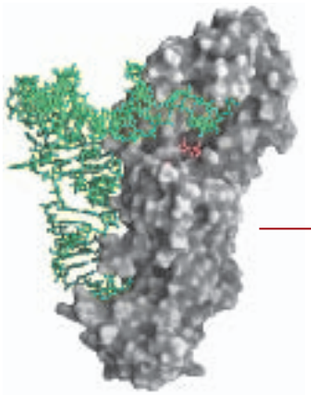


chapter 27

PROTEIN METABOLISM

Obviously, Harry [Noller]'s finding doesn't speak to how life started, and it doesn't explain what came before RNA. But as part of the continually growing body of circumstantial evidence that there was a life form before us on this planet, from which we emerged—boy, it's very strong!

—Gerald Joyce, quoted in commentary in Science, *1992*

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 their appropriate cellular locations, and degraded when no longer needed.

An understanding of protein synthesis, the most complex biosynthetic process, has been one of the greatest challenges in biochemistry. Eukaryotic protein synthesis involves more than 70 different ribosomal proteins; 20 or more enzymes to activate the amino acid precursors; a dozen or more auxiliary enzymes and other protein factors for the initiation, elongation, and termination of polypeptides; perhaps 100 additional enzymes for the final processing of different proteins; and 40 or more kinds of transfer and ribosomal RNAs. Overall, almost 300 different macromolecules cooperate to synthesize polypeptides. Many of these macromolecules are organized into the complex three-dimensional structure of the ribosome.

To appreciate the central importance of protein synthesis, consider the cellular resources devoted to this process. Protein synthesis can account for up to 90% of the chemical energy used by a cell for all biosynthetic reactions. Every prokaryotic and eukaryotic cell contains from several to thousands of copies of many different proteins and RNAs. The 15,000 ribosomes, 100,000 molecules of protein synthesis–related protein factors and enzymes, and 200,000 tRNA molecules in a typical bacterial cell can account for more than 35% of the cell's dry weight.

Despite the great complexity of protein synthesis, proteins are made at exceedingly high rates. A polypeptide of 100 residues is synthesized in an *Escherichia coli* cell (at 37 °C) in about 5 seconds. Synthesis of the thousands of different proteins in a cell is tightly regulated, so that just enough copies are made to match the current metabolic circumstances. To maintain the appropriate mix and concentration of proteins, the targeting and degradative processes must keep pace with synthesis. Research is gradually uncovering the finely coordinated cellular choreography that guides each protein to its proper cellular location and selectively degrades it when it is no longer required.

The study of protein synthesis offers another important reward: a look at a world of RNA catalysts that may have existed before the dawn of life "as we know it." Researchers have elucidated the structure of bacterial ribosomes, revealing the workings of cellular protein synthesis in beautiful molecular detail. And what did they find? Proteins are synthesized by a gigantic RNA enzyme!

27.1 The Genetic Code

Three major advances set the stage for our present knowledge of protein biosynthesis. First, in the early 1950s, Paul Zamecnik and his colleagues designed a set of experiments to investigate where in the cell proteins are synthesized. They injected radioactive amino acids into rats and, at different time intervals after the injec-

Paul Zamecnik

tion, removed the liver, homogenized it, fractionated the homogenate by centrifugation, and examined the subcellular fractions 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 only minutes had elapsed, labeled protein appeared only in a fraction containing small ribonucleoprotein particles. These particles, visible in animal tissues by electron microscopy, were therefore identified as the site of protein synthesis from amino acids, and later were named ribosomes (Fig. 27–1).

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

The third advance resulted from Francis Crick's reasoning on how the genetic information encoded in the 4-letter language of nucleic acids could be translated into

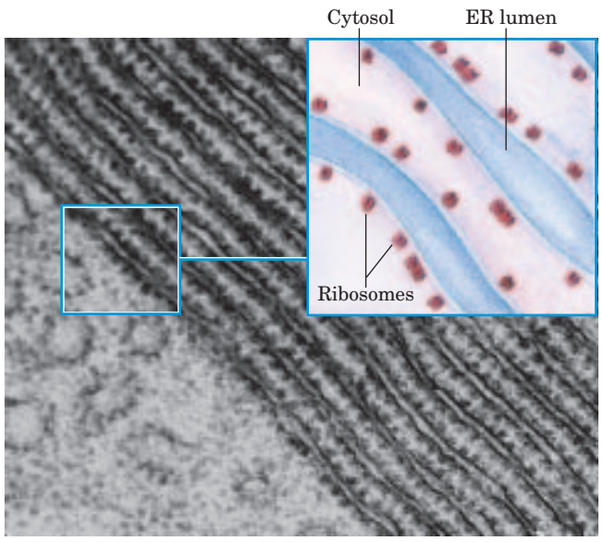


FIGURE 27–1 Ribosomes and endoplasmic reticulum. Electron micrograph and schematic drawing of a portion of a pancreatic cell, showing ribosomes attached to the outer (cytosolic) face of the endoplasmic reticulum (ER). The ribosomes are the numerous small dots bordering the parallel layers of membranes.

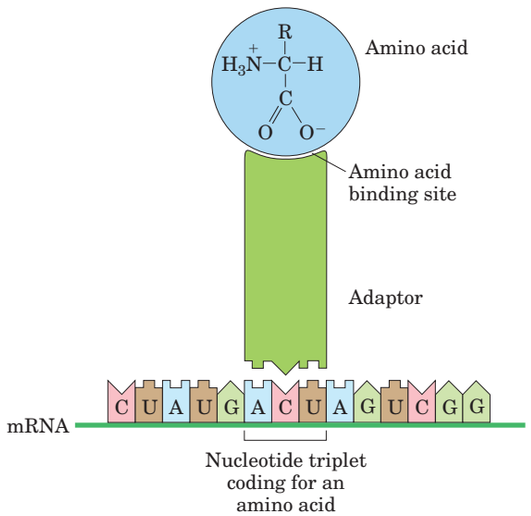


FIGURE 27–2 Crick's adaptor hypothesis. Today we know that the amino acid is covalently bound at the 3′ end of a tRNA molecule and that a specific nucleotide triplet elsewhere in the tRNA interacts with a particular triplet codon in mRNA through hydrogen bonding of complementary bases.

the 20-letter language of proteins. A small nucleic acid (perhaps RNA) could serve the role of an adaptor, one part of the adaptor molecule binding a specific amino acid and another part recognizing the nucleotide sequence encoding that amino acid in an mRNA (Fig. 27–2). This idea was soon verified. The tRNA adaptor "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 three developments soon led to recognition of the major stages of protein synthesis and ultimately to the elucidation of the genetic code that specifies each amino acid.

The Genetic Code Was Cracked Using Artificial mRNA Templates

By the 1960s it had long been apparent that at least three nucleotide residues of DNA are necessary to encode 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, insufficient to encode 20 amino acids. Groups of three, however, yield $4^3 = 64$ different combinations.

Several key properties of the genetic code were established in early genetic studies (Figs 27–3, 27–4). A **codon** is a triplet of nucleotides that codes for a specific amino acid. Translation occurs in such a way that these nucleotide triplets are read in a successive, nonoverlapping fashion. A specific first codon in the

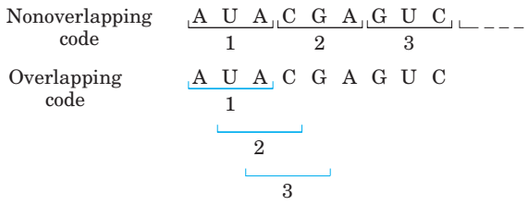


FIGURE 27–3 Overlapping versus nonoverlapping genetic codes. In a nonoverlapping code, codons (numbered consecutively) do not share nucleotides. In an overlapping code, some nucleotides in the mRNA are shared by different codons. In a triplet code with maximum overlap, many nucleotides, such as the third nucleotide from the left (A), are shared by three codons. Note that in an overlapping code, the triplet sequence of the first codon limits the possible sequences for the second codon. A nonoverlapping code provides much more flexibility in the triplet sequence of neighboring codons and therefore in the possible amino acid sequences designated by the code. The genetic code used in all living systems is now known to be nonoverlapping.

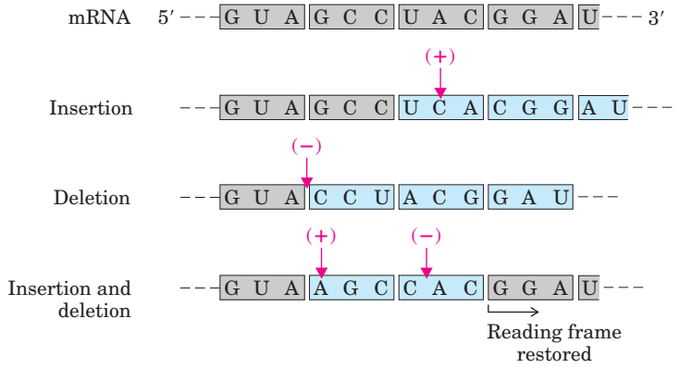
sequence establishes the **reading frame,** in which a new codon begins every three nucleotide residues. There is no punctuation between codons for successive amino acid residues. The amino acid sequence of a protein is defined by a linear sequence of contiguous triplets. In principle, any given single-stranded DNA or mRNA sequence has three possible reading frames. Each reading frame gives a different sequence of codons (Fig. 27–5), but only one is likely to encode a given protein. A key question remained: what were the three-letter code words for each amino acid?

In 1961 Marshall Nirenberg and Heinrich Matthaei reported the first breakthrough. They incubated synthetic polyuridylate, poly(U), with an *E. coli* extract, GTP, ATP, and a mixture of the 20 amino acids in 20 different tubes, each tube containing a different radioactively labeled amino acid. Because poly(U) mRNA is made up of many successive UUU triplets, it should promote the synthesis of a polypeptide containing only the amino acid encoded by the triplet UUU. A radioactive polypeptide was indeed formed in only one of the 20 tubes, the one containing radioactive phenylalanine. Nirenberg and Matthaei therefore concluded that the triplet codon UUU encodes phenylalanine. The same approach revealed that polycytidylate, poly(C), encodes a polypeptide containing only proline (polyproline), and polyadenylate, poly(A), encodes polylysine. Polyguanylate did not generate any polypeptide in this experiment because it spontaneously forms tetraplexes (see Fig. 8–22) that cannot be bound by ribosomes.


Marshall Nirenberg

The synthetic polynucleotides used in such experiments were prepared with polynucleotide phosphorylase (p. 1020), 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 only, it makes only poly(U). If it is presented with a mixture of five parts ADP and one part CDP, it makes a polymer in which about five-sixths of the residues are adenylate and one-sixth are cytidylate. This random polymer is likely to have many triplets of the sequence AAA, smaller numbers of AAC, ACA, and CAA triplets, relatively few ACC, CCA, and CAC triplets, and very few CCC triplets (Table 27–1). Using a variety of artificial mRNAs made by polynucleotide phosphorylase from different starting mixtures of ADP, GDP, UDP, and CDP, investigators soon identified the base compositions of the triplets coding for almost all the amino acids. Although these experiments revealed the base composition of the coding triplets, they could not reveal the sequence of the bases.

FIGURE 27–4 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; all amino acids coded by the mRNA following the change are affected. Combining insertion and deletion mutations affects some amino acids but can eventually restore 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.



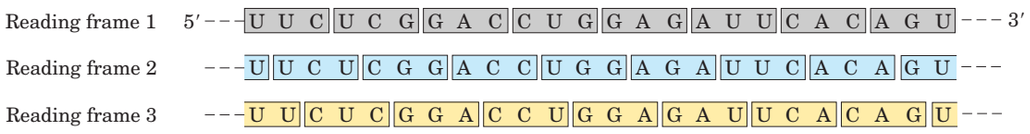


FIGURE 27–5 Reading frames in the genetic code. In a triplet, nonoverlapping code, all mRNAs have three potential reading frames, shaded here in different colors. The triplets, and hence the amino acids specified, are different in each reading frame.

TABLE 27–1 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_2C	20
Glutamine	24	A_2C	20
Histidine	6	AC_2	4
Lysine	100	AAA	100
Proline	7	AC_2, CCC	4.8
Threonine	26	A_2C, AC_2	24

Note: Presented here is a summary of data from one of the early experiments designed to elucidate the genetic code. A synthetic RNA containing only A and C residues in a 5:1 ratio directed polypeptide synthesis, and both the identity and the quantity of incorporated amino acids were determined. Based on the relative abundance of A and C residues in the synthetic RNA, and assigning the codon AAA (the most likely codon) a frequency of 100, there should be three different codons of composition A_2C, each at a relative frequency of 20; three of composition AC_2, each at a relative frequency of 4.0; and CCC at a relative frequency of 0.8. The CCC assignment 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.

*These designations of nucleotide composition contain no information on nucleotide sequence (except, of course, AAA and CCC).

In 1964 Nirenberg and Philip Leder achieved another experimental breakthrough. Isolated *E. coli* ribosomes would bind a specific aminoacyl-tRNA in the presence of the corresponding synthetic polynucleotide messenger. (By convention, the identity of a tRNA is indicated by a superscript, such as $tRNA^{Ala}$, and the aminoacylated tRNA by a hyphenated name: alanyl-$tRNA^{Ala}$ or Ala-$tRNA^{Ala}$.) For example, ribosomes incubated with poly(U) and phenylalanyl-$tRNA^{Phe}$ (Phe-$tRNA^{Phe}$) bind both RNAs, but if the ribosomes are incubated with poly(U) and some other aminoacyl-tRNA, the aminoacyl-tRNA is not bound, because it does not recognize the UUU triplets in poly(U) (Table 27–2). Even trinucleotides could promote specific binding of appropriate tRNAs, so these experiments could be carried out with chemically synthesized small oligonucleotides. With this technique researchers determined which aminoacyl-tRNA bound to about 50 of the 64 possible triplet codons. For some codons, either no aminoacyl-tRNA or more than one would bind. Another method was needed to complete and confirm the entire genetic code.

TABLE 27–2 Trinucleotides That Induce Specific Binding of Aminoacyl-tRNAs to Ribosomes+

	*Relative increase in ^{14}C-labeled aminoacyl-tRNA bound to ribosome**		
Trinucleotide	*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.

*Each number represents the factor by which the amount of bound ^{14}C increased when the indicated trinucleotide was present, relative to a control with no trinucleotide.

H. Gobind Khorana

At about this time, a complementary approach was provided by H. Gobind Khorana, who developed chemical methods to synthesize polyribonucleotides with defined, repeating sequences of two to four bases. The polypeptides produced by these mRNAs 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: ACACACACACACACA. The polypeptide synthesized on this messenger contained equal amounts of threonine and histidine. Given that a histidine codon has one A and two Cs (Table 27–1), CAC must code for histidine and ACA for threonine.

