala-tRNA synthetase as long as the RNA contains this critical G=U (Fig. 26–21b).
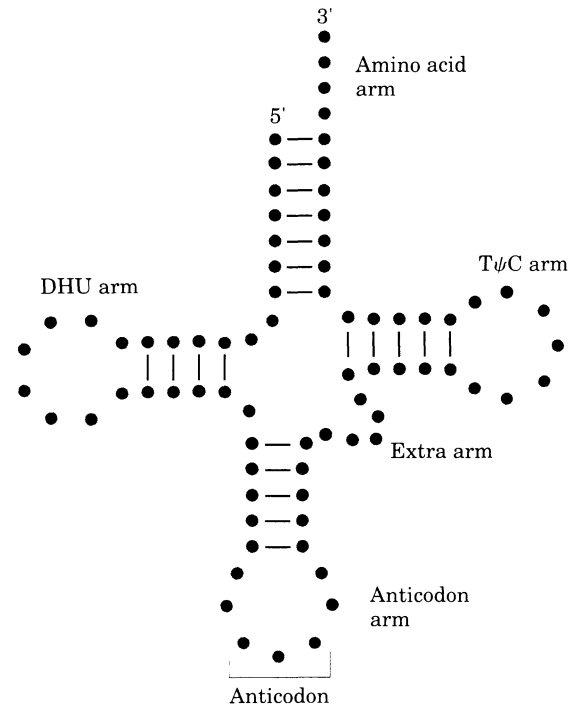

Figure 26–20 Known positions in tRNAs recognized by aminoacyl-tRNA synthetases. Positions in blue are the same in all tRNAs and therefore cannot be used to discriminate one from another. Other positions are known recognition points for one (red) or more (green) tRNA synthetases. Structural features other than sequence are important for recognition by some tRNA synthetases.
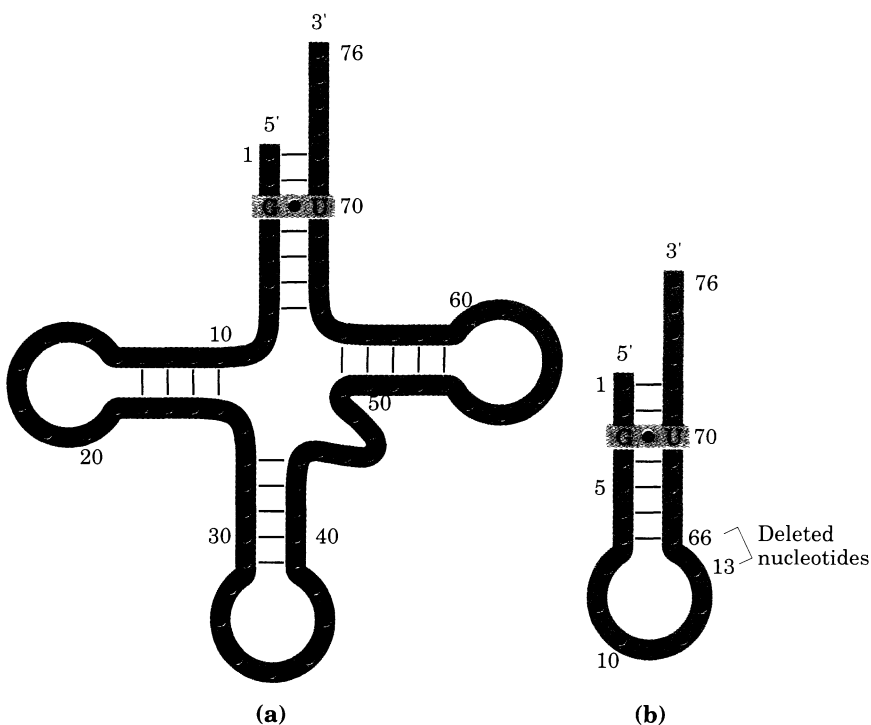

(a)   (b)

Figure 26–21 (a) The tRNA^{Ala} structural elements recognized by the Ala-tRNA synthetase are unusually simple. A single G=U base pair (red) is the only element needed for specific binding and aminoacylation. (b) A short synthetic RNA minihelix, which has the critical G=U base pair but lacks most of the remaining tRNA structure, is specifically aminoacylated with alanine almost as efficiently as the complete tRNA^{Ala}.
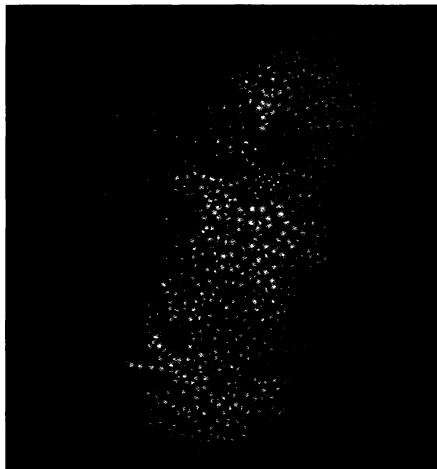
Figure 26–22 Structure of Gln-tRNA synthetase (white) bound to its cognate tRNA^{Gln} (green and red) and ATP. The three phosphate groups of the ATP, shown in yellow, are visible. In this case, bases in both the anticodon arm and the amino acid arm are the key structural features of the tRNA used for recognition by the aminoacyl-tRNA synthetase. Additional contacts between the enzyme and the tRNA revealed in this crystal structure occur along the inside of the L structure of the tRNA (see Fig. 26–17), but many of these involve residues conserved in all tRNAs and may not contribute to discrimination among different tRNAs.
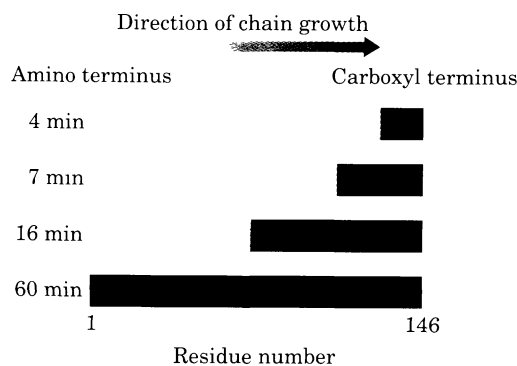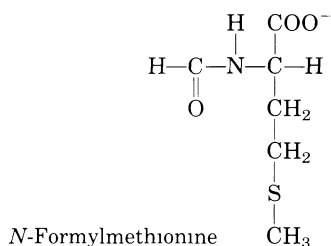

Figure 26–23 Proof that polypeptide chains grow by addition of new amino acid residues to the carboxyl end. The dark red zones show the portions of completed α-globin chains containing radioactive Leu residues at different times after addition of labeled leucine. At 4 min, only a few residues at the carboxyl end of α-globin were labeled. This is because the only *complete* globin chains that contained label after 4 min were those that had nearly completed synthesis at the time the label was added. On longer times of incubation with labeled leucine, successively longer segments of the polypeptide chain contained labeled residues, always in a block at the carboxyl end of the chain. The unlabeled end of the polypeptide (the amino terminus) was thus defined as the initiating end, and the polypeptide chain grows by successive addition of amino acids at the carboxyl end.

A complete understanding of the structural factors guiding these interactions remains an area of very active investigation. The solution of the crystal structures of two aminoacyl-tRNA synthetases (Gln and Asp) complexed with their cognate tRNAs and ATP is an important advance (Fig. 26–22). The relatively simple alanine system described above may be an evolutionary relic of a period when RNA oligonucleotides (ancestors to tRNA) were aminoacylated in a primitive system for protein synthesis.

Polypeptide Synthesis Begins at the Amino-Terminal End

Does polypeptide chain growth begin from the amino-terminal or from the carboxyl-terminal end? The answer came from isotope tracer experiments carried out by Howard Dintzis in 1961. Reticulocytes (immature erythrocytes) that were actively synthesizing hemoglobin were incubated with radioactive leucine. Leucine was chosen because it occurs frequently along both the α- and β-globin chains. Samples of completed α chains were isolated from the reticulocytes at various times after addition of radioactive leucine, and the distribution of radioactivity along the α chain was determined with the expectation that it would be concentrated in the end that was synthesized last. In those globin chains isolated after 60 min of incubation, nearly all the Leu residues were radioactive. However, in completed globin chains that were isolated only a few minutes after radioactive leucine was added, radioactive Leu residues were concentrated at the carboxyl-terminal end (Fig. 26–23). From these observations it was concluded that polypeptide chains are begun at the amino-terminal end and are elongated by sequential addition of residues to the carboxyl-terminal end. This pattern has been confirmed in innumerable additional experiments and applies to all proteins in all cells.

A Specific Amino Acid Initiates Protein Synthesis

Although there is only one codon for methionine (AUG), there are two tRNAs for methionine in all organisms. One tRNA is used exclusively when AUG represents the initiation codon for protein synthesis. The second is used when methionine is added at an internal position in a polypeptide.

In bacteria, the two separate classes of tRNA specific for methionine are designated tRNA^{Met} and tRNA^{fMet}. The starting amino acid residue at the amino-terminal end is N-formylmethionine. It enters

N-Formylmethionine

the ribosome as N-formylmethionyl-tRNA$^{\text{fMet}}$ (fMet-tRNA$^{\text{fMet}}$), which is formed in two successive reactions. First, methionine is attached to tRNA$^{\text{fMet}}$ by the Met-tRNA synthetase:

$$\text{Methionine} + \text{tRNA}^{\text{fMet}} + \text{ATP} \longrightarrow \text{Met-tRNA}^{\text{fMet}} + \text{AMP} + \text{PP}_i$$

As already noted, there is only one of these enzymes in *E. coli,* and it aminoacylates both tRNA$^{\text{fMet}}$ and tRNA$^{\text{Met}}$. Second, a formyl group is transferred to the amino group of the Met residue from N^{10}-formyltetrahydrofolate by a transformylase enzyme:

$$N^{10}\text{-Formyltetrahydrofolate} + \text{Met-tRNA}^{\text{fMet}} \longrightarrow$$
$$\text{tetrahydrofolate} + \text{fMet-tRNA}^{\text{fMet}}$$
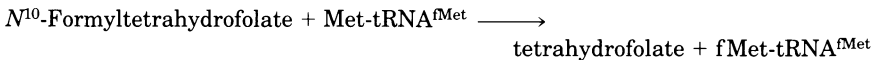
This transformylase is more selective than the Met-tRNA synthetase, and it cannot formylate free methionine or Met residues attached to tRNA$^{\text{Met}}$. Instead, it is specific for Met residues attached to tRNA$^{\text{fMet}}$, presumably recognizing some unique structural feature of that tRNA. The other Met-tRNA species, Met-tRNA$^{\text{Met}}$, is used to insert methionine in interior positions in the polypeptide chain. Blocking of the amino group of methionine by the N-formyl group not only prevents it from entering interior positions but also allows fMet-tRNA$^{\text{fMet}}$ to be bound at a specific initiation site on the ribosome that does not accept Met-tRNA$^{\text{Met}}$ or any other aminoacyl-tRNA.

In eukaryotic cells, all polypeptides synthesized by cytosolic ribosomes begin with a Met residue (as opposed to fMet), but again a specialized initiating tRNA is used that is distinct from the tRNA$^{\text{Met}}$ used at interior positions. In contrast, polypeptides synthesized by the ribosomes in the mitochondria and chloroplasts of eukaryotic cells begin with N-formylmethionine. This and other similarities in the protein-synthesizing machinery of these organelles and bacteria strongly support the view that mitochondria and chloroplasts originated from bacterial ancestors symbiotically incorporated into the precursors of eukaryotic cells at an early stage of evolution (see Fig. 2–17).

We are now left with a puzzle. There is only one codon for methionine, namely (5′)AUG. How can this single codon serve to identify both the starting N-formylmethionine (or methionine in the case of eukaryotes) and those Met residues that occur in interior positions in polypeptide chains? The answer will be found in the next section.

Initiation of Polypeptide Synthesis Has Several Steps

We now turn to a detailed examination of the second stage of protein synthesis: **initiation.** The focus here, and in the discussion of elongation and termination to follow, is on protein synthesis in bacteria; the process is not as well understood in eukaryotes. The initiation of polypeptide synthesis in bacteria requires (1) the 30S ribosomal subunit, which contains 16S rRNA, (2) the mRNA coding for the polypeptide to be made, (3) the initiating fMet-tRNA$^{\text{fMet}}$, (4) a set of three proteins called initiation factors (IF-1, IF-2, and IF-3), (5) GTP, (6) the 50S ribosomal subunit, and (7) Mg^{2+}. The formation of the initiation complex takes place in three steps (Fig. 26–24).

In the first step, the 30S ribosomal subunit binds initiation factor 3 (IF-3), which prevents the 30S and 50S subunits from combining prematurely. Binding of the mRNA to the 30S subunit then takes place in such a way that the initiation codon (AUG) binds to a precise location on the 30S subunit (Fig. 26–24).
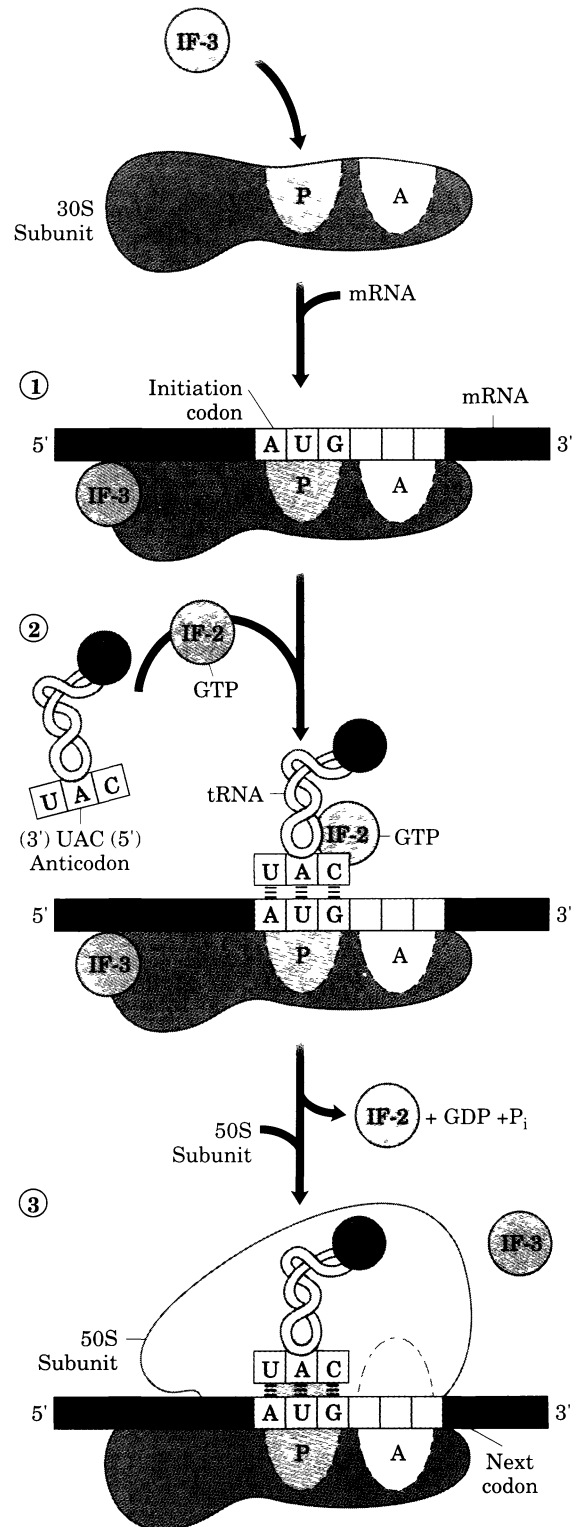

Figure 26–24 Formation of the initiation complex in three steps (described in the text) at the expense of the hydrolysis of GTP to GDP and P$_i$. IF-2 and IF-3 are initiation factors. P designates the peptidyl site, A the aminoacyl site.