Consolidation of the results from many experiments permitted the assignment of 61 of the 64 possible codons. The other three were identified as termination codons, in part because they disrupted amino acid coding patterns when they occurred in a synthetic RNA polymer (Fig. 27–6). Meanings for all the triplet codons (tabulated in Fig. 27–7) were established by 1966 and have been verified in many different ways. The cracking of the genetic code is regarded as one of the most important scientific discoveries of the twentieth century.

Codons are the key to the translation of genetic information, directing the synthesis of specific proteins. The reading frame is set when translation of an mRNA molecule begins, and it is maintained as the synthetic machinery reads sequentially from one triplet to the next. If the initial reading frame is off by one or two bases, or if translation somehow skips a nucleotide in the mRNA, all the subsequent codons will be out of register; the result is usually a "missense" protein with a garbled amino acid sequence. There are a few unusual but interesting exceptions to this rule (Box 27–1).

Several codons serve special functions (Fig. 27–7). The **initiation codon** AUG is the most common signal for the beginning of a polypeptide in all cells (some rare alternatives are discussed in Box 27–2), in addition to coding for Met residues in internal positions of polypeptides. The **termination codons** (UAA, UAG, and UGA), also called stop codons or nonsense codons, normally signal the end of polypeptide synthesis and do not code for any known amino acids.

As described in Section 27.2, initiation of protein synthesis in the cell is an elaborate process that relies on initiation codons and other signals in the mRNA. In retrospect, the experiments of Nirenberg and Khorana to identify codon function should not have worked in the absence of initiation codons. Serendipitously, experimental conditions caused the normal initiation require-

First letter of codon (5′ end)

Second letter of codon

	U		C		A		G	
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
	UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop
	UUG	Leu	UCG	Ser	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

FIGURE 27–7 "Dictionary" of amino acid code words in mRNAs. The codons are written in the 5′⟶3′ direction. The third base of each codon (in bold type) plays a lesser role in specifying an amino acid than the first two. The three termination codons are shaded in pink, the initiation codon AUG in green. 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 at the third base.

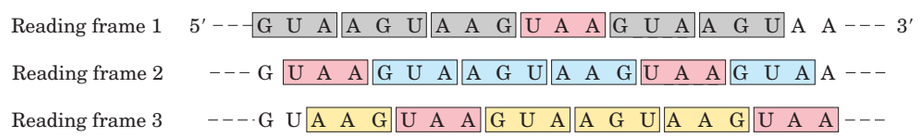


FIGURE 27–6 Effect of a termination codon in a repeating tetranucleotide. Termination codons (pink) are encountered every fourth codon in three different reading frames (shown in different colors). Dipeptides or tripeptides are synthesized, depending on where the ribosome initially binds.

ments for protein synthesis to be relaxed. Diligence combined with chance to produce a breakthrough—a common occurrence in the history of biochemistry.

In a random sequence of nucleotides, 1 in every 20 codons in each reading frame is, on average, a termination codon. In general, a reading frame without a termination codon among 50 or more codons is referred to as an **open reading frame (ORF).** Long open reading frames usually correspond to genes that encode proteins. In the analysis of sequence databases, sophisticated programs are used to search for open reading frames in order to find genes among the often huge background of nongenic DNA. 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.

A striking feature of the genetic code is that an amino acid may be specified by more than one codon, so the code is described as **degenerate.** This does *not* suggest that the code is flawed: although an amino acid may have two or more codons, each codon specifies only one amino acid. The degeneracy of the code is not uniform. Whereas methionine and tryptophan have single codons, for example, three amino acids (Leu, Ser, Arg) have six codons, five amino acids have four, isoleucine has three, and nine amino acids have two (Table 27–3).

The genetic code is nearly universal. With the intriguing exception of a few minor variations in mitochondria, some bacteria, and some single-celled eukaryotes (Box 27–2), amino acid codons are identical in all species examined so far. Human beings, *E. coli,* tobacco plants, amphibians, and viruses share the same genetic code. Thus it would appear that all life forms have a common evolutionary ancestor, whose genetic code has been preserved throughout biological evolution. Even the variations (Box 27–2) reinforce this theme.

TABLE 27–3 Degeneracy of the Genetic Code

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

Wobble Allows Some tRNAs to Recognize More than One Codon

When several different codons specify one amino acid, the difference between them usually lies at the third base position (at the 3′ end). For example, alanine is coded by the triplets GCU, GCC, GCA, and GCG. The codons for most amino acids can be symbolized by $XY\frac{A}{G}$ or $XY\frac{U}{C}$. The first two letters of each codon are the primary determinants of specificity, a feature that has some interesting consequences.

Transfer RNAs base-pair with mRNA codons at a three-base sequence on the tRNA called the **anticodon.** The first base of the codon in mRNA (read in the 5′→3′ direction) pairs with the third base of the anticodon (Fig. 27–8a). If the anticodon triplet of a tRNA recognized only one codon triplet through Watson-Crick base pairing at all three positions, cells would have a different tRNA for each amino acid codon. This is not the case, however, because the anticodons in some tRNAs include the nucleotide inosinate (designated I), which contains the uncommon base hypoxanthine (see Fig. 8–5b). Inosinate can form hydrogen bonds with three different nucleotides (U, C, and A; Fig. 27–8b), although

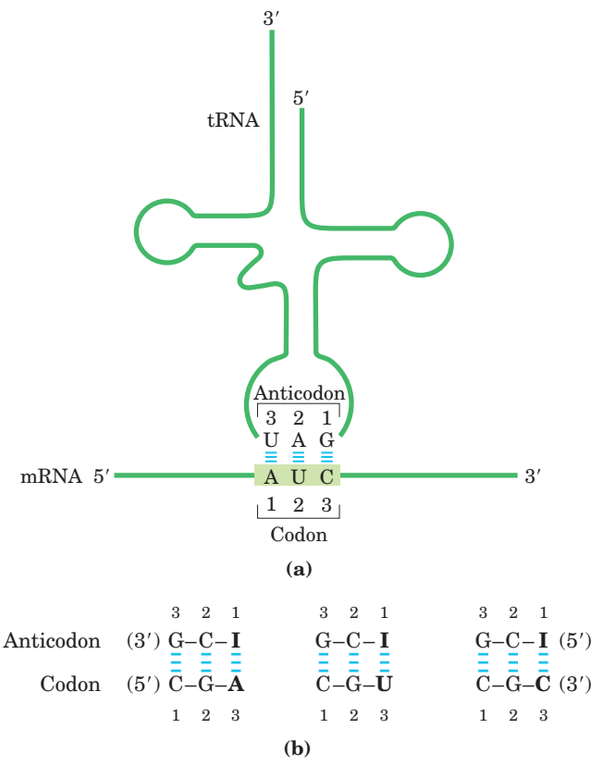


FIGURE 27–8 Pairing relationship of codon and anticodon. (a) Alignment of the two RNAs is antiparallel. The tRNA is shown in the traditional cloverleaf configuration. **(b)** Three different codon pairing relationships are possible when the tRNA anticodon contains inosinate.

BOX 27–1 WORKING IN BIOCHEMISTRY

Changing Horses in Midstream: Translational Frameshifting and mRNA Editing

Once the reading frame has been set during protein synthesis, codons are translated without overlap or punctuation until the ribosomal complex encounters a termination codon. The other two possible reading frames usually contain no useful genetic information, but a few genes are structured so that ribosomes "hiccup" at a certain point in the translation of their mRNAs, changing the reading frame from that point on. This appears to be a mechanism either to allow two or more related but distinct proteins to be produced from a single transcript or to regulate the synthesis of a protein.

One of the best-documented examples occurs in translation of the mRNA for the overlapping *gag* and *pol* genes of the Rous sarcoma virus (see Fig. 26–31). The reading frame for *pol* is offset to the left by one base pair (−1 reading frame) relative to the reading frame for *gag* (Fig. 1).

The product of the *pol* gene (reverse transcriptase) is translated as a larger polyprotein, on the same mRNA that is used for the *gag* protein alone (see Fig. 26–30). The polyprotein, or *gag-pol* protein, is then trimmed to the mature reverse transcriptase by proteolytic digestion. Production of the polyprotein requires a translational frameshift in the overlap region to allow the ribosome to bypass the UAG termination codon at the end of the *gag* gene (shaded pink in Fig. 1).

Frameshifts occur during about 5% of translations of this mRNA, and the *gag-pol* polyprotein (and ultimately reverse transcriptase) is synthesized at about one-twentieth the frequency of the *gag* protein, a level that suffices for efficient reproduction of the virus. In some retroviruses, another translational frameshift allows translation of an even larger polyprotein that includes the product of the *env* gene fused to the *gag* and *pol* gene products (see Fig. 26–30). A similar mechanism produces both the τ and γ subunits of *E. coli* DNA polymerase III from a single *dnaX* gene transcript (see Table 25–2).

This mechanism also occurs in the gene for *E. coli* release factor 2 (RF-2), discussed in Section 27.2, which is required for termination of protein synthesis at the termination codons UAA and UGA. The twenty-sixth codon in the transcript of the gene for RF-2 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. Translation pauses at this codon, but termination does not occur unless RF-2 is bound to the codon (the lower the level of RF-2, the less likely the binding). The absence of bound RF-2 prevents the termination of protein synthesis at UGA and allows time for a frameshift to occur. The UGA plus the C that follows it (UGAC) is therefore read as GAC, which translates to Asp. Translation then proceeds in the new reading frame to complete synthesis of RF-2. In this way, RF-2 regulates its own synthesis in a feedback loop.

Some mRNAs are edited before translation. The initial transcripts of the genes that encode cytochrome oxidase subunit II in some protist mitochondria do not correspond precisely to the sequence needed at the

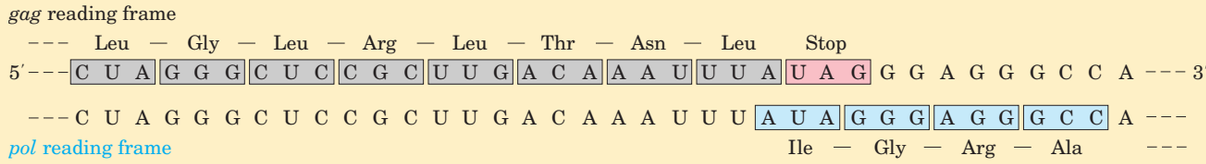


FIGURE 1 The *gag-pol* overlap region in Rous sarcoma virus RNA.

these pairings are much weaker than the hydrogen bonds of Watson-Crick base pairs (G≡C and A═U). In yeast, one $tRNA^{Arg}$ has the anticodon (5′)ICG, which recognizes three arginine codons: (5′)CGA, (5′)CGU, and (5′)CGC. The first two bases are identical (CG) and form strong Watson-Crick base pairs with the corresponding bases of the anticodon, but the third base (A, U, or C) forms 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 anticodon; to use his picturesque word, the third base of such codons (and the first base of their corre-

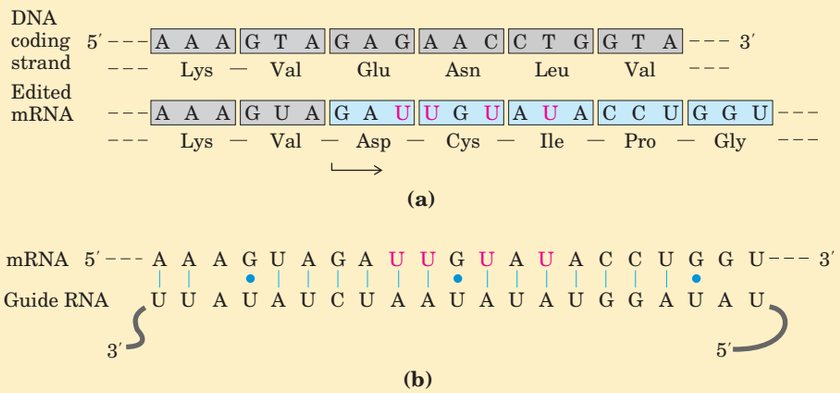


FIGURE 2 RNA editing of the transcript of the cytochrome oxidase subunit II gene from *Trypanosoma brucei* mitochondria. **(a)** Insertion of four U residues (pink) produces a revised reading frame. **(b)** A special class of guide RNAs, complementary to the edited product, may act as templates for the editing process.

carboxyl terminus of the protein product. A posttranscriptional editing process inserts four U residues that shift the translational reading frame of the transcript. Figure 2a shows the added U residues in the small part of the transcript that is affected by editing. Neither the function nor the mechanism of this editing process is understood. Investigators have detected a special class of RNA molecules encoded by these mitochondria, with sequences complementary to the edited mRNAs. These so-called guide RNAs (Fig. 2b) appear to act as templates for the editing process. Note that the base pairing involves a number of G═U base pairs (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 apolipoprotein B, 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 encoded by an mRNA produced from the gene for apoB-100. A cytosine deaminase enzyme found only in the intestine binds to the mRNA at the codon for amino acid residue 2,153 (CAA = Gln) and converts the C to a U, to introduce the termination codon UAA. The apoB-48 produced in the intestine from this modified mRNA is simply an abbreviated form (corresponding to the amino-terminal half) of apoB-100 (Fig. 3). This reaction permits tissue-specific synthesis of two different proteins from one gene.

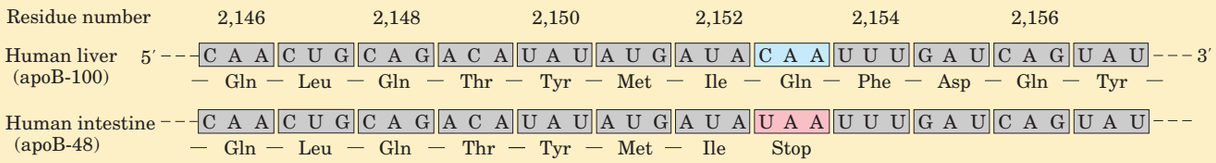


FIGURE 3 RNA editing of the transcript of the gene for the apolipoprotein B-100 component of LDL. Deamination, which occurs only in the intestine, converts a specific cytosine to uracil, changing a Gln codon to a stop codon and producing a truncated protein.

sponding anticodons) "wobbles." Crick proposed a set of four relationships called the **wobble hypothesis:**

1. The first two bases of an mRNA codon always form strong Watson-Crick base pairs with the corresponding bases of the tRNA anticodon and confer most of the coding specificity.
2. The first base of the anticodon (reading in the 5′→3′ direction; this pairs with the third base of the codon) determines the number of codons recognized by the tRNA. When the first base of the anticodon is C or A, base pairing is specific and only one codon is recognized by that tRNA. When the first base is U or G, binding is less

BOX 27–2 WORKING IN BIOCHEMISTRY

Exceptions That Prove the Rule: Natural Variations in the Genetic Code

In biochemistry, as in other disciplines, exceptions to general rules can be problematic for instructors and frustrating for students. At the same time, though, 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.

One would expect little room for variation in the genetic code. Even a single amino acid substitution can have profoundly deleterious effects on the structure of a protein. Nevertheless, variations in the code do occur in some organisms, and they are both interesting and instructive. The types of variation and their rarity provide powerful evidence for a common evolutionary origin of all living things.

To alter the code, changes must occur in one or more tRNAs, with the obvious target for alteration being the anticodon. Such a change would lead to the systematic insertion of an amino acid at a codon that, according to the normal code (see Fig. 27–7), does not specify that amino acid. The genetic code, in effect, is defined by two elements: (1) the anticodons on tRNAs (which determine where an amino acid is placed in a growing polypeptide) and (2) the specificity of the enzymes—the aminoacyl-tRNA synthetases—that charge the tRNAs, which determines the identity of the amino acid attached to a given tRNA.

Most sudden changes in the code would have catastrophic effects on cellular proteins, so code alterations are more likely where relatively few proteins would be affected—such as 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, which do not generally occur *within* genes (see Box 27–4 for exceptions to *this* rule). This pattern is in fact observed.

Of the very few variations in the genetic code that we know of, most occur in mitochondrial DNA (mtDNA), which encodes only 10 to 20 proteins. Mitochondria have their own tRNAs, so their code variations do not affect the much larger cellular genome. The most common changes in mitochondria (and the only code changes that have been observed in cellular genomes) involve termination codons. These changes affect termination in the products of only a subset of genes, and sometimes the effects are minor because the genes have multiple (redundant) termination codons.

In mitochondria, these changes can be viewed as a kind of genomic streamlining. Vertebrate mtDNAs have genes that encode 13 proteins, 2 rRNAs, and 22 tRNAs (see Fig. 19–32). An unusual set of wobble rules allows the 22 tRNAs to decode all 64 possible codon triplets; not all of the 32 tRNAs required for the normal code are needed. Four codon families (in which the amino acid is determined entirely by the first two nucleotides) are decoded by a single tRNA with a U residue in the first (or wobble) position in the anticodon. Either the U pairs somehow with any of the four possible bases in the third position of the codon or a "two out of three" mechanism is used—that is, no base pairing is needed at the third position. 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 (see Table 27–3). If all mitochondrial tRNAs recognize two codons, we would expect additional Met and Trp codons in mitochondria. And we find that the single most common code variation is the normal termination codon UGA specifying tryptophan. The $tRNA^{Trp}$ recognizes and inserts a Trp residue at either UGA or the normal Trp codon, UGG. The second most common variation is conversion of AUA from an Ile codon to a Met codon; the normal Met codon is AUG, and a single tRNA recognizes both codons. 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, oc-

specific and two different codons may be read. When inosine (I) is the first (wobble) nucleotide of an anticodon, three different codons can be recognized—the maximum number for any tRNA. These relationships are summarized in Table 27–4.

3. When an amino acid is specified by several different codons, the 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 (31 to encode the amino acids and 1 for initiation).

curring in the simplest free-living cell, *Mycoplasma capricolum*. Among eukaryotes, the only known extramitochondrial coding changes occur in a few species of ciliated protists, in which both termination codons UAA and UAG can specify glutamine.

Changes in the code need not be absolute; a codon might not always encode the same amino acid. In *E. coli* we find two examples of amino acids being inserted at positions not specified in the normal code. The first is the occasional use of GUG (Val) as an initiation codon. This occurs only for those genes in which the GUG is properly located relative to particular mRNA sequences that affect the initiation of translation (as discussed in Section 27.2).

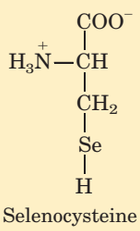
The second *E. coli* example also involves contextual signals that alter coding patterns. A few proteins in all cells (such as formate dehydrogenase in bacteria and glutathione peroxidase in mammals) require the element selenium for their activity, generally in the form of the modified amino acid selenocysteine. Although modified amino acids are generally produced in posttranslational reactions (described in Section 27.3), in *E. coli* selenocysteine is introduced into formate dehydrogenase during translation, in response to an in-frame UGA codon. A special type of serine tRNA, present at lower levels than other Ser-tRNAs, recognizes UGA and no other codons. This tRNA is charged with serine, and the serine is enzymatically converted to selenocysteine before its use at the ribosome. The charged tRNA does not recognize just any UGA codon; some contextual signal in the mRNA, still to be identified, ensures that this tRNA recognizes only the few UGA codons, within certain genes, that specify selenocysteine. In effect, *E. coli* has 21 common amino acids, and UGA doubles as a codon for both termination and (sometimes) selenocysteine.

These variations tell us that the code is not quite as universal as once believed, but that its flexibility is severely constrained. The variations are obviously derivatives of the normal code, and no example of a completely different code has been found. The limited scope of code variants strengthens the principle that all life on this planet evolved on the basis of a single (slightly flexible) genetic code.

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	?	+	+	+	?

*N indicates any nucleotide; +, codon has the same meaning as in the normal code; ?, codon not observed in this mitochondrial genome.



Selenocysteine

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 a codon engaged in strong Watson-Crick pairing with the three bases of the anticodon, tRNAs would dissociate too slowly and this would severely limit the rate of protein synthesis. Codon-anticodon interactions balance the requirements for accuracy and speed.

The genetic code tells us how protein sequence information is stored in nucleic acids and provides some

TABLE 27–4 How the Wobble Base of the Anticodon Determines the Number of Codons a tRNA Can Recognize

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–**A** or **G** (3′)	(5′) Y–X–**C** or **U** (3′)

3. Three codons recognized:

Anticodon	(3′) X–Y–**I** (5′)
Codon	(5′) Y–X–**A** or **U** or **C** (3′)

Note: X and Y denote bases complementary to and capable of strong Watson-Crick base pairing with X′ and Y′, respectively. Wobble bases—in the 3′ position of codons and 5′ position of anticodons—are shaded in pink.

clues about how that information is translated into protein. We now turn to the molecular mechanisms of the translation process.

SUMMARY 27.1 The Genetic Code

- The particular amino acid sequence of a protein is constructed through the translation of information encoded in mRNA. This process is carried out by ribosomes.
- Amino acids are specified by mRNA codons consisting of nucleotide triplets. Translation requires adaptor molecules, the tRNAs, that recognize codons and insert amino acids into their appropriate sequential positions in the polypeptide.
- The base sequences of the codons were deduced from experiments using synthetic mRNAs of known composition and sequence.
- The codon AUG signals initiation of translation. The triplets UAA, UAG, and UGA are signals for termination.
- The genetic code is degenerate: it has multiple code words for almost every amino acid.
- The standard genetic code words are universal in all species, with some minor deviations in mitochondria and a few single-celled organisms.
- The third position in each codon is much less specific than the first and second and is said to wobble.

27.2 Protein Synthesis

As we have seen for DNA and RNA (Chapters 25 and 26), the synthesis of polymeric biomolecules can be considered in terms of initiation, elongation, and termination stages. These fundamental processes are typically bracketed by two additional stages: activation of precursors before synthesis and postsynthetic processing of the completed polymer. Protein synthesis follows the same pattern. The activation of amino acids before their incorporation into polypeptides and the posttranslational processing of the completed polypeptide play particularly important roles in ensuring both the fidelity of synthesis and the proper function of the protein product. The cellular components involved in the five stages of protein synthesis in *E. coli* and other bacteria are listed in Table 27–5; the requirements in eukaryotic cells are quite similar, although the components are in some cases more numerous. An initial overview of the stages of protein synthesis provides a useful outline for the discussion that follows.

Protein Biosynthesis Takes Place in Five Stages

Stage 1: Activation of Amino Acids For the synthesis of a polypeptide with a defined sequence, 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, and (2) a link must be established between each new amino acid and the information in the mRNA that encodes it. Both these requirements are met by attaching the amino acid to a tRNA in the first stage of protein synthesis. Attaching the right amino acid to the right tRNA is critical. This reaction takes place in the cytosol, not on the ribosome. Each of the 20 amino acids is covalently attached to a specific tRNA at the expense of ATP energy, using Mg^{2+}-dependent activating enzymes known as aminoacyl-tRNA synthetases. When attached to their amino acid (aminoacylated) the tRNAs are said to be "charged."

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

Stage 3: Elongation The nascent polypeptide is 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 requires cytosolic proteins known as elongation factors. The binding of each incoming aminoacyl-tRNA and the movement of

TABLE 27–5 Components Required for the Five Major Stages of Protein Synthesis in *E. coli*

Stage	*Essential components*
1. Activation of amino acids	20 amino acids 20 aminoacyl-tRNA synthetases 32 or more tRNAs ATP Mg^{2+}
2. Initiation	mRNA *N*-Formylmethionyl-$tRNA^{fmet}$ 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) GTP Mg^{2+}
4. Termination and release	Termination codon in mRNA Release factors (RF-1, RF-2, RF-3)
5. Folding and posttranslational processing	Specific enzymes, cofactors, and other components for removal of initiating residues and signal sequences, additional proteolytic processing, modification of terminal residues, and attachment of phosphate, methyl, carboxyl, carbohydrate, or prosthetic groups

Masayasu Nomura

the ribosome along the mRNA are facilitated by the hydrolysis of GTP as each residue is added to the growing polypeptide.

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

Stage 5: Folding and Posttranslational Processing In order to achieve its biologically active form, the new polypeptide must fold into its proper three-dimensional conformation. Before or after folding, the new polypeptide may undergo enzymatic processing, including removal of one or more amino acids (usually from the amino terminus); addition of acetyl, phosphoryl, methyl, carboxyl, or other groups to certain amino acid residues; proteolytic cleavage; and/or attachment of oligosaccharides or prosthetic groups.

Before looking at these five stages in detail, we must examine two key components in protein biosynthesis: the ribosome and tRNAs.

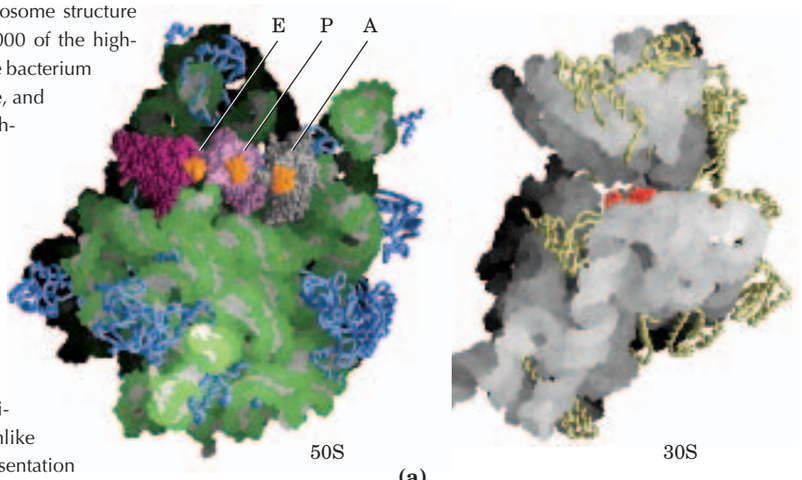
The Ribosome Is a Complex Supramolecular Machine

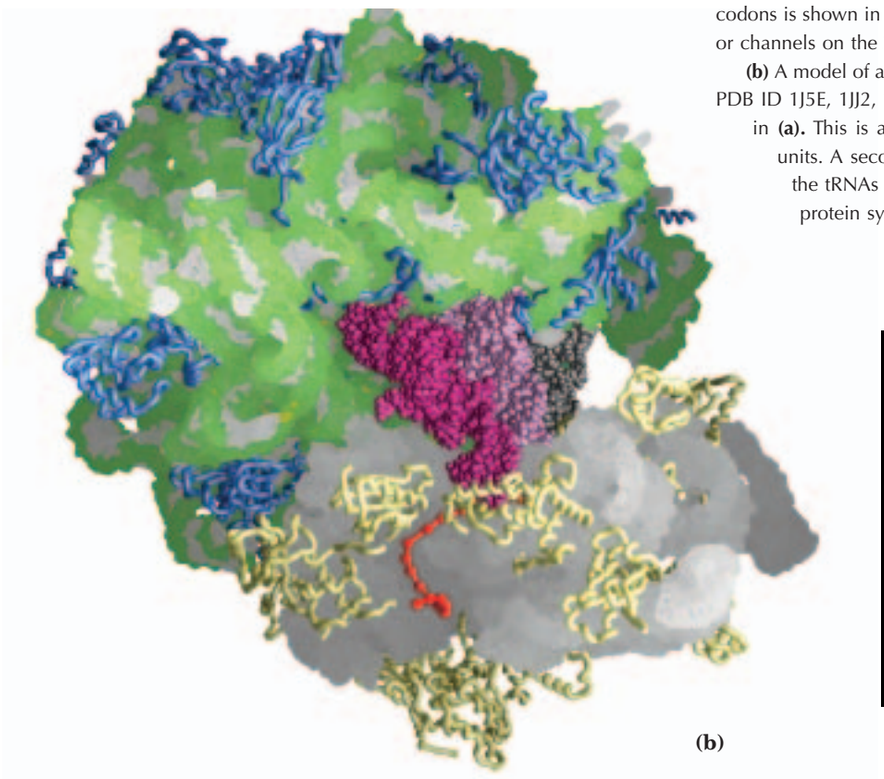
Each *E. coli* cell contains 15,000 or more ribosomes, making up almost a quarter of the dry weight of the cell. Bacterial ribosomes contain about 65% rRNA and 35% protein; they have a diameter of about 18 nm and are composed of two unequal subunits with sedimentation coefficients of 30S and 50S and a combined sedimentation coefficient of 70S. Both subunits contain dozens of ribosomal proteins and at least one large rRNA (Table 27–6).

Following Zamecnik's discovery that ribosomes are the complexes responsible for protein synthesis, and following elucidation of the genetic code, the study of ribosomes accelerated. In the late 1960s Masayasu Nomura and colleagues demonstrated that both ribosomal subunits can be broken down into their RNA and protein components, then reconstituted in vitro. Under appropriate experimental conditions, the RNA and protein spontaneously reassemble to form 30S or 50S subunits nearly identical in structure and activity to native subunits. This breakthrough fueled decades of research into

the function and structure of ribosomal RNAs and proteins. At the same time, increasingly sophisticated structural methods revealed more and more details about ribosome structure.