E. coli trp A   (5')A G C A C **G A G** G **G A** A A A U C U G A U G G A A C G C U A C (3')

E. coli ara B   U U U G G A U **G G A G** U G A A A C G A U G G C G A U U G C A

E. coli lac I   C A A U U C A G **G C U G** G **G** G A A U G U G A A A C C A G U A

φX174 phage A protein   A A U C U U **G G A G G** C U U U U U U A U G G U U C G U U C U

λ phage cro   A U G U A C **U A A G G A G G** U U G U A U G G A A C A A C G C

Shine–Dalgarno sequence;
pairs with 16S rRNA

Initiation codon;
pairs with fMet–tRNA^fMet

(a)

```
          3'
          OH              G
3' End of                 A
16S rRNA  A               U
          U               C
          U C C U C C A
```

Idealized (5')G A U U C C U **A G G A G G U** U U G A C C U A U G C G A G C U U U U A G U (3')
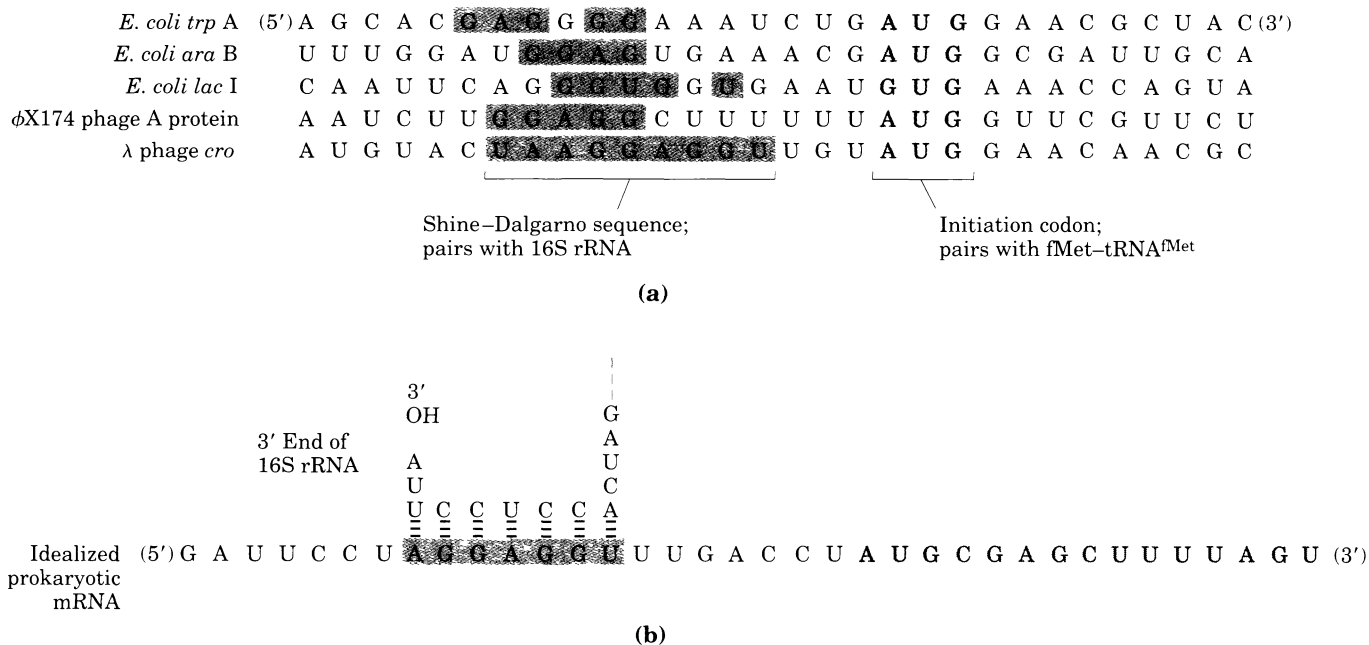prokaryotic
mRNA

(b)

Figure 26–25 Sequences on the mRNA that serve as signals for initiation of protein synthesis in prokaryotes. **(a)** Alignment of the initiating AUG (shaded in green) in the P site depends in part on Shine–Dalgarno sequences (shaded in red) upstream. Portions of the mRNA transcripts of five prokaryotic genes are shown. **(b)** The Shine–Dalgarno sequences pair with a sequence near the 3' end of the 16S rRNA, as shown for an idealized prokaryotic mRNA.

The initiating AUG is guided to the correct position on the 30S subunit by an initiating signal called the **Shine–Dalgarno sequence** in the mRNA, centered 8 to 13 base pairs to the 5' side of the initiation codon (Fig. 26–25). Generally consisting of four to nine purine residues, the Shine–Dalgarno sequence is recognized by, and base-pairs (antiparallel) with, a complementary pyrimidine-rich sequence near the 3' end of the 16S rRNA of the 30S subunit. This mRNA–rRNA interaction fixes the mRNA so that the AUG is correctly positioned for initiation of translation. The specific AUG where fMet-tRNA^fMet is to be bound is thereby distinguished from interior methionine codons by its proximity to the Shine–Dalgarno sequence in the mRNA.

Ribosomes have two sites that bind aminoacyl-tRNAs, the **aminoacyl** or **A site** and the **peptidyl** or **P site.** Both the 30S and the 50S subunits contribute to the characteristics of each site. The initiating AUG is positioned in the P site, which is the only site to which fMet-tRNA^fMet can bind (Fig. 26–24). However, fMet-tRNA^fMet is the exception: during the subsequent elongation stage, all other incoming aminoacyl-tRNAs, including the Met-tRNA^Met that binds to interior AUGs, bind to the A site. The P site is the site from which the "uncharged" tRNAs leave during elongation.

In the second step of the initiation process (Fig. 26–24), the complex consisting of the 30S subunit, IF-3, and mRNA now forms a still larger complex by binding IF-2, which already is bound to GTP and the initiating fMet-tRNA^fMet. The anticodon of this tRNA pairs correctly with the initiation codon in this step.

In the third step, this large complex combines with the 50S ribosomal subunit; simultaneously, the GTP molecule bound to IF-2 is hydrolyzed to GDP and P_i (which are released). IF-3 and IF-2 also depart from the ribosome.

A major difference in protein synthesis between prokaryotes and eukaryotes is the existence of at least nine eukaryotic initiation factors. One of these, called cap binding protein or CBPI, binds to the 5' cap of mRNA and facilitates formation of a complex between the mRNA and the 40S ribosomal subunit. The mRNA is then scanned to

Table 26–8 Protein factors required for translation initiation in bacteria and eukaryotes

Bacteria		Eukaryotes	
Factor	Function	Factor	Function
IF-1	Stimulates activities of IF-2 and IF-3		
IF-2	Facilitates binding of fMet-tRNA^{fMet} to 30S ribosomal subunit	eIF2*	Facilitates binding of initiating Met-tRNA^{Met} to 40S ribosomal subunit
IF-3	Binds to 30S subunit; prevents premature association of 50S subunit	eIF3, eIF4C	First factors to bind 40S subunit; facilitate subsequent steps
		CBPI	Binds to 5' cap of mRNA
		eIF4A, eIF4B, eIF4F	Bind to mRNA; facilitate scanning of mRNA to locate first AUG
		eIF5	Promotes dissociation of several other initiation factors from 40S subunit as prelude to association of 60S subunit to form 80S initiation complex
		eIF6	Facilitates dissociation of inactive 80S ribosome into 40S and 60S subunits

*Surprisingly, eIF2 appears to be a multifunctional protein In addition to its role in the initiation of translation, it is also involved in the splicing of mRNA precursors in the nucleus This finding provides an intriguing link between transcription and translation in eukaryotic cells

locate the first AUG codon, which signals the beginning of the reading frame. Several additional initiation factors are required in this mRNA scanning reaction, and in assembly of the complete 80S initiation complex in which the initiating Met-tRNA^{Met} and mRNA are bound and ready for elongation to proceed. The roles of the various bacterial and eukaryotic initiation factors in the overall process are summarized in Table 26–8. The mechanism by which these proteins act remains a very important area of investigation.

In bacteria, the steps in Figure 26–24 result in a functional 70S ribosome called the **initiation complex,** containing the mRNA and the initiating fMet-tRNA^{fMet}. The correct binding of the fMet-tRNA^{fMet} to the P site in the complete 70S initiation complex is assured by two points of recognition and attachment: the codon–anticodon interaction involving the initiating AUG fixed in the P site, and binding interactions between the P site and the fMet-tRNA^{fMet} The initiation complex is now ready for the elongation steps.

Peptide Bonds Are Formed during the Elongation Stage

The third stage of protein synthesis is elongation, the stepwise addition of amino acids to the polypeptide chain. Again, our discussion focuses on bacteria. Elongation requires (1) the initiation complex described above, (2) the next aminoacyl-tRNA, specified by the next codon in the mRNA, (3) a set of three soluble cytosolic proteins called **elongation factors** (EF-Tu, EF-Ts, and EF-G), and (4) GTP. Three steps take place in the addition of each amino acid residue, and this cycle is repeated as many times as there are residues to be added.
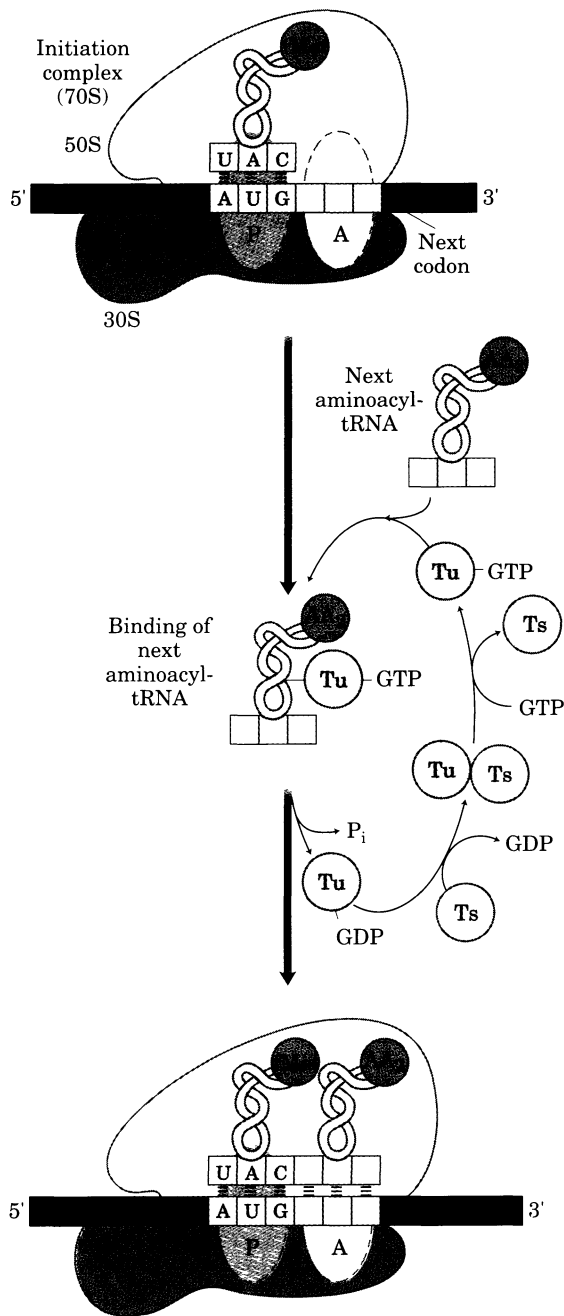
Figure 26–26 First step in elongation: the binding of the second aminoacyl-tRNA. The second amino-acyl-tRNA enters bound to EF-Tu (shown as Tu), which also contains bound GTP. Binding of the second aminoacyl-tRNA to the A site in the ribosome is accompanied by hydrolysis of the GTP to GDP and P_i, and an EF-Tu·GDP complex leaves the ribosome. The bound GDP is released when the EF-Tu·GDP complex binds to EF-Ts, and EF-Ts is subsequently released when another molecule of GTP becomes bound to EF-Tu. This recycles EF-Tu and permits it to bind another aminoacyl-tRNA.

In the first step of the elongation cycle (Fig. 26–26), the next aminoacyl-tRNA is first bound to a complex of EF-Tu containing a molecule of bound GTP. The resulting aminoacyl-tRNA–EF-Tu·GTP complex is then bound to the A site of the 70S initiation complex. The GTP is hydrolyzed, an EF-Tu·GDP complex is released from the 70S ribosome, and an EF-Tu·GTP complex is regenerated (Fig. 26–26).

In the second step, a new peptide bond is formed between the amino acids bound by their tRNAs to the A and P sites on the ribosome (Fig. 26–27). This occurs by the transfer of the initiating N-formylmethionyl group from its tRNA to the amino group of the second amino acid now in the A site. The α-amino group of the amino acid in the A site acts as nucleophile, displacing the tRNA in the P site to form the peptide bond. This reaction produces a dipeptidyl-tRNA in the A site and the now "uncharged" (deacylated) tRNA^{fMet} remains bound to the P site.

The enzymatic activity that catalyzes peptide bond formation has historically been referred to as **peptidyl transferase** and was widely assumed to be intrinsic to one or more of the proteins in the large subunit. In 1992, Harry Noller and his colleagues discovered that this activity was catalyzed not by a protein but by the 23S rRNA, adding another critical biological function for ribozymes. As indicated in Chapter 25, this startling discovery has important implications for our understanding of the evolution of life on this planet.

In the third step of the elongation cycle, called **translocation,** the ribosome moves by the distance of one codon toward the 3' end of the mRNA. Because the dipeptidyl-tRNA is still attached to the second codon of the mRNA, the movement of the ribosome shifts the dipeptidyl-tRNA from the A site to the P site, and the deacylated tRNA is released from the initial P site back into the cytosol. The third codon of the mRNA is now in the A site and the second codon in the P site. This shift of the ribosome along the mRNA requires EF-G (also called the translocase) and the energy provided by hydrolysis of another molecule of GTP (Fig. 26–28). A change in the three-dimensional conformation of the entire ribosome is believed to take place at this step in order to move the ribosome along the mRNA.

The ribosome, with its attached dipeptidyl-tRNA and mRNA, is now ready for another elongation cycle to attach the third amino acid residue. This process occurs in precisely the same way as the addition of the second. For each amino acid residue added to the chain, two GTPs are hydrolyzed to GDP and P_i. The ribosome moves from codon to codon along the mRNA toward the 3' end, adding one amino acid residue at a time to the growing chain.

The polypeptide chain always remains attached to the tRNA of the last amino acid to have been inserted. This continued attachment to a tRNA is the chemical glue that makes the entire process work. The ester linkage between the tRNA and the carboxyl terminus of the polypeptide activates the terminal carboxyl group for nucleophilic attack by the incoming amino acid to form a new peptide bond (as in Fig. 26–27). At the same time, this tRNA represents the only link between the growing polypeptide and the information in the mRNA. As the existing ester linkage between the polypeptide and tRNA is broken during peptide bond formation, a new linkage is formed because each new amino acid is itself attached to a tRNA.

In eukaryotes, the elongation cycle is quite similar. Three eukaryotic elongation factors called eEF1α, eEF1βγ, and eEF2 have functions analogous to the bacterial elongation factors EF-Tu, EF-Ts, and EF-G, respectively.
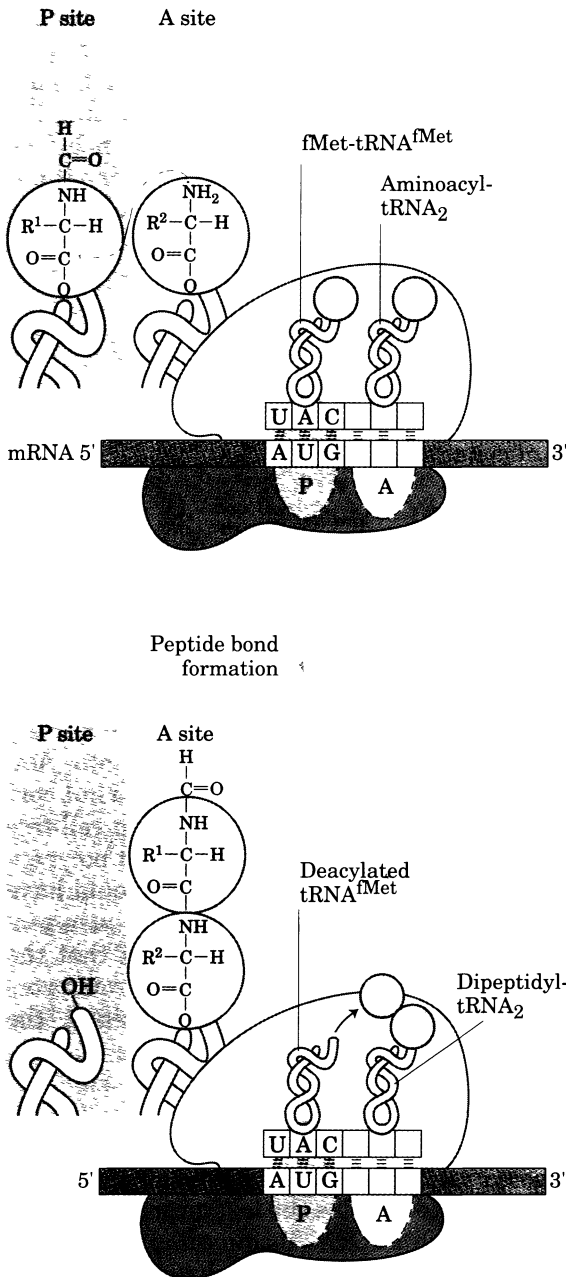