The dawn of a new millennium brought with it the elucidation of the first high-resolution structures of bacterial ribosomal subunits. The bacterial ribosome is complex, with a combined molecular weight of ~2.7 million, and it is providing a wealth of surprises (Fig. 27–9). First, the traditional focus on the protein components of ribosomes was shifted. The ribosomal subunits are huge RNA molecules. In the 50S subunit, the 5S and 23S rRNAs form the structural core. The proteins are secondary elements in the complex, decorating the surface. Second and most important, there is no protein within 18 Å of the active site for peptide bond formation. The high-resolution structure thus confirms what many had suspected for more than a decade: the ribosome is a ribozyme. In addition to the insight they provide into the mechanism of protein synthesis (as elaborated below), the detailed





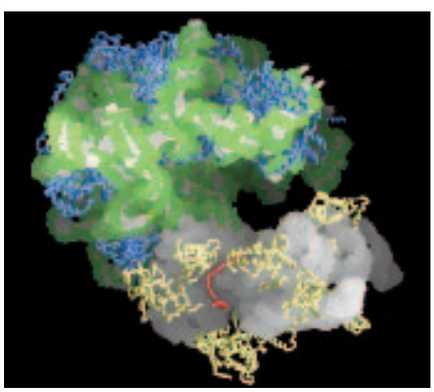


FIGURE 27–9 **Ribosomes.** Our understanding of ribosome structure took a giant step forward with the publication in 2000 of the high-resolution structure of the 50S ribosomal subunit of the bacterium *Haloarcula marismortui* by Thomas Steitz, Peter Moore, and their colleagues. This was followed by additional high-resolution structures of the ribosomal subunits from several different bacterial species, and models of the corresponding complete ribosomes. A sampling of that progress is presented here.

(a) The 50S and 30S bacterial subunits, split apart to visualize the surfaces that interact in the active ribosome. The structure on the left is the 50S subunit (derived from PDB ID 1JJ2 and 1GIY), with tRNAs (purple, mauve, and gray); bound to sites E, P, and A, described later in the text; the tRNA anticodons are in orange. Proteins appear as blue wormlike structures; the rRNA as a blended space-filling representation designed to highlight surface features, with the bases in white and the backbone in green. The structure on the right is the 30S subunit (derived from PDB ID 1J5E and 1JGO). Proteins are yellow and the rRNA white. The part of the mRNA that interacts with the tRNA anticodons is shown in red. The rest of the mRNA winds through grooves or channels on the 30S subunit surface.

(b) A model of a complete active bacterial ribosome (derived from PDB ID 1J5E, 1JJ2, 1JGO, and 1GIY). All components are colored as in **(a).** This is a view down into the groove separating the subunits. A second view (inset) is from the same angle, but with the tRNAs removed to give a better sense of the cleft where protein synthesis occurs.

TABLE 27–6 RNA and Protein Components of the *E. coli* Ribosome

Subunit	*Number of different proteins*	*Total number of proteins*	*Protein designations*	*Number and type of rRNAs*
30S	21	21	S1–S21	1 (16S rRNA)
50S	33	36	L1–L36*	2 (5S and 23S rRNAs)

*The L1 to L36 protein designations do not correspond to 36 different proteins. The protein originally designated L7 is in fact a modified form of L12, and L8 is a complex of three other proteins. Also, L26 proved to be the same protein as S20 (and not part of the 50S subunit). This gives 33 different proteins in the large subunit. There are four copies of the L7/L12 protein, with the three extra copies bringing the total protein count to 36.

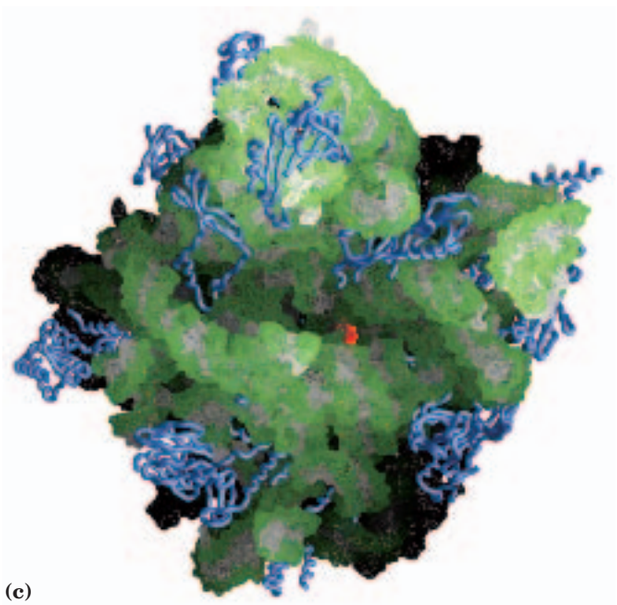
structures of the ribosome and its subunits have stimulated a new look at the evolution of life (Box 27–3).

The two irregularly shaped ribosomal subunits fit together to form a cleft through which the mRNA passes as the ribosome moves along it during translation (Fig. 27–9b). The 55 proteins in bacterial ribosomes vary enormously in size and structure. Molecular weights range from about 6,000 to 75,000. Most of the proteins have globular domains arranged on the ribosome surface. Some also have snakelike protein extensions that protrude into the rRNA core of the ribosome, stabilizing its structure. The functions of some of these proteins have not yet been elucidated in detail, although a structural role seems evident for many of them.

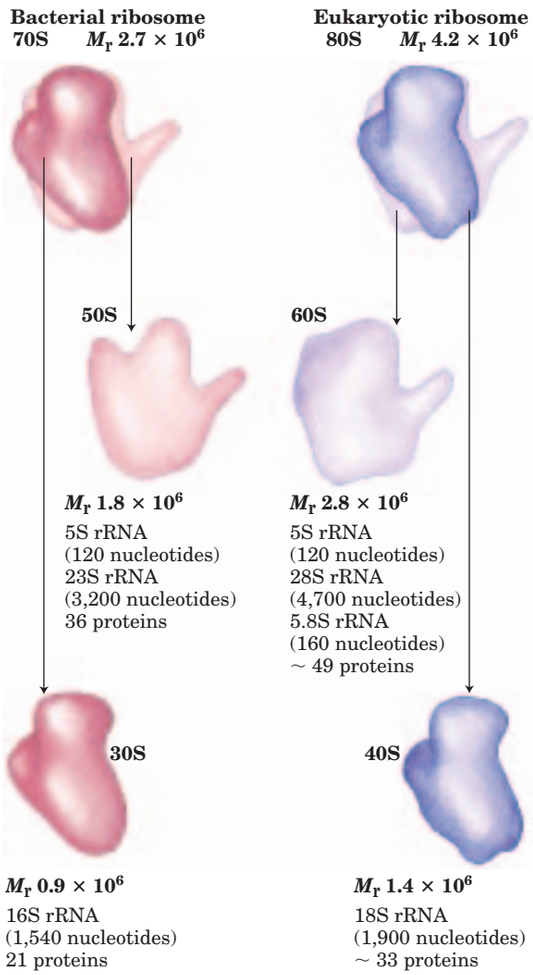
The sequences of the rRNAs of many organisms are now known. Each of the three single-stranded rRNAs of

(c) Structure of the 50S bacterial ribosome subunit (PDB ID 1Q7Y). The subunit is again viewed from the side that attaches to the 30S subunit, but is tilted down slightly compared to its orientation in **(a).** The active site for peptide bond formation (the peptidyl transferase activity), deep within a surface groove and far away from any protein, is marked by a bound inhibitor, puromycin (red).

(d) Summary of the composition and mass of ribosomes in prokaryotes and eukaryotes. Ribosomal subunits are identified by their S (Svedberg unit) values, sedimentation coefficients that refer to their rate of sedimentation in a centrifuge. The S values are not necessarily additive when subunits are combined, because rates of sedimentation are affected by shape as well as mass.



(c)



(d)

BOX 27–3 THE WORLD OF BIOCHEMISTRY

From an RNA World to a Protein World

Extant ribozymes generally promote one of two types of reactions: hydrolytic cleavage of phosphodiester bonds or phosphoryl transfers (Chapter 26). In both cases, the substrates of the reactions are also RNA molecules. The ribosomal RNAs provide an important expansion of the catalytic range of known ribozymes. Coupled to the laboratory exploration of potential RNA catalytic function (see Box 26–3), the idea of an RNA world as a precursor to current life forms becomes increasingly attractive.

A viable RNA world would require an RNA capable of self-replication, a primitive metabolism to generate the needed ribonucleotide precursors, and a cell boundary to aid in concentrating the precursors and sequestering them from the environment. The requirements for catalysis of reactions involving a growing range of metabolites and macromolecules could have led to larger and more complex RNA catalysts. The many negatively charged phosphoryl groups in the RNA backbone limit the stability of very large RNA molecules. In an RNA world, divalent cations or other positively charged groups could be incorporated into the structures to augment stability.

Certain peptides could stabilize large RNA molecules. For example, many ribosomal proteins in modern eukaryotic cells have long extensions, lacking secondary structure, that snake into the rRNAs and help stabilize them (Fig. 1). Ribozyme-catalyzed synthesis of peptides could thus initially have evolved as part of a general solution to the structural maintenance of large RNA molecules. The synthesis of peptides may have helped stabilize large ribozymes, but this advance also marked the beginning of the end for the RNA world. Once peptide synthesis was possible, the greater catalytic potential of proteins would have set in motion an irreversible transition to a protein-dominated metabolic system.

Most enzymatic processes, then, were eventually surrendered to the proteins—but not all. In every organism, the critical task of synthesizing the proteins remains, even now, a ribozyme-catalyzed process. There appears to be only one good arrangement (or just a very few) of nucleotide residues in a ribozyme active site that can catalyze peptide synthesis. The rRNA residues that seem to be involved in the peptidyl transferase activity of ribosomes are highly conserved in the large-subunit rRNAs of all species. Using in vitro evolution (SELEX; see Box 26–3), investigators have isolated artificial ribozymes that promote peptide synthesis. Intriguingly, most of them include the ribonucleotide octet (5′)AUAACAGG(3′), a highly conserved sequence found at the peptidyl transferase active site in the ribosomes of all cells. There may be just one optimal solution to the overall chemical problem of ribozyme-catalyzed synthesis of proteins of defined sequence. Evolution found this solution once, and no life form has notably improved on it.

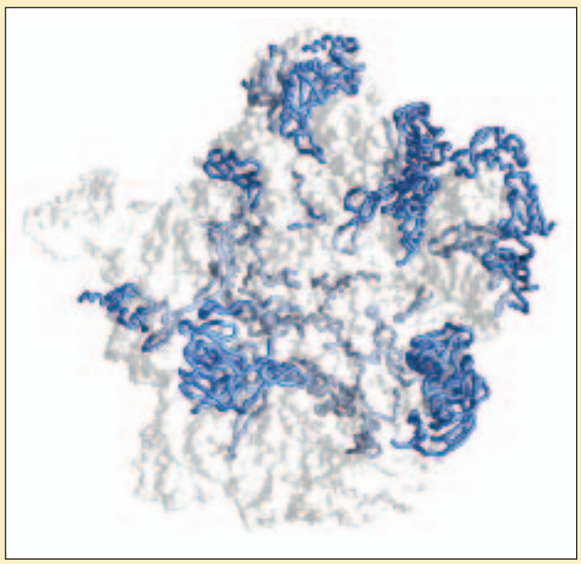


FIGURE 1 The 50S subunit of a bacterial ribosome (PDB ID 1NKW). The protein backbones are shown as blue wormlike structures; the rRNA components are transparent. The unstructured extensions of many of the ribosomal proteins snake into the rRNA structures, helping to stabilize them.

E. coli has a specific three-dimensional conformation featuring extensive intrachain base pairing. The predicted secondary structure of the rRNAs (Fig. 27–10) has largely been confirmed in the high-resolution models, but fails to convey the extensive network of tertiary interactions evident in the complete structure.

The ribosomes of eukaryotic cells (other than mitochondrial and chloroplast ribosomes) are larger and more complex than bacterial ribosomes (Fig. 27–9d), with a diameter of about 23 nm and a sedimentation coefficient of about 80S. They also have two subunits, which vary in size among species but on average are 60S

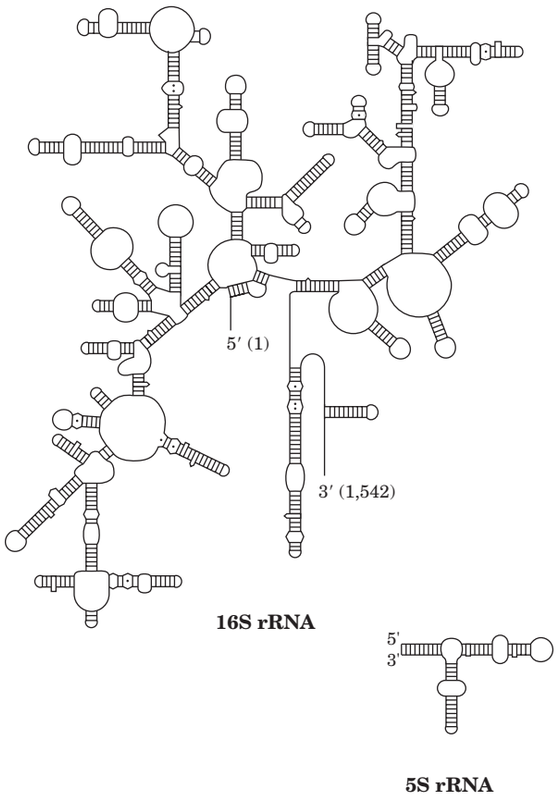


FIGURE 27–10 Bacterial rRNAs. Diagrams of the secondary structure of *E. coli* 16S and 5S rRNAs. The first (5′ end) and final (3′ end) ribonucleotide residues of the 16S rRNA are numbered.

and 40S. Altogether, eukaryotic ribosomes contain more than 80 different proteins. The ribosomes of mitochondria and chloroplasts are somewhat smaller and simpler than bacterial ribosomes. Nevertheless, ribosomal structure and function are strikingly similar in all organisms and organelles.

Transfer RNAs Have Characteristic Structural Features

To understand how tRNAs can serve as adaptors in translating the language of nucleic acids into the language of proteins, we must first examine their structure in more detail. Transfer RNAs are relatively small and consist of a single strand of RNA folded into a precise three-dimensional structure (see Fig. 8–28a). The tRNAs in bacteria and in the cytosol of eukaryotes have between 73 and 93 nucleotide residues, corresponding to molecular weights of 24,000 to 31,000. Mitochondria and chloroplasts contain distinctive, somewhat smaller tRNAs. Cells have at least one kind of tRNA for each amino acid; at least 32 tRNAs are required to recognize all the amino acid codons (some recognize more than one codon), but some cells use more than 32.

Yeast alanine tRNA ($tRNA^{Ala}$), the first nucleic acid to be completely sequenced (Fig. 27–11), contains 76 nucleotide residues, 10 of which have modified bases. Comparisons of tRNAs from various species have revealed many common denominators of structure (Fig. 27–12). Eight or more of the nucleotide residues have modified bases and sugars, 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. When

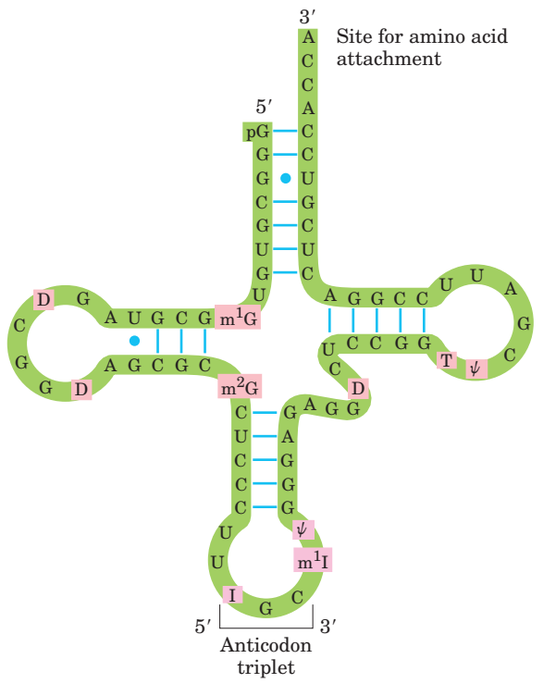


FIGURE 27–11 Nucleotide sequence of yeast $tRNA^{Ala}$. This structure was deduced in 1965 by Robert W. Holley and his colleagues; it is shown in the cloverleaf conformation in which intrastrand base pairing is maximal. The following symbols are used for the modified nucleotides (shaded pink): ψ, pseudouridine; I, inosine; T, ribothymidine; D, 5,6-dihydrouridine; m^1I, 1-methylinosine; m^1G, 1-methylguanosine; m^2G, N^2-dimethylguanosine (see Fig. 26–24). Blue lines between parallel sections indicate Watson-Crick base pairs. The anticodon can recognize three codons for alanine (GCA, GCU, and GCC). Other features of tRNA structure are shown in Figures 27–12 and 27–13. Note the presence of two G═U base pairs, signified by a blue dot to indicate non-Watson-Crick pairing. In RNAs, guanosine is often base-paired with uridine, although the G═U pair is not as stable as the Watson-Crick G≡C pair (Chapter 8).

Robert W. Holley, 1922–1993

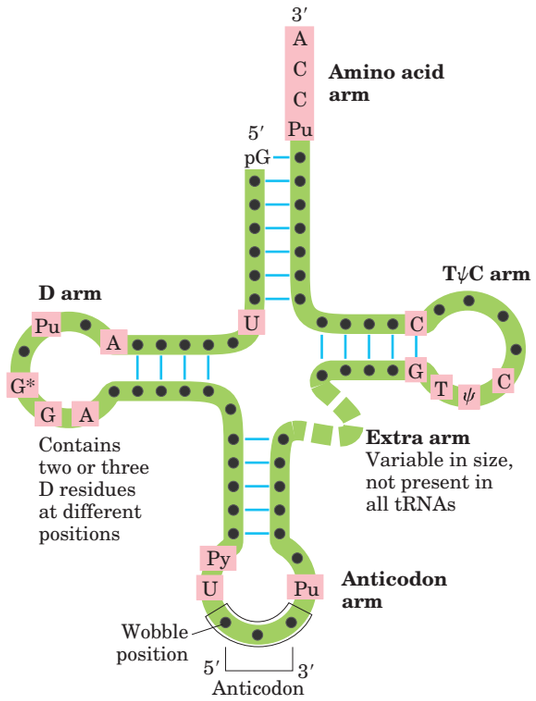
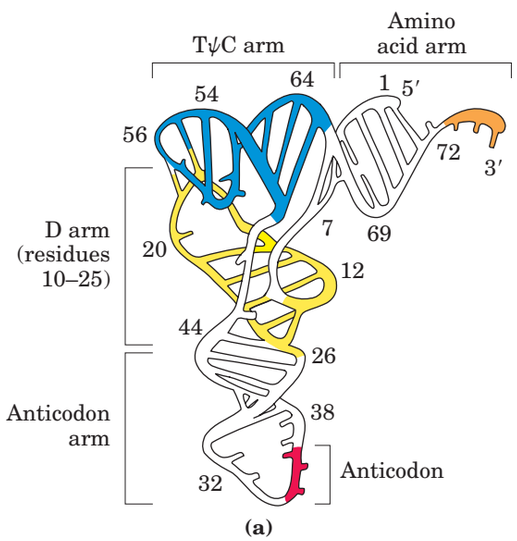


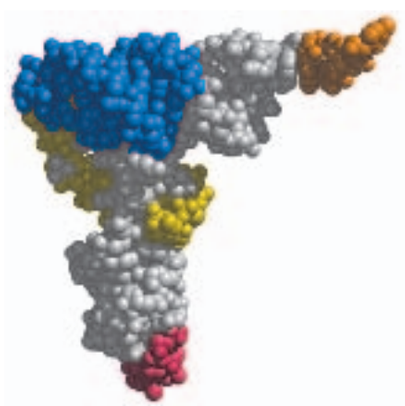
FIGURE 27–12 General cloverleaf secondary structure of tRNAs. The large dots on the backbone represent nucleotide residues; the blue lines represent base pairs. Characteristic and/or invariant residues common to all tRNAs are shaded in pink. Transfer RNAs vary in length from 73 to 93 nucleotides. Extra nucleotides occur in the extra arm or in the D arm. At the end of the anticodon arm is the anticodon loop, which always contains seven unpaired nucleotides. The D arm contains two or three D (5,6-dihydrouridine) residues, depending on the tRNA. In some tRNAs, the D arm has only three hydrogen-bonded base pairs. In addition to the symbols explained in Figure 27–11: Pu, purine nucleotide; Py, pyrimidine nucleotide; G*, guanylate *or* 2′-*O*-methylguanylate.

drawn in two dimensions, the hydrogen-bonding pattern of all tRNAs forms a cloverleaf structure with four arms; the longer tRNAs have a short fifth arm, or extra arm (Fig. 27–12). In three dimensions, a tRNA has the form of a twisted L (Fig. 27–13).

Two of the arms of a tRNA are critical for its adaptor function. The **amino acid arm** can carry a specific amino acid esterified by its carboxyl group to the 2′- or 3′-hydroxyl group of the A residue at the 3′ end of the tRNA. The **anticodon arm** contains the anticodon. The other major arms are the **D arm,** which contains the unusual nucleotide dihydrouridine (D), 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 ribose (see Fig. 26–24). The D and TψC arms contribute important



(a)



(b)

FIGURE 27–13 Three-dimensional structure of yeast tRNA^{Phe} deduced from x-ray diffraction analysis. The shape resembles a twisted L. **(a)** Schematic diagram with the various arms identified in Figure 27–12 shaded in different colors. **(b)** A space-filling model, with the same color coding (PDB ID 4TRA).The CCA sequence at the 3′ end (orange) is the attachment point for the amino acid.

interactions for the overall folding of tRNA molecules, and the TψC arm interacts with the large-subunit rRNA.

Having looked at the structures of ribosomes and tRNAs, we now consider in detail the five stages of protein synthesis.

Stage 1: Aminoacyl-tRNA Synthetases Attach the Correct Amino Acids to Their tRNAs

During the first stage of protein synthesis, taking place in the cytosol, aminoacyl-tRNA synthetases esterify the 20 amino acids to their corresponding tRNAs. Each enzyme is specific for one amino acid and one or more corresponding tRNAs. Most organisms have one aminoacyl-tRNA synthetase for each amino acid. For amino acids with two or more corresponding tRNAs, the same enzyme usually aminoacylates all of them.

The structures of all the aminoacyl-tRNA synthetases of *E. coli* have been determined. Researchers have divided them into two classes (Table 27–7) based on substantial differences in primary and tertiary structure and in reaction mechanism (Fig. 27–14); these two classes are the same in all organisms. There is no evidence for a common ancestor, and the biological, chemical, or evolutionary reasons for two enzyme classes for essentially identical processes remain obscure.

The reaction catalyzed by an aminoacyl-tRNA synthetase is

$$\text{Amino acid} + \text{tRNA} + \text{ATP} \overset{Mg^{2+}}{\rightleftharpoons} \text{aminoacyl-tRNA} + \text{AMP} + PP_i$$

This reaction occurs in two steps in the enzyme's active site. In step (1) (Fig. 27–14) an enzyme-bound intermediate, aminoacyl adenylate (aminoacyl-AMP), forms when the carboxyl group of the amino acid reacts with the α-phosphoryl group of ATP to form an anhydride linkage, with displacement of pyrophosphate. In the second step the aminoacyl group is transferred from enzyme-bound aminoacyl-AMP to its corresponding specific tRNA. The course of this second step depends on the class to which the enzyme belongs, as shown by pathways (2a) and (2b) in Figure 27–14. The resulting ester linkage between the amino acid and the tRNA (Fig. 27–15) has a highly negative standard free energy of hydrolysis ($\Delta G'^\circ = -29$ 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:

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

TABLE 27–7 The Two Classes of Aminoacyl-tRNA Synthetases

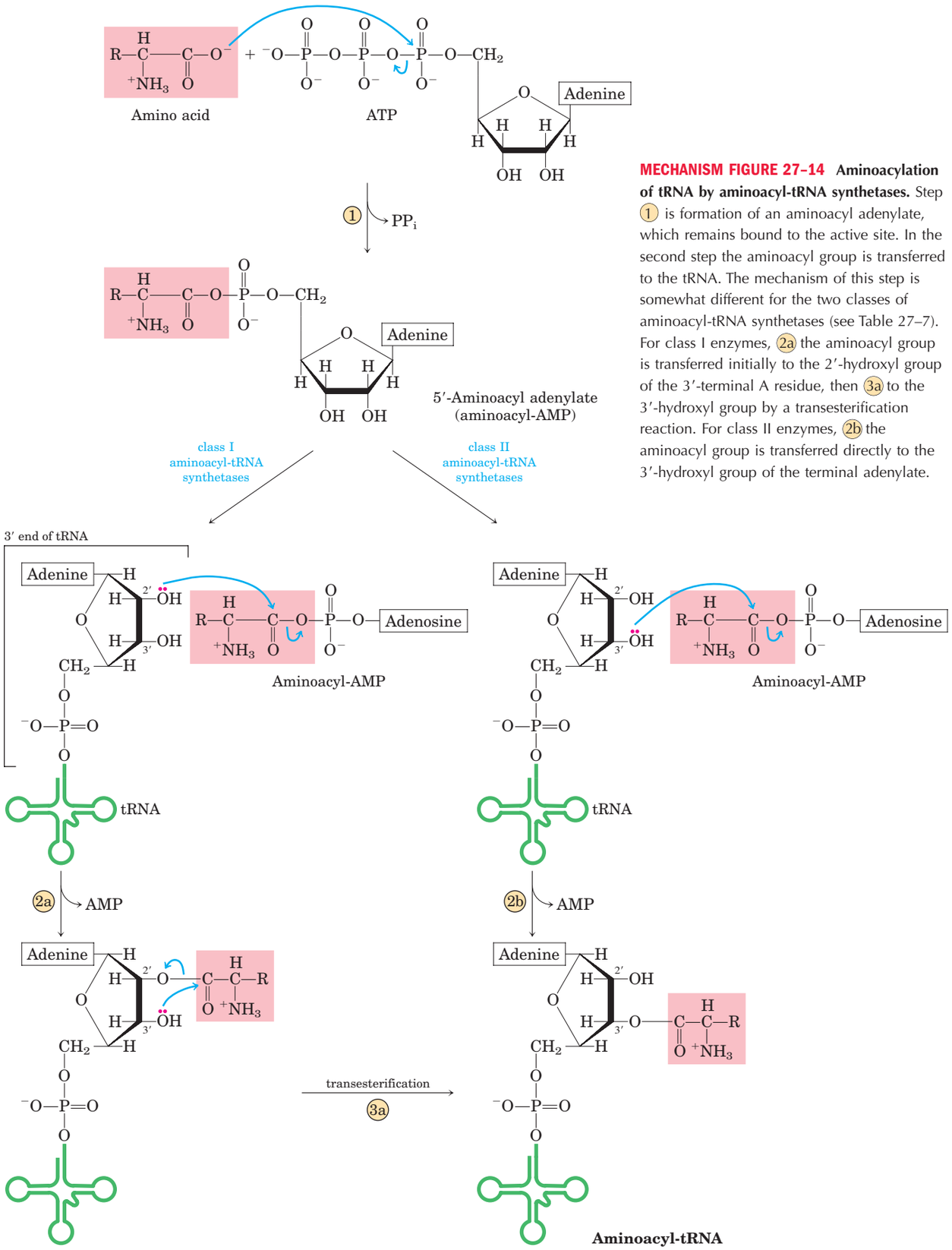
Class I		*Class II*	
Arg	Leu	Ala	Lys
Cys	Met	Asn	Phe
Gln	Trp	Asp	Pro
Glu	Tyr	Gly	Ser
Ile	Val	His	Thr

Note: Here, Arg represents arginyl-tRNA synthetase, and so forth. The 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 27–14.

Proofreading by Aminoacyl-tRNA Synthetases The aminoacylation of tRNA accomplishes two ends: (1) activation of an amino acid for peptide bond formation and (2) attachment of the amino acid to an adaptor tRNA that ensures appropriate placement of the amino acid in a growing polypeptide. The identity of the amino acid attached to a tRNA is not checked on the ribosome, so attachment of the correct amino acid to the tRNA is essential to the fidelity of protein synthesis.

As you will recall from Chapter 6, enzyme specificity is limited by the binding energy available from enzyme-substrate interactions. Discrimination between two similar amino acid substrates has been studied in detail in the case of Ile-tRNA synthetase, which distinguishes between valine and isoleucine, amino acids that differ by only a single methylene group (—CH_2—). Ile-tRNA synthetase favors activation of isoleucine (to form Ile-AMP) over valine by a factor of 200—as we would expect, given the amount by which a methylene group (in Ile) could enhance substrate binding. Yet valine is erroneously incorporated into proteins in positions normally occupied by an Ile residue at a frequency of only about 1 in 3,000. How is this greater than tenfold increase in accuracy brought about? Ile-tRNA synthetase, like some other aminoacyl-tRNA synthetases, has a proofreading function.