Figure 26–27 Second step in elongation: formation of the first peptide bond in bacterial protein synthesis, catalyzed by the 23S rRNA ribozyme. The *N*-formylmethionyl group is transferred to the amino group of the second aminoacyl-tRNA in the A site, forming a dipeptidyl-tRNA.
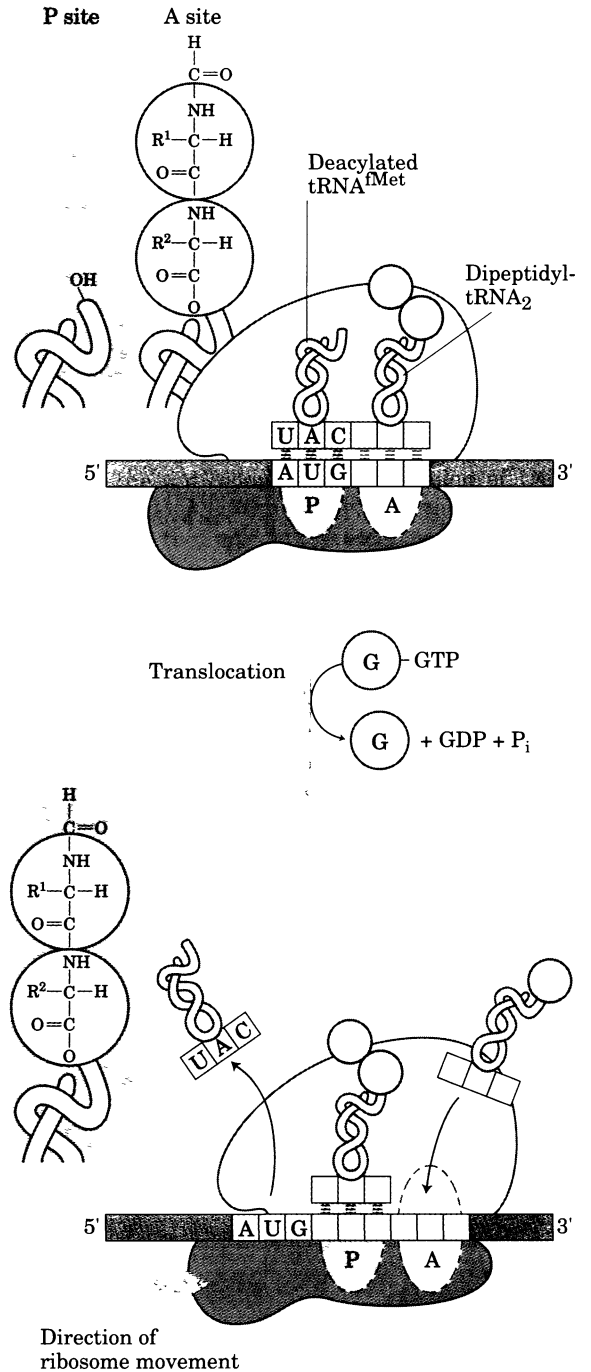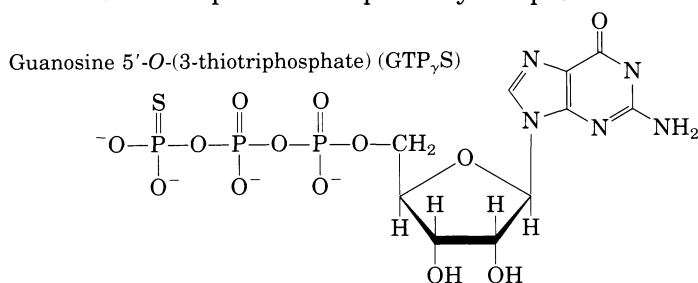

Figure 26–28 Third step in elongation: translocation. The ribosome moves one codon toward the 3′ end of mRNA, using energy provided by hydrolysis of GTP bound to EF-G (translocase; shown as G). The dipeptidyl-tRNA is now in the P site, leaving the A site open for the incoming (third) aminoacyl-tRNA.

Figure 26–29 The termination of protein synthesis in bacteria in response to a termination codon in the A site. First, a release factor, RF₁ or RF₂ depending on which termination codon is present, binds to the A site. This leads in the second step to hydrolysis of the ester linkage between the nascent polypeptide and the tRNA in the P site, and release of the completed polypeptide. Finally, the mRNA, deacylated tRNA, and release factor leave the ribosome, and the ribosome dissociates into its 30S and 50S subunits.

Proofreading on the Ribosome Is Limited to Codon–Anticodon Interactions

The GTPase activity of EF-Tu makes an important contribution to the rate and fidelity of the overall biosynthetic process. The EF-Tu•GTP complex exists for a few milliseconds, and the EF-Tu•GDP complex also exists for a similar period before it dissociates. Both of these intervals provide an opportunity for the codon–anticodon interactions to be verified (i.e., proofread). Incorrect aminoacyl-tRNAs normally dissociate during one of these periods. If the GTP analog GTPγS is used in place of GTP, hydrolysis is slowed, improving the fidelity but reducing the rate of protein synthesis. The process of protein synthesis (including the characteristics of codon–anticodon pairing already described) has clearly been optimized through evolution to balance the requirements of both speed and fidelity. Improved fidelity might diminish speed, whereas increases in speed would probably compromise fidelity.

Guanosine 5′-O-(3-thiotriphosphate) (GTPγS)

This proofreading mechanism establishes only that the proper codon–anticodon pairing has taken place. The identity of the amino acids attached to tRNAs is not checked at all on the ribosome. This was demonstrated experimentally in 1962 by two research groups led by Fritz Lipmann and Seymour Benzer. They isolated enzymatically formed Cys-tRNA^{Cys} and then chemically converted it into Ala-tRNA^{Cys}. This hybrid aminoacyl-tRNA, which carries alanine but contains the anticodon for cysteine, was then incubated with a cell-free system capable of protein synthesis. The newly synthesized polypeptide was found to contain Ala residues in positions that should have been occupied by Cys residues. This important experiment also provided timely proof for Crick's adapter hypothesis. The fact that the amino acids themselves are never checked on the ribosome reinforces the central role of aminoacyl-tRNA synthetases in maintaining the fidelity of protein biosynthesis.

Termination of Polypeptide Synthesis Requires a Special Signal

Elongation continues until the ribosome adds the last amino acid, completing the polypeptide coded by the mRNA. Termination, the fourth stage of polypeptide synthesis, is signaled by one of three termination codons in the mRNA (UAA, UAG, UGA), immediately following the last amino acid codon (Box 26–3).

In bacteria, once a termination codon occupies the ribosomal A site three **termination** or **release factors,** the proteins RF₁, RF₂, and RF₃, contribute to (1) the hydrolysis of the terminal peptidyl-tRNA bond, (2) release of the free polypeptide and the last tRNA, now uncharged, from the P site, and (3) the dissociation of the 70S ribosome into its 30S and 50S subunits, ready to start a new cycle of polypeptide synthesis (Fig. 26–29). RF₁ recognizes the termination codons UAG
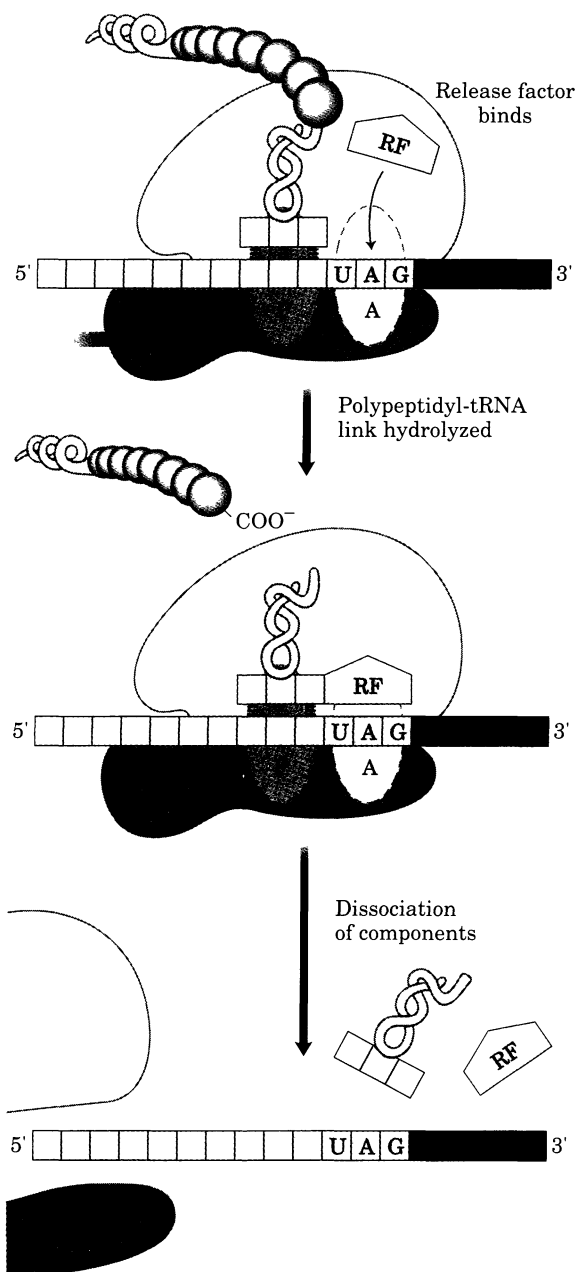

BOX 26-3   **Induced Variation in the Genetic Code: Nonsense Suppression**

When a termination codon is introduced in the interior of a gene by mutation, translation is prematurely halted and the incomplete polypeptide chains are often inactive. Such mutations are called nonsense mutations. Restoring the gene to its normal function requires a second mutation that either converts the termination codon to a codon specifying an amino acid or alternatively suppresses the effects of the termination codon. The second class of restorative mutations are called **nonsense suppressors,** and they generally involve mutations in tRNA genes that produce altered (suppressor) tRNAs that can recognize the termination codon and insert an amino acid at that position. Most suppressor tRNAs are created by single base substitutions in the anticodons of minor tRNA species.

Suppressor tRNAs constitute an experimentally induced variation in the genetic code involving the reading of what are usually termination codons, as is the case for many naturally occurring code variations described in Box 26-2. Nonsense suppression does not completely disrupt information transfer in the cell. This is because there are usually several copies of the genes for some tRNAs in any cell; some of these duplicate genes are weakly expressed and account for only a minor part of the cellular pool of a particular tRNA. Suppressor mutations usually involve these "minor" tRNA species, leaving the major tRNA to read its codon normally. For example, there are three identical genes for tRNA^{Tyr} in *E. coli,* each producing a tRNA with the anticodon (5')GUA. One of these is expressed at relatively high levels and thus represents the major tRNA^{Tyr} species; the other two genes are duplicates transcribed in only small amounts. A change in the anticodon of the tRNA product of one of these duplicate tRNA^{Tyr} genes, from (5')GUA to (5')CUA, produces a minor tRNA^{Tyr} species that will insert tyrosine at UAG stop codons. This insertion of tyrosine at UAG is inefficient, but can permit production of enough useful full-length protein from a gene with a nonsense mutation to allow the cell to live. The major tRNA^{Tyr} maintains the normal genetic code for the majority of the proteins.
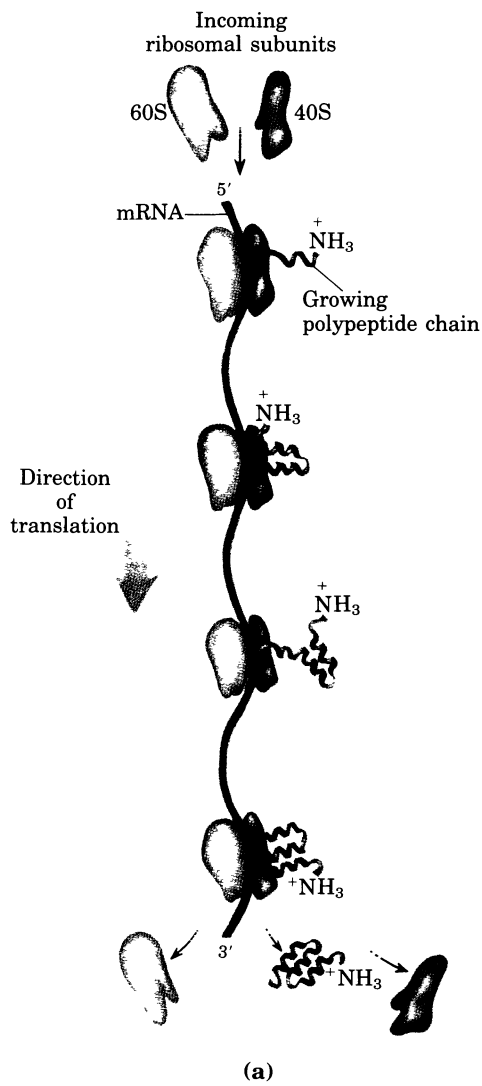
The base change in the tRNA that leads to the creation of a suppressor tRNA does not always occur in the anticodon. The suppression of UGA nonsense codons, interestingly, generally involves the tRNA^{Trp} that normally recognizes UGG. The alteration that allows it to read UGA (and insert Trp at these positions) does not occur in the anticodon. Instead, a G → A change at position 24 (in an arm of the tRNA somewhat removed from the anticodon) alters the anticodon pairing so that it can read *both* UGG and UGA. A similar change is found in tRNAs involved in the most common naturally occurring variation in the genetic code (UGA = Trp; see Box 26-2).

Suppression should lead to many abnormally long proteins, but, for reasons that are not entirely clear, this does not always occur. Many details of the molecular events that occur during translation termination and nonsense suppression are not understood.

---

and UAA, and RF$_2$ recognizes UGA and UAA. Either RF$_1$ or RF$_2$ (as appropriate, depending on which codon is present) binds at a termination codon and induces peptidyl transferase to transfer the growing peptide chain to a water molecule rather than to another amino acid. The specific function of RF$_3$ has not been firmly established. In eukaryotes, a single release factor called eRF recognizes all three termination codons.

Fidelity in Protein Synthesis Is Energetically Expensive

The enzymatic formation of each aminoacyl-tRNA used two high-energy phosphate groups. Additional ATPs are used each time incorrectly activated amino acids are hydrolyzed by the deacylation activity of some aminoacyl-tRNA synthetases (p. 914). One molecule of GTP is cleaved to GDP and P$_i$ during the first elongation step, and another GTP is hydrolyzed in the translocation step. Therefore a total of at least four high-energy bonds is ultimately required for the formation of each peptide bond of the completed polypeptide chain.

Incoming
ribosomal subunits

60S　　40S

mRNA—

$\overset{+}{N}H_3$

Growing
polypeptide chain

$\overset{+}{N}H_3$

Direction
of
translation

$\overset{+}{N}H_3$

$\overset{+}{N}H_3$

$^+NH_3$

3'

$^+NH_3$

(a)

This represents an exceedingly large thermodynamic "push" in the direction of synthesis: at least $4 \times 30.5 = 122$ kJ/mol of phosphodiester bond energy is required to generate a peptide bond having a standard free energy of hydrolysis of only about -21 kJ/mol. The net free-energy change in peptide-bond synthesis is thus -101 kJ/mol. Although this large energy expenditure may appear wasteful, it is again important to remember that proteins are information-containing polymers. The biochemical problem is not simply the formation of a peptide bond, but the formation of a peptide bond between *specific* amino acids. Each of the high-energy bonds expended in this process plays a role in a step that is critical to maintaining proper alignment between each new codon in the mRNA and the amino acid it encodes at the growing end of the polypeptide. This energy makes possible the nearly perfect fidelity in the biological translation of the genetic message of mRNA into the amino acid sequence of proteins.

Polysomes Allow Rapid Translation of a Single Message

Large clusters of 10 to 100 ribosomes can be isolated from either eukaryotic or bacterial cells that are very active in protein synthesis. Such clusters, called **polysomes,** can be fragmented into individual ribosomes by the action of ribonuclease. Furthermore, a connecting fiber between adjacent ribosomes is visible in electron micrographs (Fig. 26–30). The connecting strand is a single strand of mRNA, being translated simultaneously by many ribosomes, spaced closely together. The simultaneous translation of a single mRNA by many ribosomes allows highly efficient use of the mRNA.

In bacteria there is a very tight coupling between transcription and translation. Messenger RNAs are synthesized in the $5' \rightarrow 3'$ direction and are translated in the same direction. As shown in Figure 26–31, ribosomes begin translating the 5' end of the mRNA before transcription is complete. The situation is somewhat different in eukaryotes, where newly transcribed mRNAs must be transferred out of the nucleus before they can be translated.

0.25 μm

mRNA

Polypeptide
chains

Ribosomes

Direction of
translation
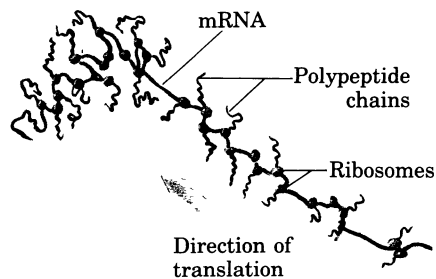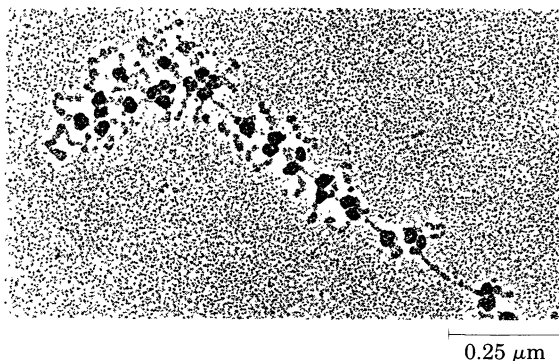
(b)

Figure 26–30 A polysome. **(a)** Four ribosomes are shown translating a eukaryotic mRNA molecule simultaneously, moving from the 5' end to the 3' end. **(b)** Electron micrograph and explanatory diagram of a polysome from the silk gland of a silk- worm larva. The mRNA is being translated by many ribosomes simultaneously. The polypeptide chains become longer as the ribosomes move toward the 3' end of the mRNA. The final product of this process is silk fibroin.
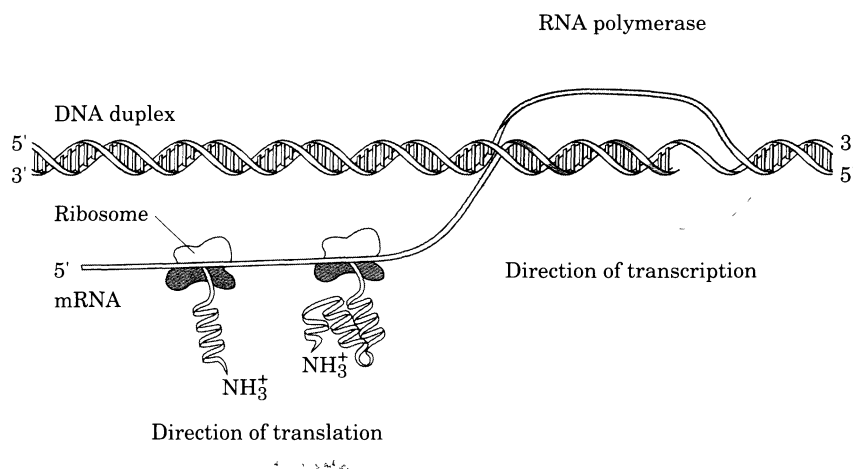
Bacterial mRNAs generally exist for only a few minutes (p. 880) before they are degraded by nucleases. Therefore, in order to maintain high rates of protein synthesis, the mRNA for a given protein or set of proteins must be made continuously and translated with maximum efficiency. The short lifetime of mRNAs in bacteria allows synthesis of a protein to cease rapidly when it is no longer needed by the cell.
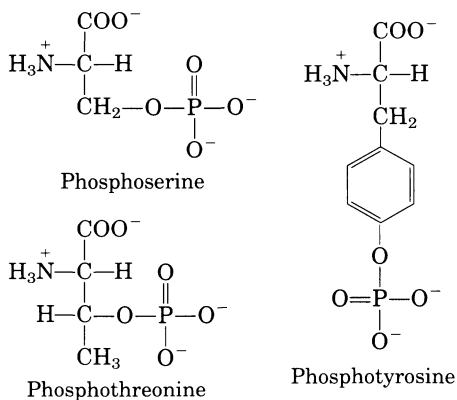
Polypeptide Chains Undergo Folding and Processing

In the fifth and final step of protein synthesis, the nascent polypeptide chain is folded and processed into its biologically active form. At some point during or after its synthesis, the polypeptide chain spontaneously assumes its native conformation, which permits the maximum number of hydrogen bonds and van der Waals, ionic, and hydrophobic interactions (see Fig. 7–22). In this way, the linear or one-dimensional genetic message in the mRNA is converted into the three-dimensional structure of the protein. Some newly made proteins do not attain their final biologically active conformation until they have been altered by one or more processing reactions called **posttranslational modifications.** Both prokaryotic and eukaryotic posttranslational modifications are considered in what follows.

Amino-Terminal and Carboxyl-Terminal Modifications Initially, all polypeptides begin with a residue of N-formylmethionine (in bacteria) or methionine (in eukaryotes). However, the formyl group, the amino-terminal Met residue, and often additional amino-terminal and carboxyl-terminal residues may be removed enzymatically and thus do not appear in the final functional proteins.

In as many as 50% of eukaryotic proteins, the amino group of the amino-terminal residue is acetylated after translation. Carboxyl-terminal residues are also sometimes modified.

Loss of Signal Sequences As we shall see, the 15 to 30 residues at the amino-terminal end of some proteins play a role in directing the protein to its ultimate destination in the cell. Such **signal sequences** are ultimately removed by specific peptidases.

$$COO^-$$
$$H_3\overset{+}{N}-\overset{|}{C}-H$$
$$CH_2-O-\overset{\overset{O}{\|}}{P}-O^-$$
$$\underset{O^-}{}$$

Phosphoserine

$$COO^-$$
$$H_3\overset{+}{N}-\overset{|}{C}-H \quad O$$
$$H-\overset{|}{C}-O-\overset{\|}{P}-O^-$$
$$\underset{CH_3 \quad O^-}{}$$

Phosphothreonine

$$COO^-$$
$$H_3\overset{+}{N}-\overset{|}{C}-H$$
$$CH_2$$

(benzene ring)

$$O$$
$$O=\overset{\|}{P}-O^-$$
$$O^-$$

Phosphotyrosine

(a)

$$COO^-$$
$$H_3\overset{+}{N}-\overset{|}{C}-H$$
$$CH_2$$
$$CH$$
$$^-OOC \quad COO^-$$

γ-Carboxyglutamate
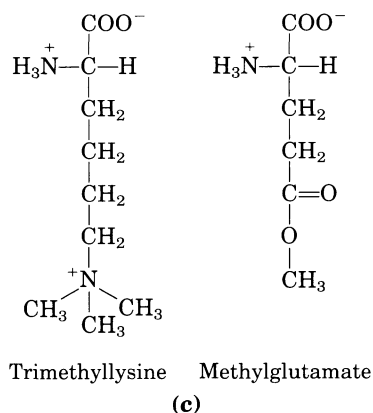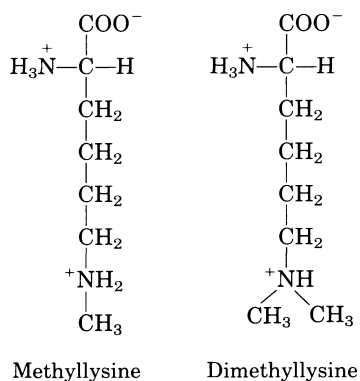
(b)

$$COO^- \qquad\qquad COO^-$$
$$H_3\overset{+}{N}-\overset{|}{C}-H \qquad H_3\overset{+}{N}-\overset{|}{C}-H$$
$$CH_2 \qquad\qquad CH_2$$
$$CH_2 \qquad\qquad CH_2$$
$$CH_2 \qquad\qquad CH_2$$
$$CH_2 \qquad\qquad CH_2$$
$$^+NH_2 \qquad\qquad ^+NH$$
$$CH_3 \qquad\qquad CH_3 \quad CH_3$$

Methyllysine   **Dimethyllysine**

$$COO^- \qquad\qquad COO^-$$
$$H_3\overset{+}{N}-\overset{|}{C}-H \qquad H_3\overset{+}{N}-\overset{|}{C}-H$$
$$CH_2 \qquad\qquad CH_2$$
$$CH_2 \qquad\qquad CH_2$$
$$CH_2 \qquad\qquad C=O$$
$$CH_2 \qquad\qquad O$$
$$^+N \qquad\qquad CH_3$$
$$CH_3 \quad CH_3$$
$$CH_3$$

Trimethyllysine   **Methylglutamate**
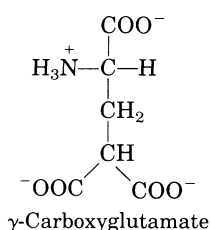
(c)

Figure 26–32 Some modified amino acid residues. **(a)** Phosphorylated amino acids. **(b)** A carboxylated amino acid. **(c)** Some methylated amino acids.

Modification of Individual Amino Acids The hydroxyl groups of certain Ser, Thr, and Tyr residues of some proteins are enzymatically phosphorylated by ATP (Fig. 26–32a); the phosphate groups add negative charges to these polypeptides. The functional significance of this modification varies from one protein to the next. For example, the milk protein casein has many phosphoserine groups, which function to bind Ca^{2+}. Given that Ca^{2+} and phosphate, as well as amino acids, are required by suckling young, casein provides three essential nutrients. The phosphorylation and dephosphorylation of the hydroxyl group of certain Ser residues is required to regulate the activity of some enzymes, such as glycogen phosphorylase (see Fig. 14–17). Phosphorylation of specific Tyr residues of some proteins is an important step in the transformation of normal cells into cancer cells (see Fig. 22–37).

Extra carboxyl groups may be added to Asp and Glu residues of some proteins. For example, the blood-clotting protein prothrombin contains a number of γ-carboxyglutamate residues (Fig. 26–32b) in its amino-terminal region, introduced by a vitamin K–requiring enzyme. These groups bind Ca^{2+}, required to initiate the clotting mechanism.

In some proteins certain Lys residues are methylated enzymatically (Fig. 26–32c). Monomethyl- and dimethyllysine residues are present in some muscle proteins and in cytochrome c. The calmodulin of most organisms contains one trimethyllysine residue at a specific position. In other proteins the carboxyl groups of some Glu residues undergo methylation (Fig. 26–32c), which removes their negative charge.

Attachment of Carbohydrate Side Chains The carbohydrate side chains of glycoproteins are attached covalently during or after the synthesis of the polypeptide chain. In some glycoproteins the carbohydrate side chain is attached enzymatically to Asn residues (N-linked oligosaccharides), in others to Ser or Thr residues (O-linked oligosaccharides; see Fig. 11–23). Many proteins that function extracellularly, as well as the "lubricating" proteoglycans coating mucous membranes, contain oligosaccharide side chains (see Fig. 11–21).

Addition of Isoprenyl Groups A number of eukaryotic proteins are isoprenylated; a thioether bond is formed between the isoprenyl group and a Cys residue of the protein (see Fig. 10–3). The isoprenyl groups are derived from pyrophosphate intermediates of the cholesterol biosynthetic pathway (see Fig. 20–34), such as farnesyl pyrophosphate (Fig. 26–33). Proteins modified in this way include the products of the *ras* oncogenes and proto-oncogenes (Chapter 22), G proteins (Chapter 22), and proteins called lamins, found in the nuclear matrix. In some
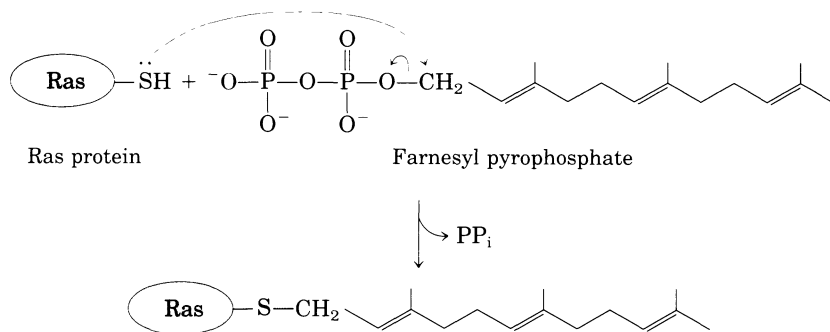
Figure 26–33 Farnesylation of a Cys residue on a protein. The thioether linkage is shown in red. The ras protein is the product of the *ras* oncogene.

cases the isoprenyl group serves to help anchor the protein in a membrane. The transforming (carcinogenic) activity of the *ras* oncogene is lost when isoprenylation is blocked, stimulating great interest in identifying inhibitors of this posttranslational modification pathway for use in cancer chemotherapy.

Addition of Prosthetic Groups Many prokaryotic and eukaryotic proteins require for their activity covalently bound prosthetic groups; these are attached to the polypeptide chain after it leaves the ribosome. Two examples are the covalently bound biotin molecule in acetyl-CoA carboxylase and the heme group of cytochrome *c*.

Proteolytic Processing Many proteins—for example, insulin (see Fig. 22–20), some viral proteins, and proteases such as trypsin and chymotrypsin (see Fig. 8–30)—are initially synthesized as larger, inactive precursor proteins. These precursors are proteolytically trimmed to produce their final, active forms.

Formation of Disulfide Cross-Links Proteins to be exported from eukaryotic cells, after undergoing spontaneous folding into their native conformations, are often covalently cross-linked by the formation of intrachain or interchain disulfide bridges between Cys residues. The cross-links formed in this way help to protect the native conformation of the protein molecule from denaturation in an extracellular environment that can differ greatly from that inside the cell.

Protein Synthesis Is Inhibited by Many Antibiotics and Toxins

Protein synthesis is a central function in cellular physiology, and as such it is the primary target of a wide variety of naturally occurring antibiotics and toxins. Except as noted, these antibiotics inhibit protein synthesis in bacteria. The differences between bacterial and eukaryotic protein synthesis are sufficient that most of these compounds are relatively harmless to eukaryotic cells. Antibiotics are important "biochemical weapons," synthesized by some microorganisms and extremely toxic to others. Antibiotics have become valuable tools in the study of protein synthesis; nearly every step in protein synthesis can be specifically inhibited by one antibiotic or another.

One of the best-understood inhibitory antibiotics is **puromycin,** made by the mold *Streptomyces alboniger*. Puromycin has a structure
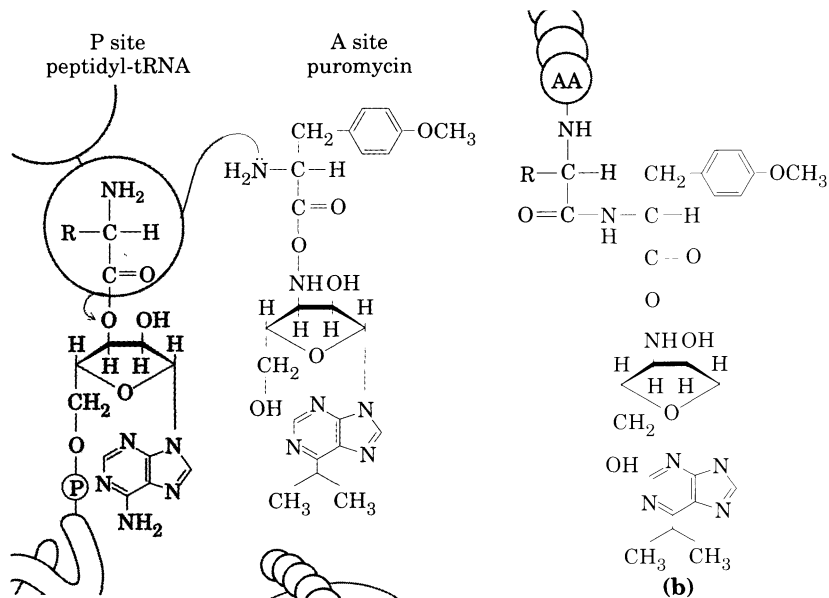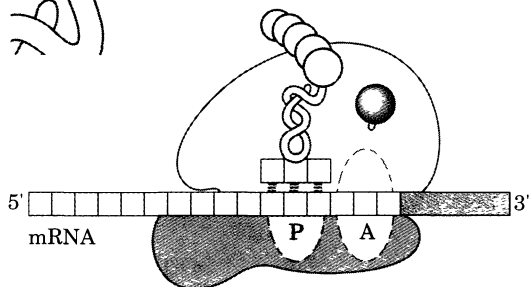
P site
peptidyl-tRNA

A site
puromycin

NH₂

R—C—H

C=O

O

H₂N—C—H

CH₂———OCH₃

C=O

O

NHOH

H   H

H H

O

CH₂

OH

N   N

N

N

CH₃   CH₃

NH₂

Figure 26–34 Puromycin resembles the aminoacyl end of a charged tRNA and can bind to the ribosomal A site, where it can participate in peptide bond formation **(a).** The product of this reaction, instead of being translocated to the P site, dissociates from the ribosome, causing premature chain termination. **(b)** Peptidyl puromycin.