Recall a general principle from the discussion of proofreading by DNA polymerases (p. 955): if available binding interactions do not provide sufficient discrimination between two substrates, the necessary specificity can be achieved by substrate-specific binding in *two successive* steps. The effect of forcing the system through two successive filters is multiplicative. In the case of Ile-tRNA synthetase, the first filter is the initial binding of the amino acid to the enzyme and its activation to aminoacyl-AMP. The second is the binding of any



MECHANISM FIGURE 27–14 Aminoacylation of tRNA by aminoacyl-tRNA synthetases. Step (1) is formation of an 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 27–7). For class I enzymes, (2a) the aminoacyl group is transferred initially to the 2′-hydroxyl group of the 3′-terminal A residue, then (3a) to the 3′-hydroxyl group by a transesterification reaction. For class II enzymes, (2b) the aminoacyl group is transferred directly to the 3′-hydroxyl group of the terminal adenylate.

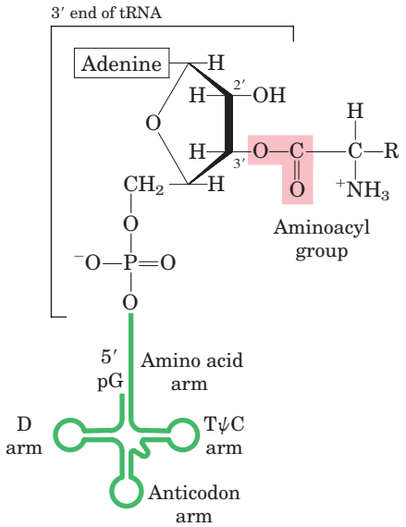
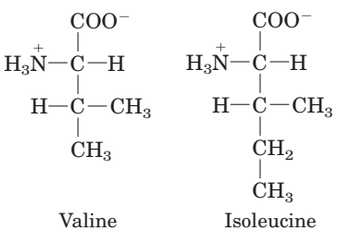


FIGURE 27–15 General structure of aminoacyl-tRNAs. The aminoacyl group is esterified to the 3′ position of the terminal A residue. The ester linkage that both activates the amino acid and joins it to the tRNA is shaded pink.

incorrect aminoacyl-AMP products to a separate active site on the enzyme; a substrate that binds in this second active site is hydrolyzed. The R group of valine is slightly smaller than that of isoleucine, so Val-AMP fits the hydrolytic (proofreading) site of the Ile-tRNA synthetase but Ile-AMP does not. Thus Val-AMP is hydrolyzed to valine and AMP in the proofreading active site, and tRNA bound to the synthetase does not become aminoacylated to the wrong amino acid.



In addition to proofreading after formation of the aminoacyl-AMP intermediate, most aminoacyl-tRNA synthetases can also hydrolyze the ester linkage between amino acids and tRNAs in the aminoacyl-tRNAs. This hydrolysis is greatly accelerated for incorrectly charged tRNAs, providing yet a third filter to enhance the fidelity of the overall process. The few aminoacyl-tRNA synthetases that activate amino acids with no close structural relatives (Cys-tRNA synthetase, for example) demonstrate little or no proofreading activity; in these cases, the active site for aminoacylation can sufficiently discriminate between the proper substrate and any incorrect amino acid.

The overall error rate of protein synthesis (~1 mistake per 10^4 amino acids incorporated) is not nearly as low as that of DNA replication. Because flaws in a protein are eliminated when the protein is degraded and are not passed on to future generations, they have less biological significance. The degree of fidelity in protein synthesis 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. One defective protein molecule is usually unimportant when many correct copies of the same protein are present.

Interaction between an Aminoacyl-tRNA Synthetase and a tRNA: A "Second Genetic Code" An individual aminoacyl-tRNA synthetase must be specific not only for a single amino acid but for certain tRNAs as well. Discriminating among dozens of 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," reflecting its critical role in maintaining the accuracy of protein synthesis. The "coding" rules appear to be more complex than those in the "first" code.

Figure 27–16 summarizes what we know about the nucleotides involved in recognition by some aminoacyl-tRNA synthetases. Some nucleotides are conserved in all tRNAs and therefore cannot be used for discrimination.

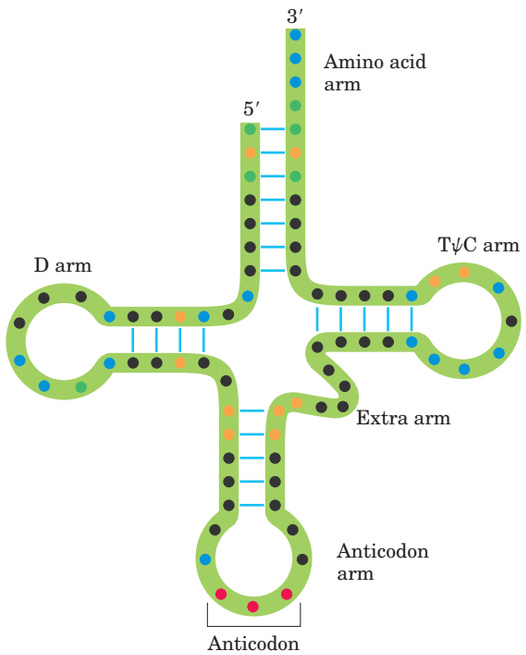
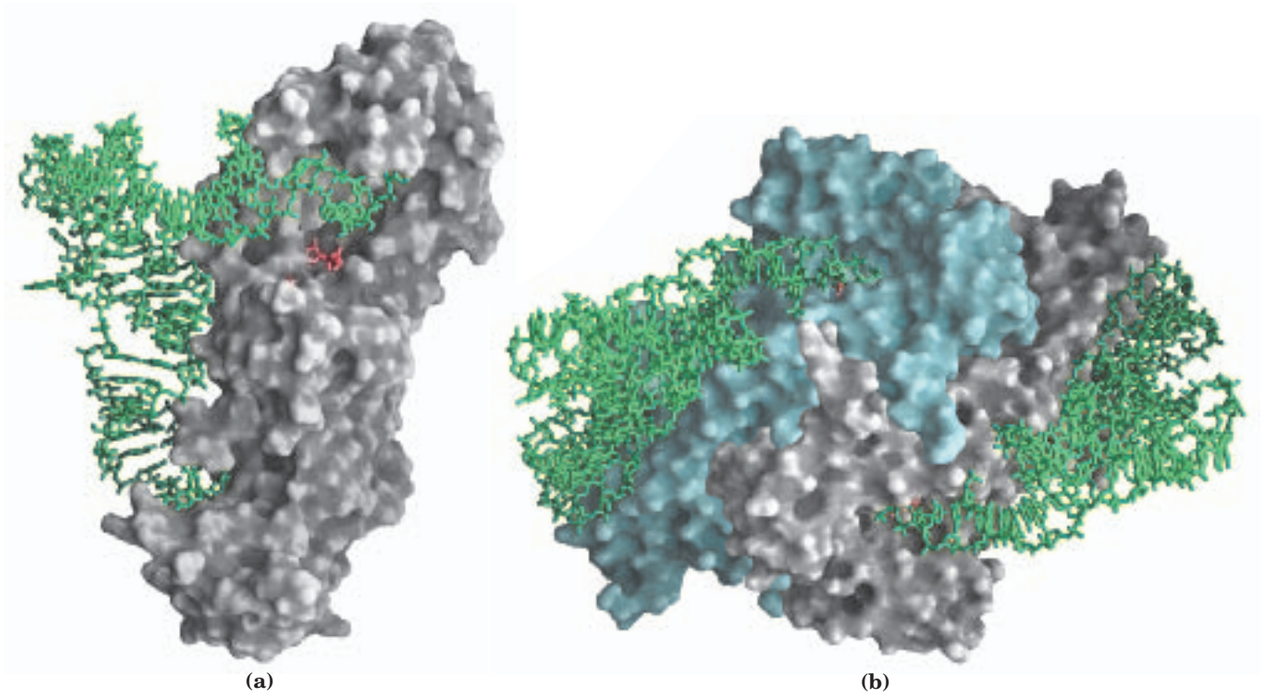


FIGURE 27–16 Nucleotide positions in tRNAs that are recognized by aminoacyl-tRNA synthetases. Some positions (blue dots) are the same in all tRNAs and therefore cannot be used to discriminate one from another. Other positions are known recognition points for one (orange) or more (green) aminoacyl-tRNA synthetases. Structural features other than sequence are important for recognition by some of the synthetases.



(a) (b)

FIGURE 27–17 Aminoacyl-tRNA synthetases. Both synthetases are complexed with their cognate tRNAs (green stick structures). Bound ATP (red) pinpoints the active site near the end of the aminoacyl arm. **(a)** Gln-tRNA synthetase from *E. coli,* a typical monomeric type I synthetase (PDB ID 1QRT). **(b)** Asp-tRNA synthetase from yeast, a typical dimeric type II synthetase (PDB ID 1ASZ).

By observing changes in nucleotides that alter substrate specificity, researchers have identified nucleotide positions that are involved in discrimination by the aminoacyl-tRNA synthetases. These nucleotide positions seem to be concentrated in the amino acid arm and the anticodon arm, including the nucleotides of the anticodon itself, but are also located in other parts of the tRNA molecule. Determination of the crystal structures of aminoacyl-tRNA synthetases complexed with their cognate tRNAs and ATP has added a great deal to our understanding of these interactions (Fig. 27–17).

Ten or more specific nucleotides may be involved in recognition of a tRNA by its specific aminoacyl-tRNA synthetase. But in a few cases the recognition mechanism is quite simple. Across a range of organisms from bacteria to humans, the primary determinant of tRNA recognition by the Ala-tRNA synthetases is a single G═U base pair in the amino acid arm of $tRNA^{Ala}$ (Fig. 27–18a). A short RNA with as few as 7 bp arranged in a simple hairpin minihelix is efficiently aminoacylated by the Ala-tRNA synthetase, as long as the RNA contains the critical G═U (Fig. 27–18b). This relatively simple alanine system may be an evolutionary relic of a period when RNA oligonucleotides, ancestors to tRNA, were aminoacylated in a primitive system for protein synthesis.

Stage 2: A Specific Amino Acid Initiates Protein Synthesis

Protein synthesis begins at the amino-terminal end and proceeds by the stepwise addition of amino acids to the carboxyl-terminal end of the growing polypeptide, as determined by Howard Dintzis in 1961 (Fig. 27–19). The AUG initiation codon thus specifies an *amino-terminal* methionine residue. Although methionine has only one codon, (5′)AUG, all organisms have two tRNAs for methionine. One is used exclusively when (5′)AUG is the initiation codon for protein synthesis. The other is used to code for a Met residue in an internal position in a polypeptide.

The distinction between an initiating (5′)AUG and an internal one is straightforward. In bacteria, the two types of tRNA specific for methionine are designated $tRNA^{Met}$ and $tRNA^{fMet}$. The amino acid incorporated in response to the (5′)AUG initiation codon is *N*-formylmethionine (fMet). It arrives at the ribosome as *N*-formylmethionyl-$tRNA^{fMet}$ (fMet-$tRNA^{fMet}$), which is formed in two successive reactions. First, methionine is attached to $tRNA^{fMet}$ by the Met-tRNA synthetase (which in *E. coli* aminoacylates both $tRNA^{fMet}$ and $tRNA^{Met}$):

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

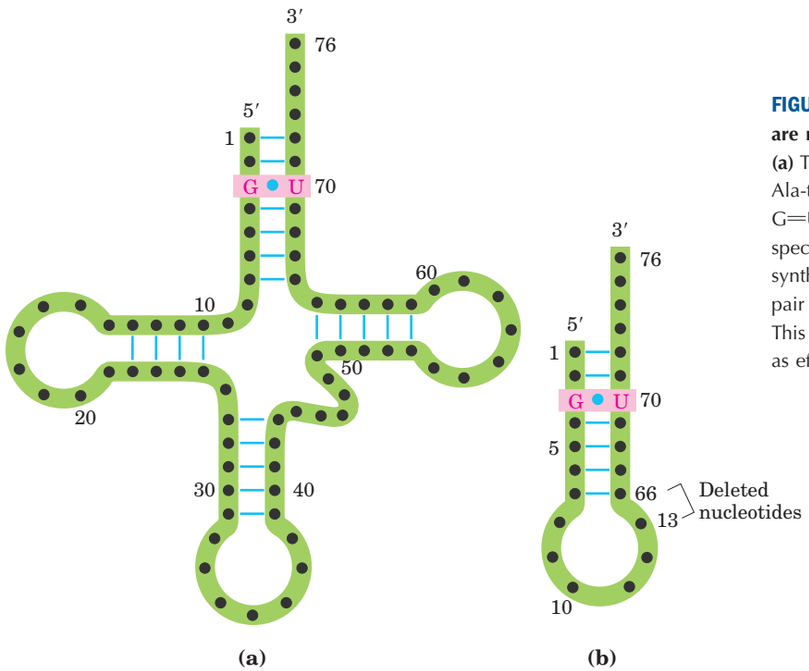


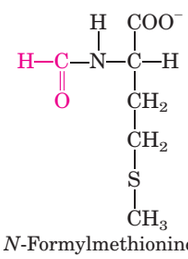
FIGURE 27–18 Structural elements of $tRNA^{Ala}$ that are required for recognition by Ala-tRNA synthetase. **(a)** The $tRNA^{Ala}$ structural elements recognized by the Ala-tRNA synthetase are unusually simple. A single G═U base pair (pink) is the only element needed for specific binding and aminoacylation. **(b)** A short synthetic RNA minihelix, with the critical G═U base pair but lacking most of the remaining tRNA structure. This is specifically aminoacylated with alanine almost as efficiently as the complete $tRNA^{Ala}$.

Next, a transformylase transfers a formyl group from N^{10}-formyltetrahydrofolate to the amino group of the Met residue:

$$N^{10}\text{-Formyltetrahydrofolate} + \text{Met-tRNA}^{fMet} \longrightarrow \text{tetrahydrofolate} + \text{fMet-tRNA}^{fMet}$$

The transformylase is more selective than the Met-tRNA synthetase; it is specific for Met residues attached to $tRNA^{fMet}$, presumably recognizing some unique structural feature of that tRNA. By contrast, Met-$tRNA^{Met}$ inserts methionine in interior positions in polypeptides.

Addition of the *N*-formyl group to the amino group of methionine by the transformylase prevents fMet from entering interior positions in a polypeptide while also allowing fMet-$tRNA^{fMet}$ to be bound at a specific ribosomal initiation site that accepts neither Met-$tRNA^{Met}$ nor any other aminoacyl-tRNA.