AA

NH

R—C—H

O=C—N—C—H
         H

CH₂———OCH₃

C—O

O

NHOH

H   H

H H

O

CH₂

OH

N   N

N   N

CH₃   CH₃

(b)

5'   3'
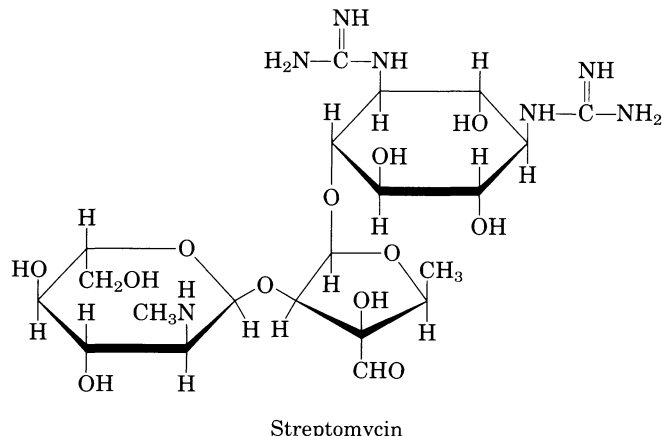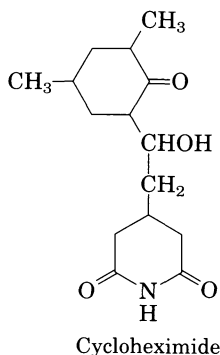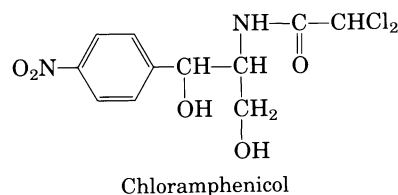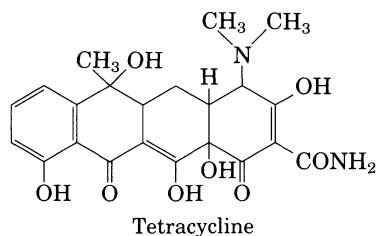mRNA   P   A

5'   3'
P   A

(a)

very similar to the 3′ end of an aminoacyl-tRNA (Fig. 26–34). It binds to the A site and participates in all elongation steps up to and including peptide bond formation, producing a peptidyl puromycin. However, puromycin will not bind to the P site, nor does it engage in translocation. It dissociates from the ribosome shortly after it is linked to the carboxyl terminus of the peptide, prematurely terminating synthesis of the polypeptide.
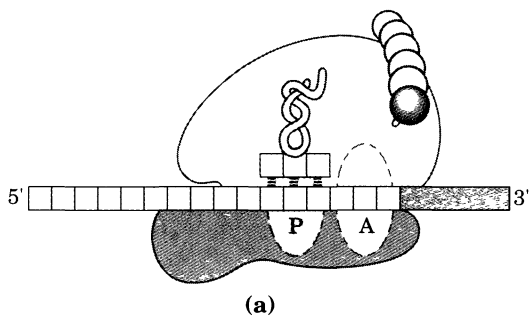
Tetracyclines inhibit protein synthesis in bacteria by blocking the A site on the ribosome, inhibiting binding of aminoacyl-tRNAs. **Chloramphenicol** inhibits protein synthesis by bacterial (and mitochondrial and chloroplast) ribosomes by blocking peptidyl transfer but does not affect cytosolic protein synthesis in eukaryotes. Conversely, **cycloheximide** blocks the peptidyl transferase of 80S eukaryotic ribosomes but not that of 70S bacterial (and mitochondrial and chloroplast) ribosomes. **Streptomycin,** a basic trisaccharide, causes misreading of the genetic code in bacteria at relatively low concentrations and inhibits initiation at higher concentrations.

Several other inhibitors of protein synthesis are notable because of their toxicity to humans and other mammals. **Diphtheria toxin** (M_r 65,000) catalyzes the ADP-ribosylation of a diphthamide (a modified histidine) residue on eukaryotic elongation factor eEF2, thereby inactivating it (see Box 8–4). **Ricin,** an extremely toxic protein of the castor bean, inactivates the 60S subunit of eukaryotic ribosomes.

CH₃   CH₃

CH₃ OH
         H   N   OH

OH

O₂N———CH—CH—CH₂
         | |
         OH   CH₂
                OH

CONH₂

OH   O   OH   O

Tetracycline

NH—C—CHCl₂
         ‖
         O

Chloramphenicol

CH₃

CH₃
         =O

CHOH

CH₂

O   N   O
         H

Cycloheximide

NH

H₂N—C—NH   H

H

HO

OH

NH

NH—C—NH₂

H

OH

O

H

HO   CH₂OH   H

H   CH₃N

OH   H

O

CH₃

OH

H

CHO

Streptomycin

928

Protein Targeting and Degradation

The eukaryotic cell is made up of many structures, compartments, and organelles, each with specific functions requiring distinct sets of proteins and enzymes. The synthesis of almost all these proteins begins on free ribosomes in the cytosol. How are these proteins directed to their final cellular destinations?

The answer to this question is at once complex, fascinating, and unfortunately incomplete. Enough is known, however, to outline many key steps in this process. Proteins destined for secretion, integration in the plasma membrane, or inclusion in lysosomes generally share the first few steps of a transport pathway that begins in the endoplasmic reticulum. Proteins destined for mitochondria, chloroplasts, or the nucleus each use separate mechanisms, and proteins destined for the cytosol simply remain where they are synthesized. The pathways by which proteins are sorted and transported to their proper cellular location are often referred to as **protein targeting** pathways.

The most important element in all of these targeting systems (with the exception of cytosolic and nuclear proteins) is a short amino acid sequence at the amino terminus of a newly synthesized polypeptide called the **signal sequence.** This signal sequence, whose function was first postulated by David Sabatini and Gunter Blobel in 1970, directs a protein to its appropriate location in the cell and is removed during transport or when the protein reaches its final destination. In many cases, the targeting capacity of particular signal sequences has been confirmed by fusing the signal sequence from one protein, say protein A, to a different protein B, and showing that the signal directs protein B to the location where protein A is normally found.

The selective degradation of proteins no longer needed in the cell also relies largely on a set of molecular signals embedded in each protein's structure; most of these signals are not yet understood. The final part of this chapter is devoted to the processes of targeting and degradation, with emphasis on the underlying signals and molecular regulation that are so crucial to cellular metabolism. Except where noted, the focus is on eukaryotic cells.
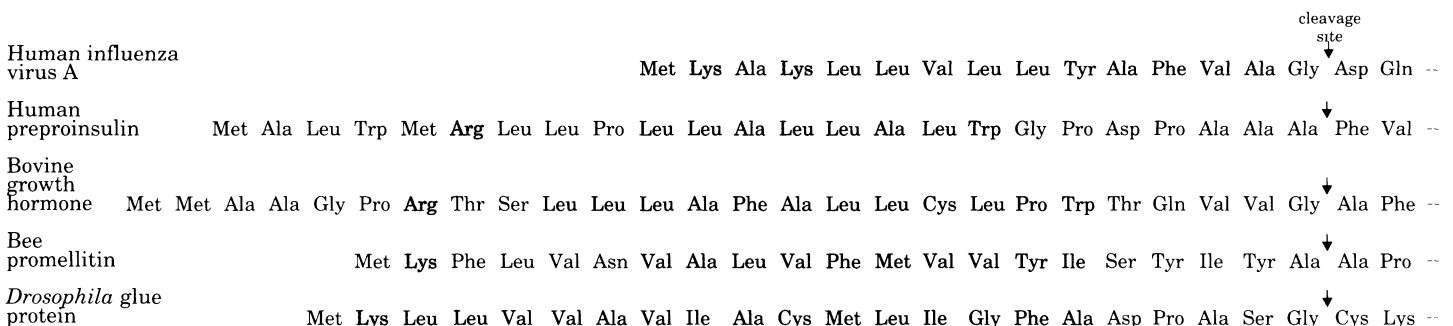
Posttranslational Modification of Many Eukaryotic Proteins Begins in the Endoplasmic Reticulum

Perhaps the best-characterized targeting system begins in the endoplasmic reticulum (ER). Most lysosomal, membrane, or secreted proteins have an amino-terminal signal sequence that marks them for translocation into the lumen of the ER. More than 100 signal sequences for proteins in this group have been determined (Fig. 26–35). The se-

Gunter Blobel

Figure 26–35 Amino-terminal signal sequences of some eukaryotic proteins, directing translocation into the endoplasmic reticulum. The hydrophobic core (yellow) is preceded by one or more basic residues (blue). Note the presence of polar and short-side-chain residues immediately preceding the cleavage sites (indicated by red arrows).

cleavage site

|																							
Human influenza virus A									Met	Lys	Ala	Lys	Leu	Leu	Val	Leu	Leu	Tyr	Ala	Phe	Val	Ala	Gly	Asp	Gln --
Human preproinsulin			Met	Ala	Leu	Trp	Met	Arg	Leu	Leu	Pro	Leu	Leu	Ala	Leu	Leu	Ala	Leu	Trp	Gly	Pro	Asp	Pro	Ala	Ala	Ala	Phe	Val --
Bovine growth hormone	Met	Met	Ala	Ala	Gly	Pro	Arg	Thr	Ser	Leu	Leu	Leu	Ala	Phe	Ala	Leu	Leu	Cys	Leu	Pro	Trp	Thr	Gln	Val	Val	Gly	Ala	Phe --
Bee promellitin			Met	Lys	Phe	Leu	Val	Asn	Val	Ala	Leu	Val	Phe	Met	Val	Val	Tyr	Ile	Ser	Tyr	Ile	Tyr	Ala	Ala	Pro --
Drosophila glue protein		Met	Lys	Leu	Leu	Val	Val	Ala	Val	Ile	Ala	Cys	Met	Leu	Ile	Gly	Phe	Ala	Asp	Pro	Ala	Ser	Gly	Cys	Lys --

George Palade

Figure 26–36 Directing eukaryotic proteins with the appropriate signals to the endoplasmic reticulum: the SRP cycle and nascent polypeptide translocation and cleavage. ① The ribosomal subunits assemble in an initiation complex at the initiation codon and begin protein synthesis. ② If an appropriate signal sequence appears at the amino terminus of the nascent polypeptide, ③ the SRP binds to the ribosome and halts elongation. ④ The ribosome–SRP complex is bound by receptors on the ER, and ⑤ the SRP dissociates and is recycled. ⑥ Protein synthesis resumes, coupled to translocation of the polypeptide chain into the lumen of the ER. ⑦ The signal sequence is cleaved by a signal peptidase within the lumen of the ER. ⑧ The ribosome is recycled.

The SRP is a rod-shaped complex containing a 300 nucleotide RNA (called 7SL-RNA) and six different proteins, with a combined molecular weight of 325,000. One protein subunit of the SRP binds directly to the signal sequence, inhibiting elongation by sterically blocking entry of aminoacyl-tRNAs and inhibiting peptidyl transferase. The SRP receptor is a heterodimer of α (M_r 69,000) and β (M_r 30,000) subunits.

quences vary in length (13 to 36 amino acid residues), but all have (1) a sequence of hydrophobic amino acids, typically 10 to 15 residues long, (2) one or more positively charged amino acid residues, usually near the amino terminus preceding the hydrophobic sequence, and (3) a short sequence at the carboxyl terminus (near the cleavage site) that is relatively polar, with amino acid residues having short side chains (especially Ala) predominating in the positions closest to the cleavage site.
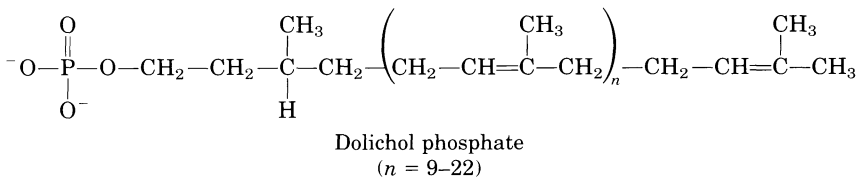
As originally demonstrated by George Palade, proteins with these signal sequences are synthesized on ribosomes attached to the ER. The signal sequence itself is instrumental in directing the ribosome to the ER. The overall pathway summarized in Figure 26–36 begins with the initiation of protein synthesis on free ribosomes. The signal sequence appears early in the synthetic process because it is at the amino terminus. As it leaves the ribosome, this sequence and the ribosome itself are rapidly bound by a large complex called the **signal recognition particle** (SRP). This binding event halts elongation when the peptide is about 70 amino acids long and the signal sequence has emerged completely from the ribosome. The bound SRP directs the ribosome with the incomplete polypeptide to a specific set of SRP receptors in the cytosolic face of the ER. The nascent polypeptide is delivered to a **peptide translocation complex** in the ER, the SRP dissociates from the ribosome, and synthesis of the protein resumes. The translocation complex feeds the growing polypeptide into the lumen of the ER in a reaction that is driven by the energy of ATP. The signal sequence is removed by a signal peptidase within the lumen of the ER. Once the complete protein has been synthesized, the ribosome dissociates from the ER.

In the lumen of the ER, newly synthesized proteins are modified in several ways. In addition to the removal of signal sequences, polypeptide chains fold and disulfide bonds form. Many proteins are also glycosylated.
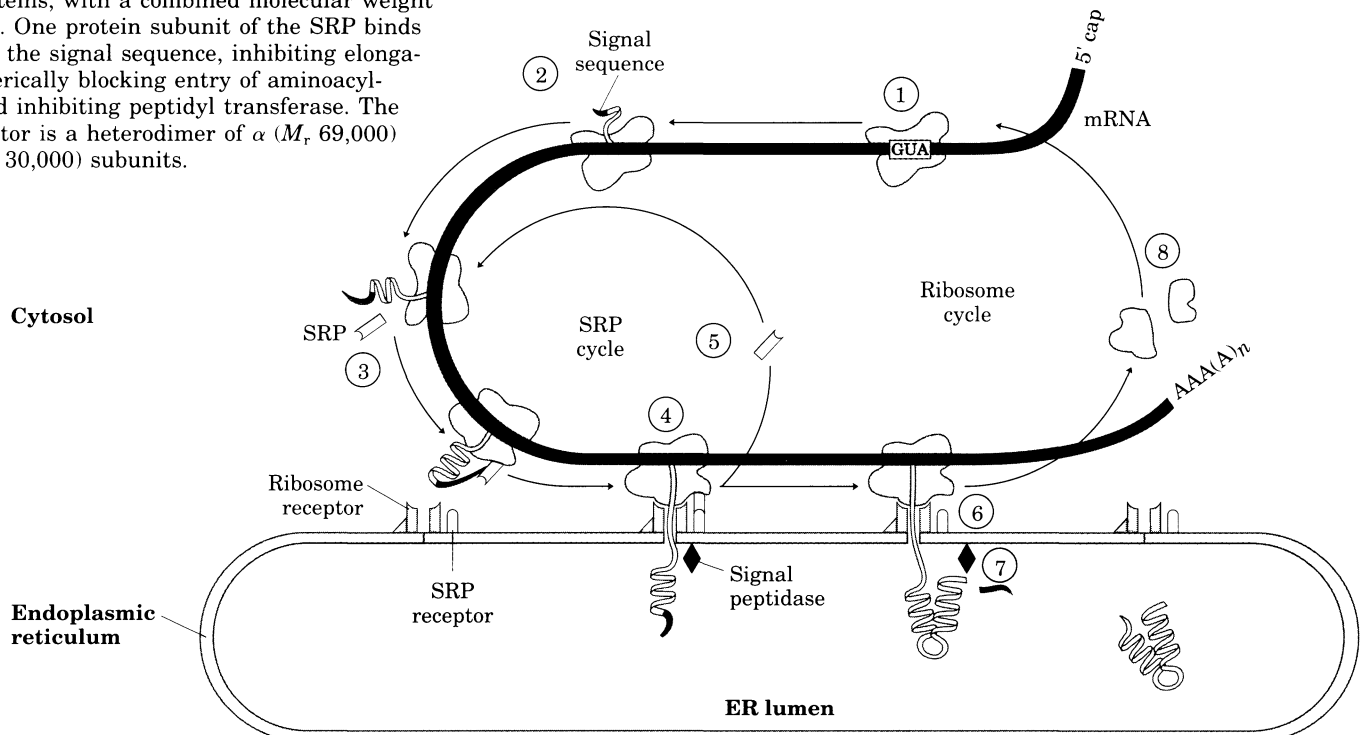
Glycosylation Plays a Key Role in Protein Targeting

Glycosylated proteins, or glycoproteins, often are linked to their oligosaccharides through Asn residues. These N-linked oligosaccharides are very diverse (Chapter 11), but the many pathways by which they form all have a common first step. A 14 residue core oligosaccharide (containing two N-acetylglucosamine, nine mannose, and three glucose residues) is transferred from a dolichol phosphate donor molecule to certain Asn residues on the proteins.

$$^-O-\overset{\overset{\displaystyle O}{\|}}{\underset{\underset{\displaystyle O^-}{|}}{P}}-O-CH_2-CH_2-\overset{\overset{\displaystyle CH_3}{|}}{\underset{\underset{\displaystyle H}{|}}{C}}-CH_2-\left(CH_2-CH=\overset{\overset{\displaystyle CH_3}{|}}{C}-CH_2\right)_n-CH_2-CH=\overset{\overset{\displaystyle CH_3}{|}}{C}-CH_3$$

Dolichol phosphate
($n = 9$–22)

The core oligosaccharide is built up on the phosphate group of dolichol phosphate (an isoprenoid derivative) by the successive addition of monosaccharide units. Once this core oligosaccharide is complete, it is enzymatically transferred from dolichol phosphate to the protein (Fig. 26–37). The transferase is located on the lumenal face of the ER and thus does not catalyze glycosylation of cytosolic proteins. After the transfer, the core oligosaccharide is trimmed and elaborated in different ways on different proteins, but all N-linked oligosaccharides retain a pentasaccharide core derived from the original 14 resi-

Figure 26–37 Synthesis of the core oligosaccharide of glycoproteins. The core oligosaccharide is built up in a series of steps as shown. The first few steps occur on the cytosolic face of the ER. Completion occurs within the lumen of the ER after a translocation step (upper left) in which the incomplete oligosaccharide is moved across the membrane. The mechanism of this translocation is not shown. The synthetic precursors that contribute additional mannose and glucose residues to the growing oligosaccharide in the lumen are themselves dolichol phosphate derivatives. The dolichol—(P)—Man and dolichol—(P)—Glc are synthesized from dolichol phosphate and GDP-mannose or UDP-glucose, respectively. After it is transferred to the protein, the core oligosaccharide is further modified in the ER and the Golgi complex in pathways that differ for different proteins. The five sugar residues enclosed in a beige screen (lower right) are retained in the final structure of all N-linked oligosaccharides. In the first step in the construction of the N-linked oligosaccharide moiety of a glycoprotein, the core oligosaccharide is transferred from dolichol phosphate to an Asn residue of the protein within the lumen of the ER. The released dolichol pyrophosphate is recycled.
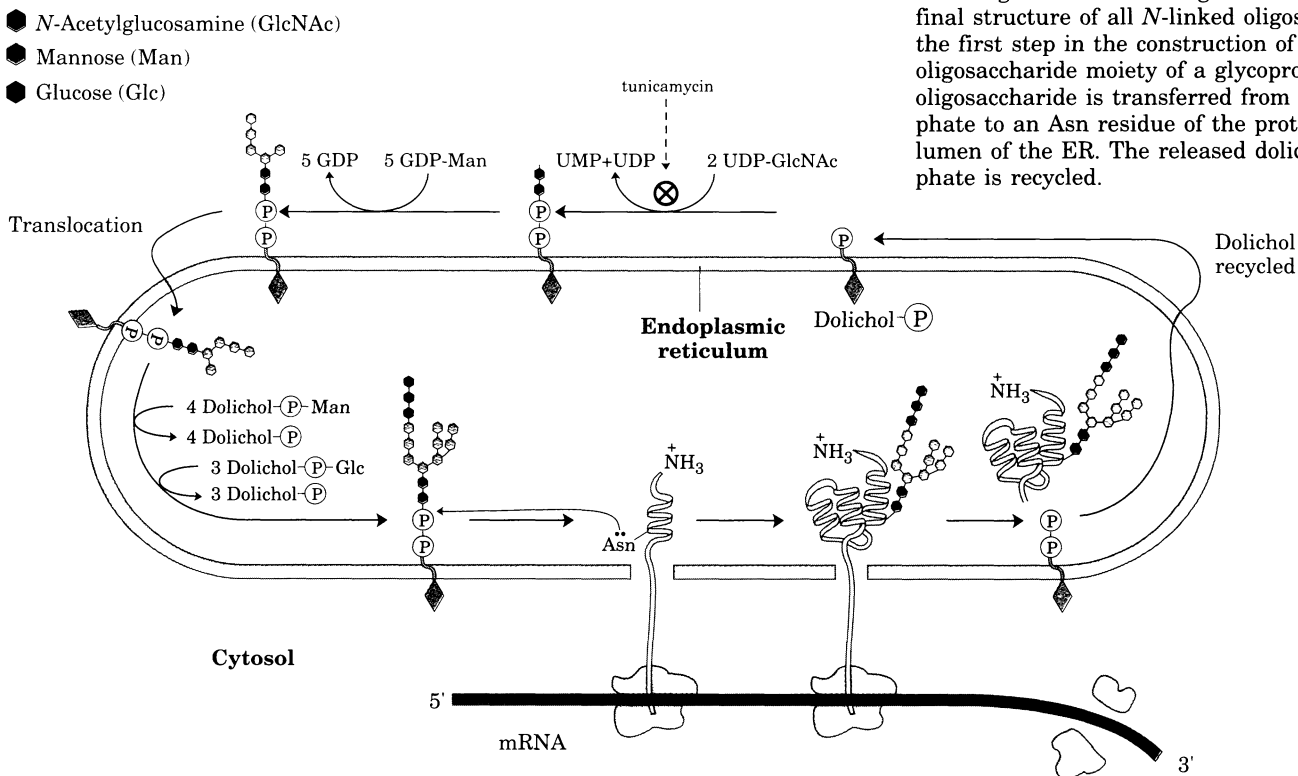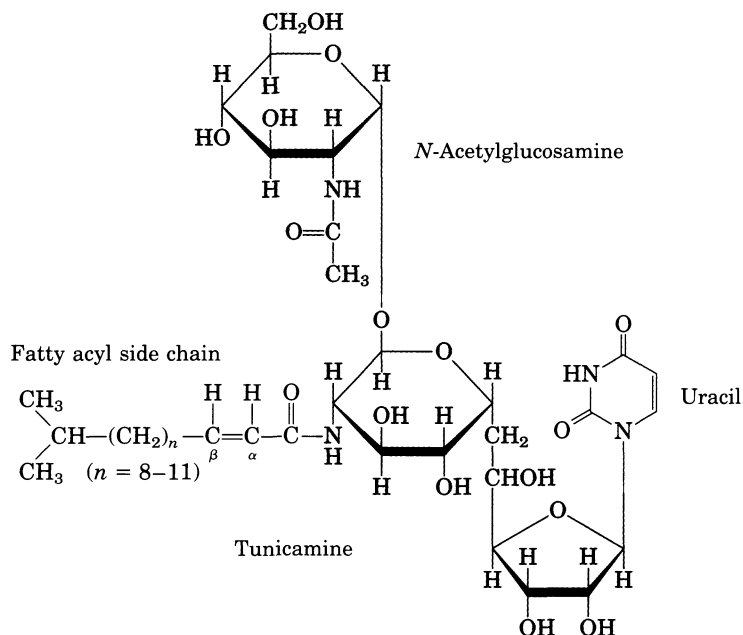

● N-Acetylglucosamine (GlcNAc)
● Mannose (Man)
● Glucose (Glc)

Figure 26–38 The structure of tunicamycin, an antibiotic produced by *Streptomyces* that mimics UDP–*N*-acetylglucosamine and blocks the first step in the synthesis of the core oligosaccharide of glycoproteins on dolichol phosphate (see Fig. 26–37). Tunicamycin is actually a family of antibiotics produced by (and isolated as a mixture from) *Streptomyces lysosuperficens*. They all contain uracil, *N*-acetylglucosamine, an 11 carbon aminodialdose called tunicamine, and a fatty acyl side chain. The structure of the fatty acyl side chain varies in the different compounds within the family. In addition to the variation in length of the fatty acyl side chain (indicated in the figure), some homologs lack the isopropyl group at the end and/or α,β-unsaturation.

Figure 26–39 The pathway taken by proteins destined for lysosomes, the plasma membrane, or secretion. Proteins are moved from the ER to the cis side of the Golgi complex in transport vesicles. Sorting occurs primarily in the trans side of the Golgi complex.

due oligosaccharide (Fig. 26–37). Several antibiotics interfere with one or more steps in this process. The best-characterized is **tunicamycin** (Fig. 26–38), which blocks the first step.

Proteins are moved from the ER to the Golgi complex in transport vesicles (Fig. 26–39). In the Golgi complex, *O*-linked oligosaccharides are added and *N*-linked oligosaccharides are further modified. By mechanisms only partially understood, proteins are also sorted here and sent to their final destinations (Fig. 26–39). Within the Golgi complex, the processes that segregate proteins destined for the cell exterior from those destined for the plasma membrane or lysosomes must distinguish between proteins on the basis of structural features other than the signal sequence, which was removed in the lumen of the ER.

This sorting process is perhaps best understood in the case of hydrolases destined for transport to lysosomes. Upon arrival in the Golgi complex from the ER, some as yet undetermined feature of the three-dimensional structure of these hydrolases (sometimes called a "signal patch") is recognized by a phosphotransferase that catalyzes the phosphorylation of certain mannose residues in the enzymes' oligosaccharides (Fig. 26–40). The presence of one or more mannose-6-phosphate residues in their *N*-linked oligosaccharides is the structural signal that targets these proteins to lysosomes. A receptor protein in the membrane of the Golgi complex recognizes this mannose-6-phosphate signal and binds the hydrolases so marked. Vesicles containing these receptor–hydrolase complexes bud from the trans side of the Golgi complex and make their way to sorting vesicles. Here, the receptor–hydrolase complexes dissociate in a process facilitated by the lower pH within the sorting vesicles and by a phosphatase-catalyzed removal of phosphate groups from the mannose-6-phosphate residues. The receptor is returned to the Golgi complex, and vesicles containing the hydrolases bud from the sorting vesicles and move to the lysosomes. In cells treated with tunicamycin (Fig. 26–38), hydrolases normally targeted for lysosomes do not reach their destination but are secreted instead, confirming that the *N*-linked oligosaccharide plays a key role in targeting these enzymes to lysosomes.
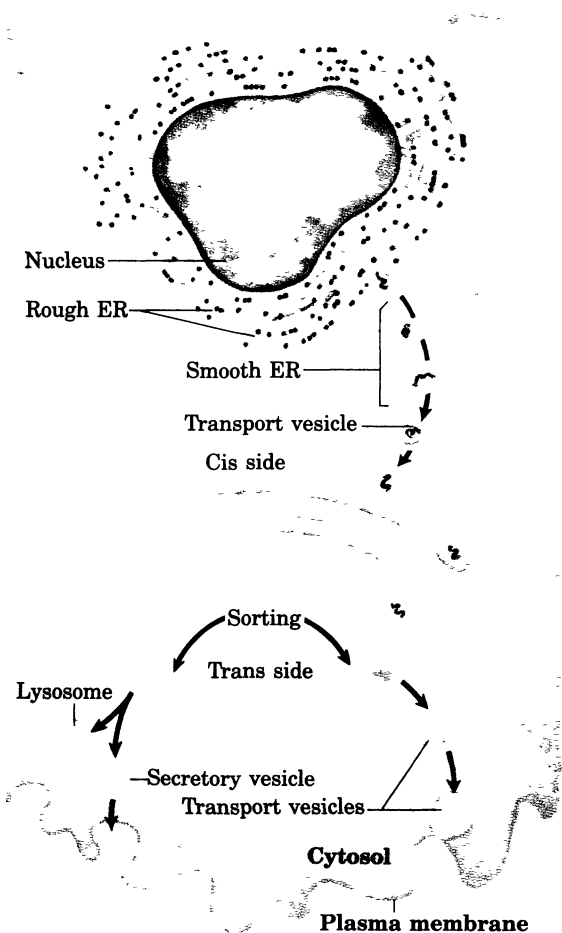
The chemical structures show:

UDP-N-Acetylglucosamine
(UDP-GlcNAc)

Mannose

Hydrolase

N-acetylglucosamine
phosphotransferase → UMP

phosphodiesterase → GlcNAc

Mannose-6-phosphate

For proteins destined for the plasma membrane or for secretion, and for those destined to reside permanently in the ER or the Golgi complex, the signals are less well understood. These targeting pathways are not impeded by tunicamycin, indicating that the signals are not carbohydrates.

Cellular proteins targeted to the mitochondria, chloroplasts, or nucleus use their own distinct signal sequences. For mitochondria and chloroplasts, the signal sequences are again found at the amino terminus of the proteins and are cleaved once the proteins arrive at their final destinations. The signal sequences that target some proteins to the nucleus (an example is the sequence –Pro–Lys–Lys–Lys–Arg–Lys–Val–) are located internally and are not cleaved. These signals permit proteins such as DNA polymerases and RNA polymerases to enter the nucleus rapidly through nuclear pores.

Bacteria Also Use Signal Sequences for Protein Targeting

Bacteria must also target some proteins to the inner or outer membranes, the periplasmic space between the membranes, or the extracellular medium (secretion). This targeting uses signal sequences at the amino terminus of the proteins much like those found on eukaryotic proteins targeted to the ER (Fig. 26–41).
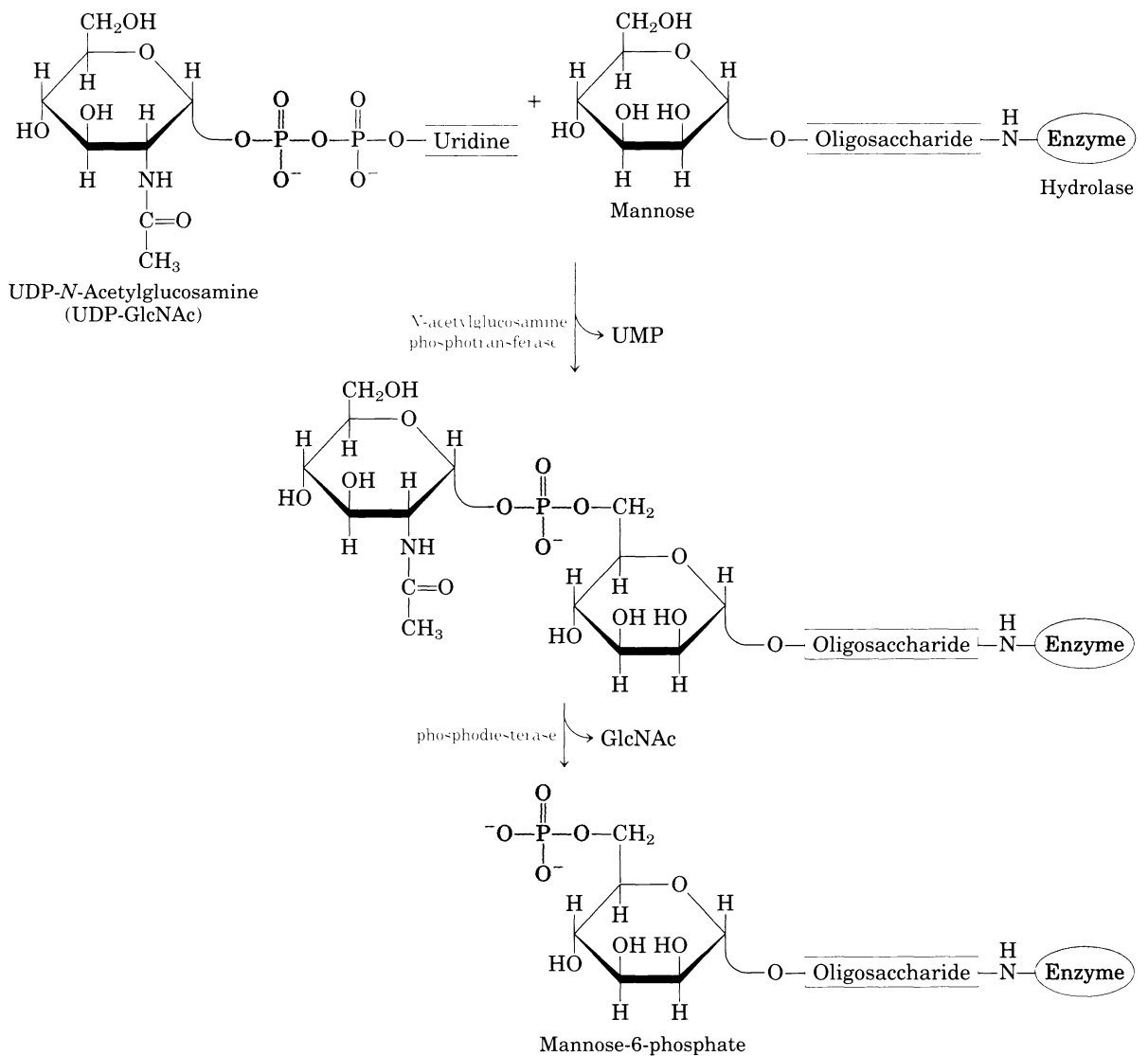
Figure 26–40 The two-step process by which mannose residues on lysosome-targeted enzymes, such as hydrolases, are phosphorylated. N-Acetylglucosamine phosphotransferase recognizes some as yet unidentified structural feature of lysosome-destined hydrolases.