N-Formylmethionine

In eukaryotic cells, all polypeptides synthesized by cytosolic ribosomes begin with a Met residue (rather than fMet), but, again, the cell uses a specialized initiating

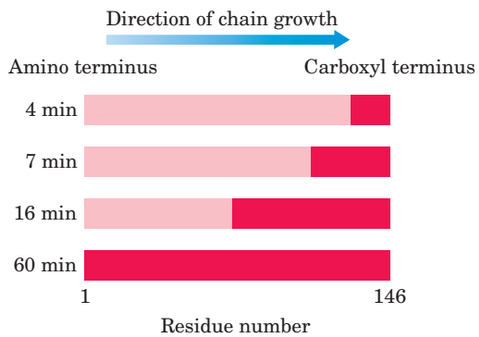


FIGURE 27–19 Proof that polypeptides grow by addition of amino acid residues to the carboxyl end: the Dintzis experiment. Reticulocytes (immature erythrocytes) actively synthesizing hemoglobin were incubated with radioactive leucine (selected because it occurs frequently in both the α- and β-globin chains). Samples of completed α chains were isolated from the reticulocytes at various times afterward, and the distribution of radioactivity was determined. The dark red zones show the portions of completed α-globin chains containing radioactive Leu residues. At 4 min, only a few residues at the carboxyl end of α-globin were labeled, because the only *complete* globin chains with incorporated label after 4 min were those that had nearly completed synthesis at the time the label was added. With longer incubation times, successively longer segments of the polypeptide 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, which means that polypeptides grow by successive addition of amino acids to the carboxyl end.

tRNA that is distinct from the $tRNA^{Met}$ used at (5′)AUG codons at interior positions in the mRNA. Polypeptides synthesized by mitochondrial and chloroplast ribosomes, however, begin with *N*-formylmethionine. This strongly supports the view that mitochondria and chloroplasts originated from bacterial ancestors that were symbiotically incorporated into precursor eukaryotic cells at an early stage of evolution (see Fig. 1–36).

How can the single (5′)AUG codon distinguish between the starting *N*-formylmethionine (or methionine, in eukaryotes) and interior Met residues? The details of the initiation process provide the answer.

The Three Steps of Initiation The **initiation** of polypeptide synthesis in bacteria requires (1) the 30S ribosomal subunit, (2) the mRNA coding for the polypeptide to be made, (3) the initiating fMet-$tRNA^{fMet}$, (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+}. Formation of the initiation complex takes place in three steps (Fig. 27–20).

In step ① the 30S ribosomal subunit binds two initiation factors, IF-1 and IF-3. Factor IF-3 prevents the 30S and 50S subunits from combining prematurely. The mRNA then binds to the 30S subunit. The initiating (5′)AUG is guided to its correct position by the **Shine-Dalgarno sequence** (named for Australian researchers John Shine and Lynn Dalgarno, who identified it) in the mRNA. This consensus sequence is an initiation signal of four to nine purine residues, 8 to 13 bp to the 5′ side of the initiation codon (Fig. 27–21a). The sequence base-pairs with a complementary pyrimidine-rich sequence near the 3′ end of the 16S rRNA of the 30S ribosomal subunit (Fig. 27–21b). This mRNA-rRNA interaction positions the initiating (5′)AUG sequence of the mRNA in the precise position on the 30S subunit where it is required for initiation of translation. The particular (5′)AUG where fMet-$tRNA^{fMet}$ is to be bound is distinguished from other methionine codons by its proximity to the Shine-Dalgarno sequence in the mRNA.

Bacterial ribosomes have three sites that bind aminoacyl-tRNAs, the **aminoacyl (A) site,** the **peptidyl (P) site,** and the **exit (E) site.** Both the 30S and the 50S subunits contribute to the characteristics of the A and P sites, whereas the E site is largely confined to the 50S subunit. The initiating (5′)AUG is positioned at the P site, the only site to which fMet-$tRNA^{fMet}$ can bind (Fig. 27–20). The fMet-$tRNA^{fMet}$ is the only aminoacyl-tRNA that binds first to the P site; during the subsequent elongation stage, all other incoming aminoacyl-tRNAs (including the Met-$tRNA^{Met}$ that binds to interior AUG codons) bind first to the A site and only subsequently to the P and E sites. The E site is the site from which the "uncharged" tRNAs leave during elongation. Factor IF-1 binds at the A site and prevents tRNA binding at this site during initiation.

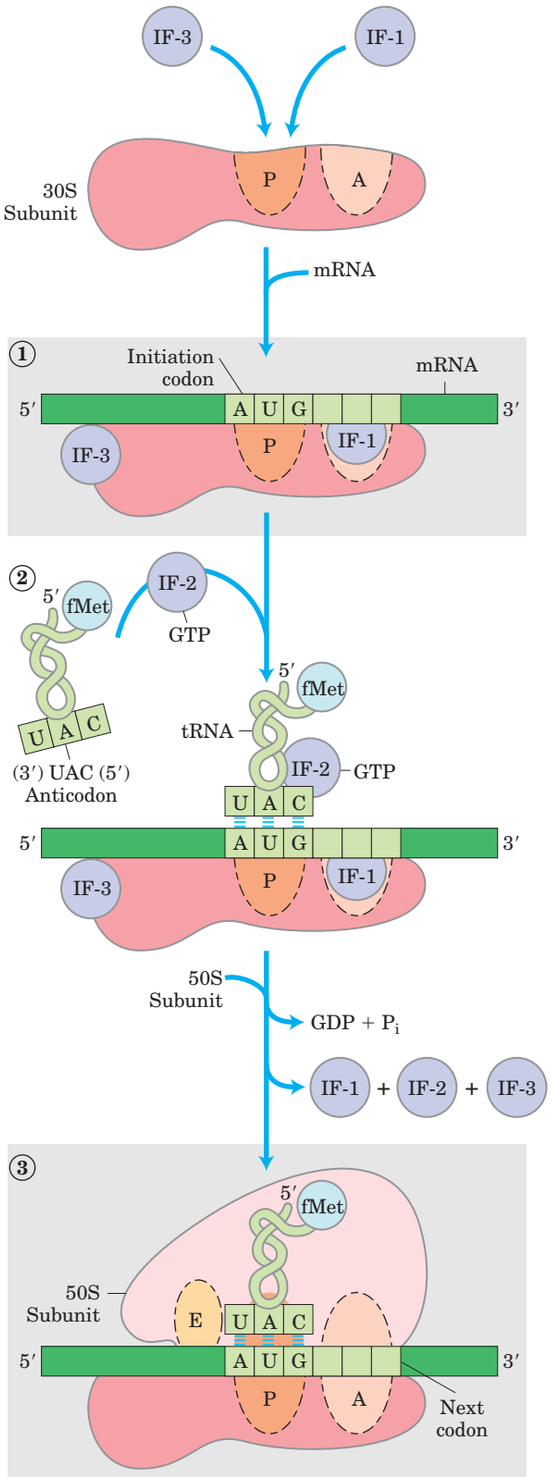
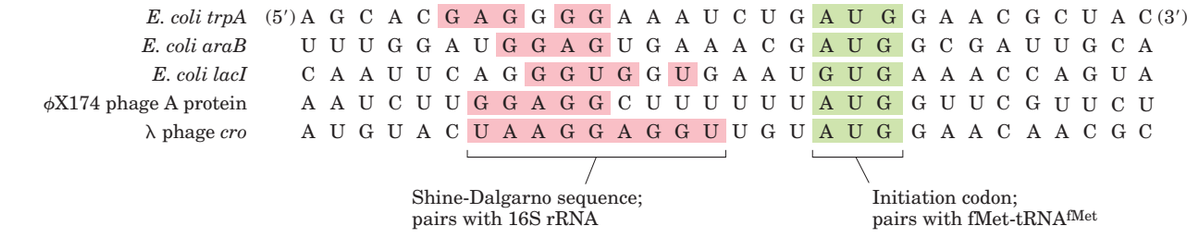
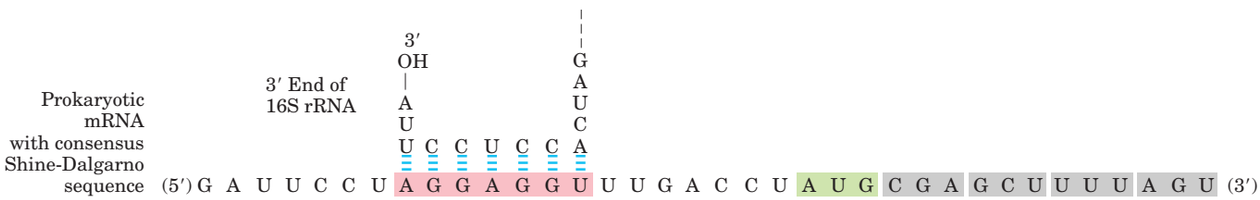


FIGURE 27–20 Formation of the initiation complex in bacteria. The complex forms in three steps (described in the text) at the expense of the hydrolysis of GTP to GDP and P_i. IF-1, IF-2, and IF-3 are initiation factors. P designates the peptidyl site, A the aminoacyl site, and E the exit site. Here the anticodon of the tRNA is oriented 3′ to 5′, left to right, as in Figure 27–8 but opposite to the orientation in Figures 27–16 and 27–18.



(a)



(b)

FIGURE 27–21 Messenger RNA sequences that serve as signals for initiation of protein synthesis in bacteria. (a) Alignment of the initiating AUG (shaded in green) at its correct location on the 30S ribosomal subunit depends in part on upstream Shine-Dalgarno sequences (pink). Portions of the mRNA transcripts of five prokaryotic genes are shown. Note the unusual example of the *E. coli* LacI protein, which initiates with a GUG (Val) codon (see Box 27–2). **(b)** The Shine-Dalgarno sequence of the mRNA pairs with a sequence near the 3′ end of the 16S rRNA.

In step ② of the initiation process (Fig. 27–20), the complex consisting of the 30S ribosomal subunit, IF-3, and mRNA is joined by both GTP-bound IF-2 and the initiating fMet-tRNA^{fMet}. The anticodon of this tRNA now pairs correctly with the mRNA's initiation codon.

In step ③ this large complex combines with the 50S ribosomal subunit; simultaneously, the GTP bound to IF-2 is hydrolyzed to GDP and P_i, which are released from the complex. All three initiation factors depart from the ribosome at this point.

Completion of the steps in Figure 27–20 produces 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 at least three points of recognition and attachment: the codon-anticodon interaction involving the initiation AUG fixed in the P site; interaction between the Shine-Dalgarno sequence in the mRNA and the 16S rRNA; and binding interactions between the ribosomal P site and the fMet-tRNA^{fMet}. The initiation complex is now ready for elongation.

Initiation in Eukaryotic Cells Translation is generally similar in eukaryotic and bacterial cells; most of the significant differences are in the mechanism of initiation. Eukaryotic mRNAs are bound to the ribosome as a complex with a number of specific binding proteins. Several of these tie together the 5′ and 3′ ends of the message. At the 3′ end, the mRNA is bound by the poly(A) binding protein (PAB). Eukaryotic cells have at least nine initiation factors. A complex called eIF4F, which includes the proteins eIF4E, eIF4G, and eIF4A, binds to the 5′ cap (see Fig. 26–12) through eIF4E. The protein eIF4G binds to both eIF4E and PAB, effectively tying them together (Fig. 27–22). The protein eIF4A has an RNA helicase activity. It is the eIF4F complex that associates

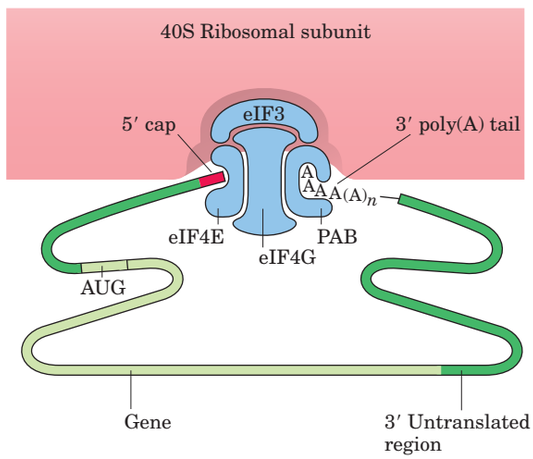


FIGURE 27–22 Protein complexes in the formation of a eukaryotic initiation complex. The 3′ and 5′ ends of eukaryotic mRNAs are linked by a complex of proteins that includes several initiation factors and the poly(A) binding protein (PAB). The factors eIF4E and eIF4G are part of a larger complex called eIF4F. This complex binds to the 40S ribosomal subunit.

TABLE 27–8 Protein Factors Required for Initiation of Translation in Bacterial and Eukaryotic Cells

Factor	*Function*
Bacterial	
IF-1	Prevents premature binding of tRNAs to A site
IF-2	Facilitates binding of fMet-tRNA^{fMet} to 30S ribosomal subunit
IF-3	Binds to 30S subunit; prevents premature association of 50S subunit; enhances specificity of P site for fMet-tRNA^{fMet}
Eukaryotic*	
eIF2	Facilitates binding of initiating Met-tRNA^{Met} to 40S ribosomal subunit
eIF2B, eIF3	First factors to bind 40S subunit; facilitate subsequent steps
eIF4A	RNA helicase activity removes secondary structure in the mRNA to permit binding to 40S subunit; part of the eIF4F complex
eIF4B	Binds to mRNA; facilitates scanning of mRNA to locate the first AUG
eIF4E	Binds to the 5′ cap of mRNA; part of the eIF4F complex
eIF4G	Binds to eIF4E and to poly(A) binding protein (PAB); part of the eIF4F complex
eIF5	Promotes dissociation of several other initiation factors from 40S subunit as a prelude to association of 60S subunit to form 80S initiation complex
eIF6	Facilitates dissociation of inactive 80S ribosome into 40S and 60S subunits

*The prefix "e" identifies these as eukaryotic factors.

with another factor, eIF3, and with the 40S ribosomal subunit. The efficiency of translation is affected by many properties of the mRNA and proteins in this complex, including the length of the 3′ poly(A) tract (in most cases, longer is better). The end-to-end arrangement of the eukaryotic mRNA facilitates translational regulation of gene expression, considered in Chapter 28.

The initiating (5′)AUG is detected within the mRNA not by its proximity to a Shine-Dalgarno-like sequence but by a scanning process: a scan of the mRNA from the 5′ end until the first AUG is encountered, signaling the beginning of the reading frame. The eIF4F complex is probably involved in this process, perhaps using the RNA helicase activity of eIF4A to eliminate secondary structure in the 5′ untranslated portion of the mRNA. Scanning is also facilitated by another protein, eIF4B.