933

Inner membrane proteins

Phage fd, major coat protein
Met **Lys Lys** Ser Leu Val Leu **Lys** Ala Ser Val Ala Val Ala Thr Leu Val Pro Met Leu Ser Phe **Ala**↓Ala Glu --

Phage fd, minor coat protein
Met **Lys Lys** Leu Leu Phe Ala Ile Pro Leu Val Val Pro Phe Tyr Ser His **Ser**↓Ala Glu --

Periplasmic proteins

Alkaline phosphatase
Met **Lys** Gln Ser Thr Ile Ala Leu Ala Leu Leu Pro Leu Leu Phe Thr Pro Val Thr Lys **Ala**↓Arg Thr --

Leucine-specific binding protein
Met **Lys** Ala Asn Ala **Lys** Thr Ile Ile Ala Gly Met Ile Ala Leu Ala Ile Ser His Thr Ala Met **Ala**↓Asp Asp --

β-Lactamase of pBR322
Met Ser Ile Gln **His** Phe **Arg** Val Ala Leu Ile Pro Phe Phe Ala Ala Phe Cys Leu Pro Val Phe **Ala**↓His Pro --

Outer membrane proteins

Lipoprotein
Met **Lys** Ala Thr **Lys** Leu Val Leu Gly Ala Val Ile Leu Gly Ser Thr Leu Leu Ala **Gly**↓Cys Ser --

LamB
Leu **Arg Lys** Leu Pro Leu Ala Val Ala Val Ala Ala Gly Val Met Ser Ala Gln Ala Met **Ala**↓Val Asp --

OmpA
Met Met Ile Thr Met **Lys Lys** Thr Ala Ile Ala Ile Ala Val Ala Leu Ala Gly Phe Ala Thr Val Ala Gln **Ala**↓Ala Pro --

cleavage site

Figure 26–41 Signal sequences used for targeting to different locations in bacteria. Basic amino acids (blue) near the amino terminus and hydrophobic core amino acids (yellow) are highlighted. The cleavage sites marking the ends of the signal sequences are marked by red arrows. Note that the inner membrane (see Fig. 2–6) is where phage fd coat proteins and DNA are assembled into phage particles.

(a)

~80 nm

(b)

Some proteins that are translocated through one or more membranes to reach their final destinations must be maintained in a distinct "translocation-competent" conformation until this process is complete. The functional conformation is assumed after translocation, and proteins purified in this final form are often found to be no longer capable of translocation. There is growing evidence that the translocation conformation is stabilized by a specialized set of proteins in all bacterial cells. These bind to the protein to be translocated while it is being synthesized, preventing it from folding into its final three-dimensional structure. In *E. coli,* a protein called trigger factor (M_r 63,000) appears to facilitate the translocation of at least one outer membrane protein through the inner membrane.

Cells Import Proteins by Receptor-Mediated Endocytosis

Some proteins are imported into certain cells from the surrounding medium; these include low-density lipoprotein (LDL), the iron-carrying protein transferrin, peptide hormones, and circulating proteins that are destined to be degraded. These proteins bind to receptors on the outer face of the plasma membrane. The receptors are concentrated in invaginations of the membrane called **coated pits,** which are coated on their cytosolic side with a lattice made up of the protein **clathrin** (Fig. 26–42). Clathrin forms closed polyhedral structures, and as more of the receptors become occupied with target proteins, the clathrin lattice grows until a complete membrane-bounded endocytic vesicle buds off the plasma membrane and moves into the cytoplasm.
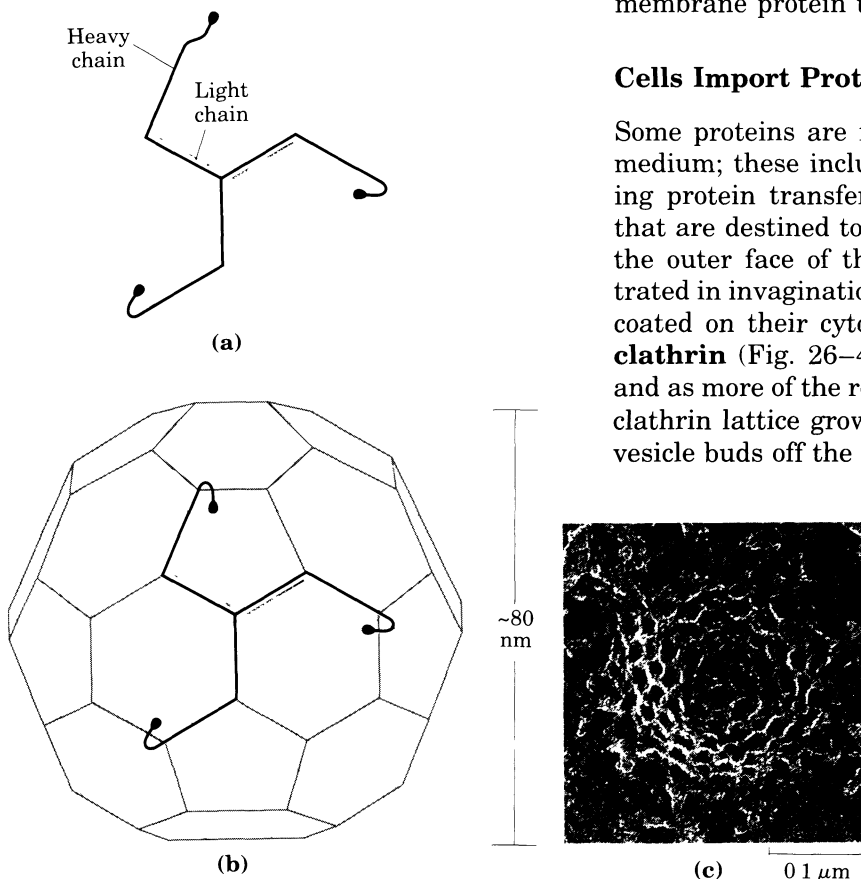
(c) 0.1 μm

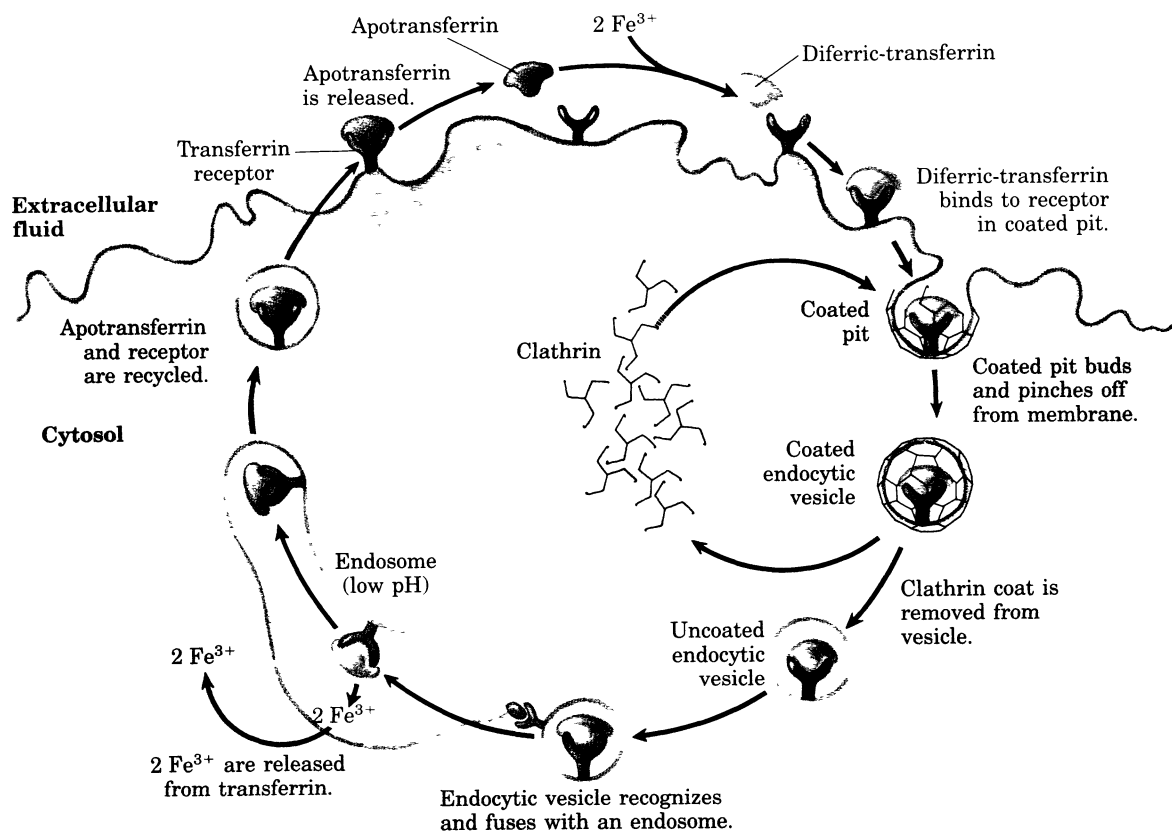**Figure 26–42** Clathrin is a trimer of three light (L) chains (M_r 35,000) and three heavy (H) chains (M_r 180,000). **(a)** The (HL)$_3$ clathrin unit is organized as a three-legged structure called a triskelion. **(b)** Triskelions have a propensity to assemble into polyhedral lattices. **(c)** Electron micrograph of a coated pit on the cytosolic face of the plasma membrane of a fibroblast.

Figure 26–43 The transferrin cycle transports iron into cells. Diferric-transferrin (transferrin containing two bound Fe^{3+} ions) is bound by receptors in coated pits (top right), which form endocytic vesicles coated with clathrin. Uncoating is catalyzed by ATP-dependent enzymes. This is followed by receptor-mediated fusion of the vesicles with endosomes (bottom). The low pH within the endosome causes dissociation of the Fe^{3+}. At low pH, the receptor retains a high affinity for apotransferrin, which is returned to the cell surface still bound to the receptors. Here the neutral pH lowers the affinity of the receptor for apotransferrin, permitting its dissociation. At neutral pH, the receptor has a high affinity for diferric-transferrin, allowing more molecules of diferric-transferrin to bind, thereby continuing the cycle.

The clathrin is quickly removed by uncoating enzymes, and the vesicles fuse with endosomes. The pH of endosomes is lowered by the activity of V-type ATPases in their membranes (see Table 10–5), producing an environment that facilitates dissociation of receptors from their target proteins. Proteins and receptors then go their separate ways, their fates varying according to the system. Transferrin and its receptor are eventually recycled (Fig. 26–43). Some hormones, growth factors, and immune complexes are degraded along with their receptors after they have elicited the appropriate response. LDL is degraded after the associated cholesterol has been delivered to its destination, but its receptor is recycled (see Fig. 20–39).

Receptor-mediated endocytosis is exploited by some viruses to gain entry to cells. Influenza virus enters cells this way. HIV, the virus that causes AIDS, also binds to specific receptors on the cell surface and may gain entry by endocytosis. In humans, the receptor that binds HIV, known as CD4, is a glycoprotein found primarily on the surface of immune system cells called helper T cells. CD4 is normally involved in the complex communication between cells of the immune system that is required to execute the immune response.

Protein Degradation Is Mediated by Specialized Systems in All Cells

Proteins are constantly being degraded in all cells to prevent the buildup of abnormal or unwanted proteins and to facilitate the recycling of amino acids. Degradation is a selective process. The lifetime of any particular protein is regulated by proteolytic systems specialized for this task, as opposed to proteolytic events that might occur during posttranslational processing. The half-lives of different proteins can vary from half a minute to many hours or even days in eukaryotes.

$$\overline{\text{Ubiquitin}} - \text{C} \underset{\text{O}^-}{\overset{\text{O}}{\diagup}}$$

Enz₁ —SH + ATP

→ AMP + PP$_i$

$$\overline{\text{Ubiquitin}} - \overset{\text{O}}{\overset{\|}{\text{C}}} - \text{S} - \text{Enz}_1$$

Enz₂—SH

Enz₁—SH

$$\overline{\text{Ubiquitin}} - \overset{\text{O}}{\overset{\|}{\text{C}}} - \text{S} - \text{Enz}_2$$

Target protein —Lys —NH₂

Enz₃

Enz₂—SH

$$\overline{\text{Ubiquitin}} - \overset{\text{O}}{\overset{\|}{\text{C}}} - \text{NH—Lys} - \text{Target protein}$$
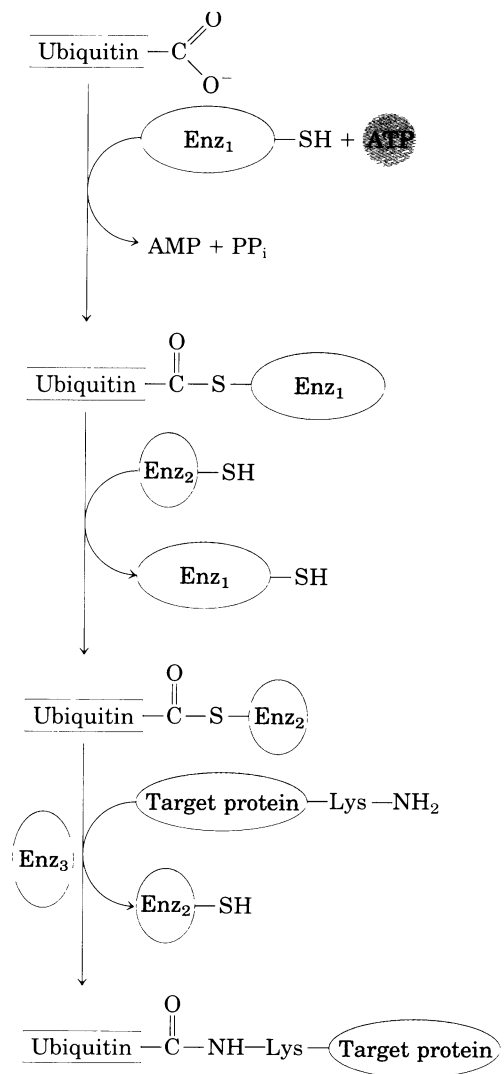
Figure 26–44 The three-step process by which ubiquitin is attached to a protein targeted for destruction in eukaryotes. Two different enzyme–ubiquitin intermediates are involved. The free carboxyl of ubiquitin's carboxyl-terminal Gly residue is ultimately linked through an amide (isopeptide) bond to an ε-amino group of a Lys residue of the target protein.

Most proteins are turned over rapidly in relation to the lifetime of a cell, although a few stable proteins (such as hemoglobin) can last for the life span of a cell (about 110 days for an erythrocyte). Proteins that are degraded rapidly include those that are defective because of one or more incorrect amino acids inserted during synthesis or because of damage that occurs during normal functioning. Also targeted for rapid turnover are many enzymes that act at key regulatory points in metabolic pathways.

Defective proteins and those with characteristically short half-lives are generally degraded in both bacteria and eukaryotes by ATP-dependent cytosolic systems. A second system in vertebrates operates in lysosomes and serves to recycle membrane proteins, extracellular proteins, and proteins with characteristically long half-lives.

In *E. coli,* many proteins are degraded by an ATP-dependent protease called La. The ATPase is activated only in the presence of defective proteins or those slated for rapid turnover; two ATP molecules are hydrolyzed for every peptide bond cleaved. The precise molecular function of ATP hydrolysis during peptide-bond cleavage is unclear. Once a protein is reduced to small inactive peptides, other ATP-independent proteases complete the degradation process.

In eukaryotes, the ATP-dependent pathway is quite different. A key component in this system is the 76 amino acid protein **ubiquitin,** so named because of its presence throughout the eukaryotic kingdoms. One of the most highly conserved proteins known, ubiquitin is essentially identical in organisms as different as yeasts and humans. Ubiquitin is covalently linked to proteins slated for destruction via an ATP-dependent pathway involving three separate enzymes (Fig. 26–44). How attachment of one or more molecules of ubiquitin to a protein targets that protein for proteolysis is not yet understood. The ATP-dependent proteolytic system in eukaryotes is a large complex ($M_r \geq 1 \times 10^6$). The mode of action of the protease component of the system and the role of ATP are unknown.

The signals that trigger ubiquitination are also not all understood, but one simple one has been found. The amino-terminal residue (i.e., the residue remaining after removal of methionine and any other proteolytic processing of the amino-terminal end) has a profound influence on the half-lives of many proteins (Table 26–9). These amino-terminal signals have evidently been conserved during billions of years of evolution; the signals are the same in bacterial protein degradation systems and in the human ubiquitination pathway. The degradation of proteins is as important to a cell's survival in a changing environment as is the protein synthetic process, and much remains to be learned about these interesting pathways.

Table 26–9 Relationship between the half-life of a protein and its amino-terminal amino acid

Amino-terminal residue	Half-life*
Stabilizing	
Met, Gly, Ala, Ser, Thr, Val	>20 h
Destabilizing	
Ile, Gln	~30 min
Tyr, Glu	~10 min
Pro	~7 min
Leu, Phe, Asp, Lys	~3 min
Arg	~2 min

Source Modified from Bachmair, A , Finley, D , & Varshavsky, A (1986) In vivo half-life of a protein is a function of its amino-terminal residue *Science* **234,** 179–186

* Half-lives were measured in yeast for a single protein that was modified so that in each experiment it had a different amino-terminal amino acid residue (See Chapter 28 for a discussion of techniques used to engineer proteins with altered amino acid sequences) Half-lives may vary for different proteins and in different organisms, but this general pattern appears to hold for all organisms amino acids listed here as stabilizing when present at the amino terminus have a stabilizing effect on proteins in all cells

Proteins are synthesized with a particular amino acid sequence through the translation of information encoded in messenger RNA by an RNA–protein complex called a ribosome. Amino acids are specified by informational units in the mRNA called codons. Translation requires adapter molecules, the transfer RNAs, which recognize codons and insert amino acids into their appropriate sequential positions in the polypeptide.

The codons for the amino acids consist of specific nucleotide triplets. The base sequences of the codons were deduced from experiments using synthetic mRNAs of known composition and sequence. The genetic code is degenerate: it has multiple code words for nearly all the amino acids. The third position in each codon is much less specific than the first and second and is said to wobble. The standard genetic code words are probably universal in all species, although some minor deviations exist in mitochondria and a few single-celled organisms. The initiating amino acid, N-formylmethionine in bacteria, is coded by AUG. Recognition of a particular AUG as the initiation codon requires a purine-rich initiating signal (the Shine–Dalgarno sequence) on the 5' side of the AUG. The triplets UAA, UAG, and UGA do not code for amino acids but are signals for chain termination. In some viruses two different proteins may be coded by the same nucleotide sequence but translated with different reading frames.

Protein synthesis occurs on the ribosomes. Bacteria have 70S ribosomes, with a large (50S) subunit and a small (30S) subunit. Ribosomes of eukaryotes are significantly larger and contain more proteins than do bacterial ribosomes.

In stage 1 of protein synthesis, amino acids are activated by specific aminoacyl-tRNA synthetases in the cytosol. These enzymes catalyze the formation of aminoacyl-tRNAs, with simultaneous cleavage of ATP to AMP and PP_i. The fidelity of protein synthesis depends to a large extent on the accuracy of this reaction, and some of these enzymes carry out proofreading steps at separate active sites. Transfer RNAs have 73 to 93 nucleotide units, several of which have modified bases. They have an amino acid arm with the terminal sequence CCA(3') to which an amino acid is esterified, an anticodon arm, a TψC arm, and a DHU arm; some tRNAs have a fifth or extra arm. The anticodon nucleotide triplet of tRNA is responsible for the specificity of interaction between the aminoacyl-tRNA and the complementary codon on the mRNA. The growth of polypeptide chains on ribosomes

begins with the amino-terminal amino acid and proceeds by successive additions of new residues to the carboxyl-terminal end.

In bacteria, the initiating aminoacyl-tRNA in all proteins is N-formylmethionyl-tRNA^{fMet}. Initiation of protein synthesis (stage 2) involves formation of a complex between the 30S ribosomal subunit, mRNA, GTP, fMet-tRNA^{fMet}, two initiation factors, and the 50S subunit; GTP is hydrolyzed to GDP and P_i. In the subsequent elongation steps (stage 3), GTP and three elongation factors are required for binding the incoming aminoacyl-tRNA to the aminoacyl site on the ribosome. In the first peptidyl transfer reaction, the fMet residue is transferred to the amino group of the incoming aminoacyl-tRNA. Movement of the ribosome along the mRNA then translocates the dipeptidyl-tRNA from the aminoacyl site to the peptidyl site, a process requiring hydrolysis of GTP. After many such elongation cycles, synthesis of the polypeptide chain is terminated (stage 4) with the aid of release factors. A polysome consists of an mRNA molecule to which are attached several or many ribosomes, each independently reading the mRNA and forming a polypeptide. At least four high-energy phosphate bonds are required to generate each peptide bond, an energy investment required to guarantee fidelity of translation. In stage 5 of protein synthesis, polypeptides undergo folding into their active, three-dimensional forms. Many proteins also are further processed by posttranslational modification reactions.

After synthesis, many proteins are directed to particular locations in the cell. One targeting mechanism involves peptide signal sequences generally found at the amino terminus of newly synthesized proteins. In eukaryotes, one class of these signal sequences is recognized and bound by a large protein–RNA complex called the signal recognition particle (SRP). The SRP binds the signal sequence as soon as it appears on the ribosome and transfers the entire ribosome and incomplete polypeptide to the endoplasmic reticulum. Polypeptides with these signal sequences are moved into the lumen of the endoplasmic reticulum as they are synthesized; there they may be modified and moved to the Golgi complex, and then sorted and sent to lysosomes, the plasma membrane, or secretory vesicles. Other known targeting signals include carbohydrates (mannose-6-phosphate targets proteins to lysosomes) and three-dimensional structural features of the proteins called signal patches. Some proteins are imported into the cell

by receptor-mediated endocytosis. These receptors are also used by some toxins and viruses to gain entry into cells.

Proteins are eventually degraded by specialized proteolytic systems present in all cells. Defective proteins and those slated for rapid turnover are generally degraded by an ATP-dependent proteolytic system. In eukaryotes, proteins to be broken down by this system are first tagged by linking them to a highly conserved protein called ubiquitin.

General

Hill, W.E., Dahlberg, A., Garrett, R.A., Moore, P.B., Schlessinger, D., & Warner, J.R. (1990) *The Ribosome: Structure, Function, and Evolution,* The American Society for Microbiology, Washington, DC.
Many good articles covering a wide range of topics.

Spirin, A.S. (1986) *Ribosome Structure and Protein Biosynthesis,* The Benjamin/Cummings Publishing Company, Menlo Park, CA.

The Genetic Code

Barrell, B.G., Air, G.M., & Hutchison, C.A., III (1976) Overlapping genes in bacteriophage ϕX174. *Nature* **264,** 34–41.

Crick, F.H.C. (1966) The genetic code: III. *Sci. Am.* **215** (October), 55–62.
An insightful overview of the genetic code at a time when the code words had just been worked out.

Fox, T.D. (1987) Natural variation in the genetic code. *Annu. Rev. Genet.* **21,** 67–91.

Hatfield, D. & Oroszlan, S. (1990) The *where, what* and *how* of ribosomal frameshifting in retroviral protein synthesis. *Trends Biochem. Sci.* **15,** 186–190.

Nirenberg, M.W. (1963) The genetic code: II. *Sci. Am.* **208** (March), 80–94.
A description of the original experiments.

Stuart, K. (1991) RNA editing in mitochondrial mRNA of trypanosomatids. *Trends Biochem. Sci.* **16,** 68–72.

Weiner, A.M. & Maizels, N. (1990) RNA editing: guided but not templated? *Cell* **61,** 917–920.

Protein Synthesis

Björk, G.R., Ericson, J.U., Gustafsson, C.E.D., Hagervall, T.G., Jonsson, Y.H., & Wikström, P.M. (1987) Transfer RNA modification. *Annu. Rev. Biochem.* **56,** 263–288.

Burbaum, J.J. & Schimmel, P. (1991) Structural relationships and the classification of aminoacyl-tRNA synthetases. *J. Biol. Chem.* **266,** 16965–16968.

Chapeville, F., Lipmann, F., von Ehrenstein, G., Weisblum, B., Ray, W.J., Jr., & Benzer, S. (1962) On the role of soluble ribonucleic acid in coding for amino acids. *Proc. Natl. Acad. Sci. USA* **48,** 1086–1092.
Classic experiments providing proof for Crick's adapter hypothesis and showing that amino acids are not checked after they are linked to tRNAs.

Clarke, S. (1992) Protein isoprenylation and methylation at carboxyl-terminal cysteine residues. *Annu. Rev. Biochem.* **61,** 355–386.

Dahlberg, A.E. (1989) The functional role of ribosomal RNA in protein synthesis. *Cell* **57,** 525–529.

Dintzis, H.M. (1961) Assembly of the peptide chains of hemoglobin. *Proc. Natl. Acad. Sci. USA* **47,** 247–261.
A classic experiment establishing that proteins are assembled beginning at the amino terminus.

Fersht, A. (1985) *Enzyme Structure and Mechanism,* W.H. Freeman and Company, New York.
See Chapter 13 for a discussion of proofreading in aminoacyl-tRNA synthetases.

Gualerzi, C.O. & Pon, C.L. (1990) Initiation of mRNA translation in prokaryotes. *Biochemistry* **29,** 5881–5889.

Lake, J.A. (1981) The ribosome. *Sci. Am.* **245** (August), 84–97.

Maden, B.E.H. (1990) The numerous modified nucleotides in eukaryotic ribosomal RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **39,** 241–303.

Moldave, K. (1985) Eukaryotic protein synthesis. *Annu. Rev. Biochem.* **54,** 1109–1149.

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

Normanly, J. & Abelson, J. (1989) tRNA identity. *Annu. Rev. Biochem.* **58,** 1029–1049.

Rich, A. & Kim, S.H. (1978) The three-dimensional structure of transfer RNA. *Sci. Am.* **238,** (January), 52–62.

Riis, B., Rattan, S.I.S., Clark, B.F.C., & Merrick, W.C. (1990) Eukaryotic protein elongation factors. *Trends Biochem. Sci.* **15,** 420–424.

Schimmel, P. (1989) Parameters for the molecular recognition of transfer RNAs. *Biochemistry* **28,** 2747–2759.

Protein Targeting and Secretion

Bachmair, A., Finley, D., & Varshavsky, A. (1986) In vivo half-life of a protein is a function of its amino-terminal residue. *Science* **234,** 179–186.

Balch, W.E. (1989) Biochemistry of interorganelle transport: a new frontier in enzymology emerges from versatile in vitro model systems. *J. Biol. Chem.* **264,** 16965–16968.

Dahms, N.M., Lobel, P., & Kornfeld, S. (1989) Mannose 6-phosphate receptors and lysosomal enzyme targeting. *J. Biol. Chem.* **264,** 12115–12118.

Goldstein, J.L., Brown, M.S., Anderson, R.G.W., Russell, D.W., & Schneider, W.J. (1985) Receptor-mediated endocytosis: concepts emerging from the LDL receptor system. *Annu. Rev. Cell Biol.* **1,** 1–39.

Hershko, A. & Ciechanover, A. (1992) The ubiquitin system for protein degradation. *Annu. Rev. Biochem.* **61,** 761–807.

Hurt, E.C. & van Loon, A.P.G.M. (1986) How proteins find mitochondria and intramitochondrial compartments. *Trends Biochem. Sci.* **11,** 204–207.

Mellman, I., Fuchs, R., & Helenius, A. (1986) Acidification of the endocytic and exocytic pathways. *Annu. Rev. Biochem.* **55,** 663–700.

Meyer, D.I. (1988) Preprotein conformation: the year's major theme in translocation studies. *Trends Biochem. Sci.* **13,** 471–474.

Pfeffer, S.R. & Rothman, J.E. (1987) Biosynthetic protein transport and sorting by the endoplasmic reticulum and Golgi. *Annu. Rev. Biochem.* **56,** 829–852.

Pryer, N.K., Wuestehube, L.J. & Schekman, R. (1992) Vesicle-mediated protein sorting. *Annu. Rev. Biochem.* **61,** 471–516.

Randall, L.L. & Hardy, S.J.S. (1984) Export of protein in bacteria. *Microbiol. Rev.* **48,** 290–298.

Rapoport, T.A. (1990) Protein transport across the ER membrane. *Trends Biochem. Sci.* **15,** 355–358.

Rothman, J.E. (1985) The compartmental organization of the Golgi apparatus. *Sci. Am.* **253** (September), 74–89.

Schmidt, G.W. & Mishkind, M.L. (1986) The transport of proteins into chloroplasts. *Annu. Rev. Biochem.* **55,** 879–912.

Silver, P.A. (1991) How proteins enter the nucleus. *Cell* **64,** 489–497.

Ward, W.H.J. (1987) Diphtheria toxin: a novel cytocidal enzyme. *Trends Biochem. Sci.* **12,** 28–31.

Wickner, W.T. & Lodish, H.F. (1985) Multiple mechanisms of protein insertion into and across membranes. *Science* **230,** 400–407.

Problems

1. *Messenger RNA Translation* Predict the amino acid sequences of peptides formed by ribosomes in response to the following mRNAs, assuming that the initial codon is the first three bases in each sequence.
  (a) GGUCAGUCGCUCCUGAUU
  (b) UUGGAUGCGCCAUAAUUUGCU
  (c) CAUGAUGCCUGUUGCUAC
  (d) AUGGACGAA

2. *How Many mRNAs Can Specify One Amino Acid Sequence?* Write all the possible mRNA sequences that can code for the simple tripeptide segment Leu–Met–Tyr. Your answer will give you some idea as to the number of possible mRNAs that can code for one polypeptide.

3. *Can the Base Sequence of an mRNA Be Predicted from the Amino Acid Sequences of Its Polypeptide Product?* A given sequence of bases in an mRNA will code for one and only one sequence of amino acids in a polypeptide, if the reading frame is specified. From a given sequence of amino acid residues in a protein such as cytochrome *c*, can we predict the base sequence of the unique mRNA that coded for it? Give reasons for your answer.

4. *Coding of a Polypeptide by Duplex DNA* The template strand of a sample of double-helical DNA contains the sequence

(5′)CTTAACACCCCTGACTTCGCGCCGTCG

(a) What is the base sequence of mRNA that can be transcribed from this strand?

(b) What amino acid sequence could be coded by the mRNA base sequence in (a), starting from the 5′ end?

(c) Suppose the other (nontemplate) strand of this DNA sample is transcribed and translated. Will the resulting amino acid sequence be the same as in (b)? Explain the biological significance of your answer.

5. *Methionine Has Only One Codon* Methionine is one of the two amino acids having only one codon. Yet the single codon for methionine can specify both the initiating residue and interior Met residues of polypeptides synthesized by *E. coli*. Explain exactly how this is possible.

6. *Synthetic mRNAs* How would you make a polyribonucleotide that could serve as an mRNA coding predominantly for many Phe residues and a small number of Leu and Ser residues? What other amino acid(s) would be coded for by this polyribonucleotide but in smaller amounts?

7. *The Direct Energy Cost of Protein Biosynthesis* Determine the minimum energy cost, in terms of high-energy phosphate groups expended, required for the biosynthesis of the β-globin chain of hemoglobin (146 residues), starting from a pool including all necessary amino acids, ATP, and GTP. Compare your answer with the direct energy cost of the biosynthesis of a linear glycogen chain of 146 glucose residues in (α1→4) linkage, starting from a pool including glucose, UTP, and ATP (Chapter 19). From your data, what is the *extra* energy cost of imparting the genetic information inherent in the β-globin molecule?

8. *Indirect Costs of Protein Synthesis* In addition to the direct energy cost for the synthesis of a protein, as developed in Problem 7, there are indirect energy costs—those required for the cell to make the necessary biocatalysts for protein synthesis. Contrast the relative magnitude of the indirect costs to a eukaryotic cell of the biosynthesis of linear (α1→4) glycogen chains versus the indirect costs of the biosynthesis of polypeptides. (Compare the enzymatic machinery used to synthesize proteins and glycogen.)

9. *Predicting Anticodons from Codons* Most amino acids have more than one codon and will be attached to more than one tRNA, each with a different anticodon. Write all possible anticodons for the four codons for glycine: (5′)GGU, GGC, GGA, and GGG.

(a) From your answer, which of the positions in the anticodons are primary determinants of their codon specificity in the case of glycine?

(b) Which of these anticodon–codon pairings have a wobbly base pair?

(c) In which of the anticodon–codon pairings do all three positions exhibit strong Watson–Crick hydrogen bonding?

10. *The Effect of Single-Base Changes on Amino Acid Sequence* Much important confirmatory evidence on the genetic code has come from the nature of single-residue changes in the amino acid sequence of mutant proteins. Which of the following single-residue amino acid replacements would be consistent with the genetic code? Which cannot be the result of single-base mutations? Why?

(a) Phe → Leu   (e) Ile → Leu
(b) Lys → Ala   (f) His → Glu
(c) Ala → Thr   (g) Pro → Ser
(d) Phe → Lys

11. *The Basis of the Sickle-Cell Mutation* In sickle-cell hemoglobin there is a Val residue at position 6 of the β-globin chain, instead of the Glu residue found in this position in normal hemoglobin A. Can you predict what change took place in the DNA codon for glutamate to account for its replacement by valine?

12. *Importance of the "Second Genetic Code"* Some aminoacyl-tRNA synthetases do not bind the anticodon of their cognate tRNAs but instead use other structural features of the tRNAs to impart binding specificity. The tRNAs for alanine apparently fall into this category. Describe the consequences of a C → G mutation in the third position of the anticodon of tRNA^Ala. What other kinds of mutations might have similar effects? Mutations of these kinds are never found in natural populations of any organism. Why? (Hint: Consider what might happen both to individual proteins and to the organism as a whole.)

13. *Maintaining the Fidelity of Protein Synthesis* The chemical mechanisms used to avoid errors in protein synthesis are different from those used during DNA replication. DNA polymerases utilize a 3′→5′ exonuclease proofreading activity to remove mispaired nucleotides incorrectly inserted into a growing DNA strand. There is no analogous proofreading function on ribosomes; and, in fact, the identity of amino acids attached to incoming tRNAs and added to the growing polypeptide is never checked. A proofreading step that hydrolyzed the last peptide bond formed when an incorrect amino acid was inserted into a growing polypeptide (analogous to the proofreading step of DNA polymerases) would actually be chemically impractical. Why? (Hint: Consider how the link between the growing polypeptide and the mRNA is maintained during the elongation phase of protein synthesis; see Figs. 26–27 and 26–28.)