The roles of the various bacterial and eukaryotic initiation factors in the overall process are summarized in Table 27–8. The mechanism by which these proteins act is an important area of investigation.

Stage 3: Peptide Bonds Are Formed in the Elongation Stage

The third stage of protein synthesis is **elongation.** Again, our initial focus is on bacterial cells. Elongation requires (1) the initiation complex described above, (2) aminoacyl-tRNAs, (3) a set of three soluble cytosolic proteins called **elongation factors** (EF-Tu, EF-Ts, and EF-G in bacteria), and (4) GTP. Cells use three steps to add each amino acid residue, and the steps are repeated as many times as there are residues to be added.

Elongation Step 1: Binding of an Incoming Aminoacyl-tRNA In the first step of the elongation cycle (Fig. 27–23), the appropriate incoming aminoacyl-tRNA binds to a complex of GTP-bound EF-Tu. The resulting aminoacyl-tRNA–EF-Tu–GTP complex binds to the A site of the 70S initiation complex. The GTP is hydrolyzed and an EF-Tu–GDP complex is released from the 70S ribosome. The EF-Tu–GTP complex is regenerated in a process involving EF-Ts and GTP.

Elongation Step 2: Peptide Bond Formation A peptide bond is now formed between the two amino acids bound by their tRNAs to the A and P sites on the ribosome. 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 (Fig. 27–24). The α-amino group of the amino acid in the A site acts as a 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 tRNAs then shift to a hybrid binding state, with elements of each spanning two different sites on the ribosome, as shown in Figure 27–24.

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 ribosomal subunit. We now know that this reaction is catalyzed by the 23S rRNA (Fig. 27–9), adding to the known catalytic repertoire of ribozymes. This discovery has interesting implications for the evolution of life (Box 27–3).

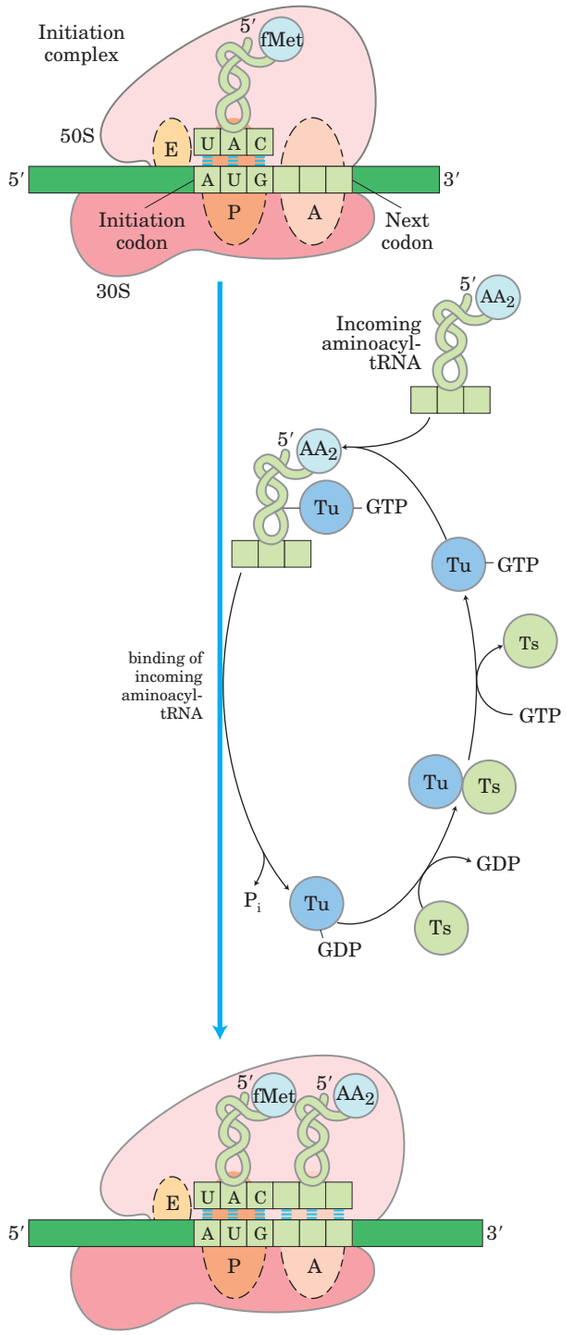


FIGURE 27–23 First elongation step in bacteria: binding of the second aminoacyl-tRNA. The second aminoacyl-tRNA enters the A site of the ribosome bound to EF-Tu (shown here as Tu), which also contains GTP. Binding of the second aminoacyl-tRNA to the A site is accompanied by hydrolysis of the GTP to GDP and P_i and release of the EF-Tu–GDP complex from 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 binds to EF-Tu. This recycles EF-Tu and makes it available to repeat the cycle.

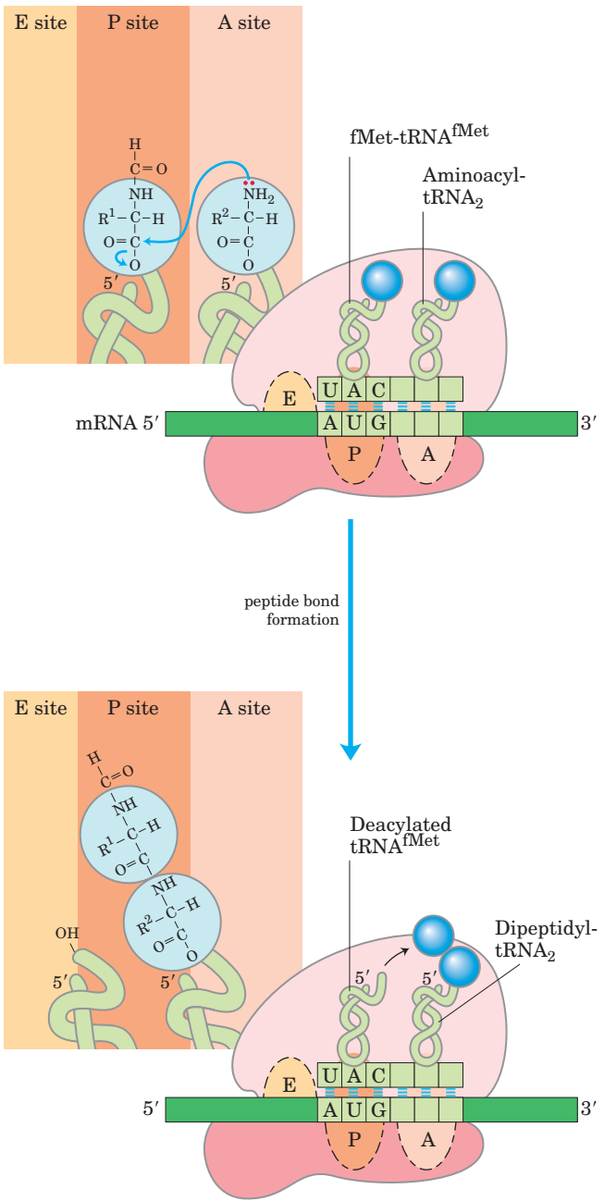
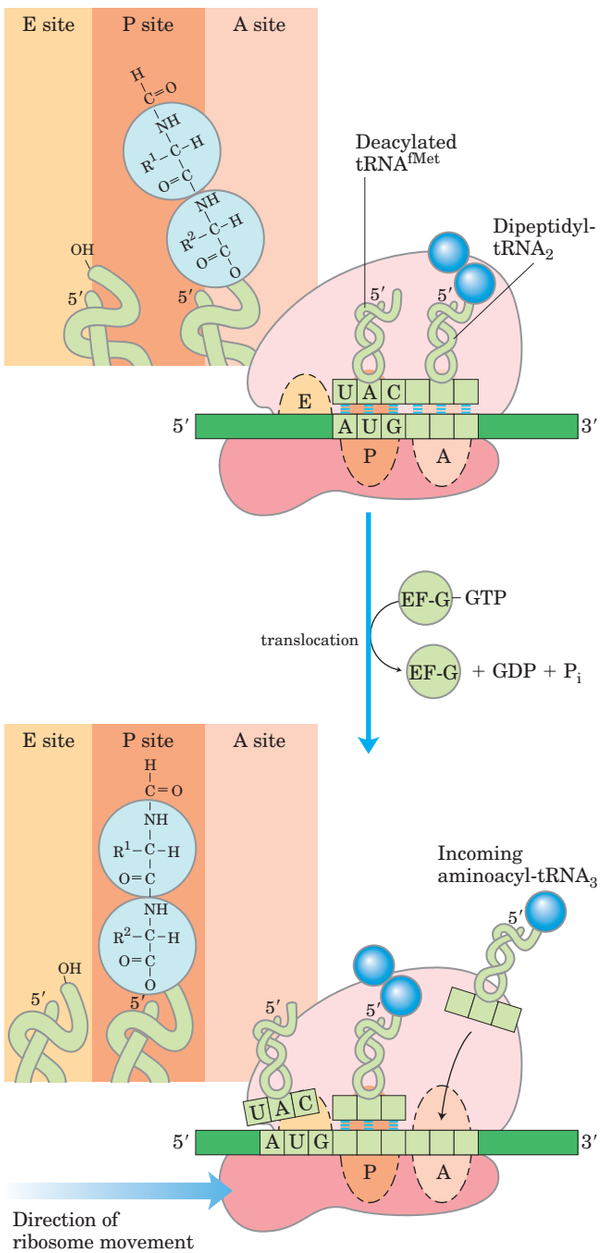


FIGURE 27–24 Second elongation step in bacteria: formation of the first peptide bond. The peptidyl transferase catalyzing this reaction is the 23S rRNA ribozyme. The *N*-formylmethionyl group is transferred to the amino group of the second aminoacyl-tRNA in the A site, forming a dipeptidyl-tRNA. At this stage, both tRNAs bound to the ribosome shift position in the 50S subunit to take up a hybrid binding state. The uncharged tRNA shifts so that its 3′ and 5′ ends are in the E site. Similarly, the 3′ and 5′ ends of the peptidyl tRNA shift to the P site. The anticodons remain in the A and P sites.

Elongation Step 3: Translocation In the final step of the elongation cycle, **translocation,** the ribosome moves one codon toward the 3′ end of the mRNA (Fig. 27–25a). This movement shifts the anticodon of the dipeptidyl-tRNA, which is still attached to the second codon of the mRNA, from the A site to the P site, and shifts the deacylated tRNA from the P site to the E site, from where the tRNA is released into the cytosol. The third codon of the mRNA now lies in the A site and the second codon in the P site. Movement of the ribosome along the mRNA requires EF-G (also known as translocase) and the energy provided by hydrolysis of another molecule of GTP.



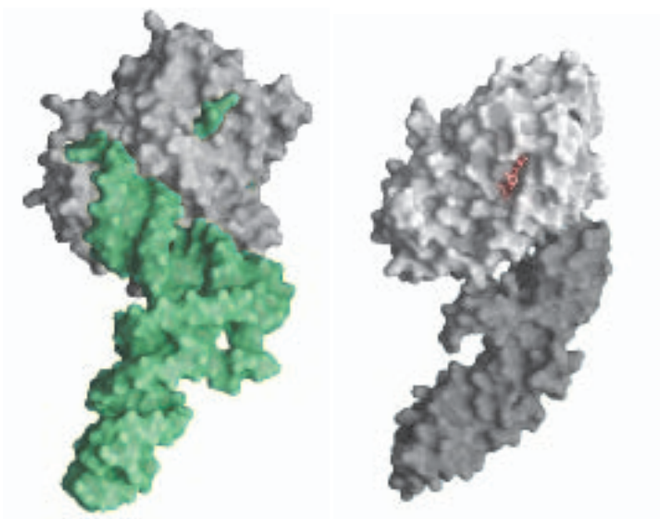
(a)

A change in the three-dimensional conformation of the entire ribosome results in its movement along the mRNA. Because the structure of EF-G mimics the structure of the EF-Tu–tRNA complex (Fig. 27–25b), EF-G can bind the A site and presumably displace the peptidyl-tRNA.

The ribosome, with its attached dipeptidyl-tRNA and mRNA, is now ready for the next elongation cycle and attachment of a third amino acid residue. This process occurs in the same way as addition of the second residue (as shown in Figs 27–23, 27–24, and 27–25). For each amino acid residue correctly added to the growing polypeptide, two GTPs are hydrolyzed to GDP and P_i as the ribosome moves from codon to codon along the mRNA toward the 3′ end.

The polypeptide remains attached to the tRNA of the most recent amino acid to be inserted. This association maintains the functional connection between the information in the mRNA and its decoded polypeptide output. At the same time, the ester linkage between this tRNA and the carboxyl terminus of the growing polypeptide activates the terminal carboxyl group for nucleophilic attack by the incoming amino acid to form a new peptide bond (Fig. 27–24). As the existing ester linkage between the polypeptide and tRNA is broken during

FIGURE 27–25 Third elongation step in bacteria: translocation. **(a)** The ribosome moves one codon toward the 3′ end of the mRNA, using energy provided by hydrolysis of GTP bound to EF-G (translocase). The dipeptidyl-tRNA is now entirely in the P site, leaving the A site open for the incoming (third) aminoacyl-tRNA. The uncharged tRNA dissociates from the E site, and the elongation cycle begins again. **(b)** The structure of EF-G mimics the structure of EF-Tu complexed with tRNA. Shown here are (left) EF-Tu complexed with tRNA (green) (PDB ID 1B23) and (right) EF-G complexed with GDP (red) (PDB ID 1DAR). The carboxyl-terminal part of EF-G (dark gray) mimics the structure of the anticodon loop of tRNA in both shape and charge distribution.



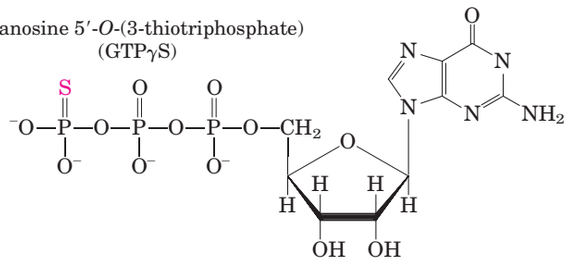
(b)

peptide bond formation, the linkage between the polypeptide and the information in the mRNA persists, because each newly added amino acid is still attached to its tRNA.

The elongation cycle in eukaryotes is quite similar to that in prokaryotes. Three eukaryotic elongation factors (eEF1α, eEF1$\beta\gamma$, and eEF2) have functions analogous to those of the bacterial elongation factors (EF-Tu, EF-Ts, and EF-G, respectively). Eukaryotic ribosomes do not have an E site; uncharged tRNAs are expelled directly from the P site.

Proofreading on the Ribosome The GTPase activity of EF-Tu during the first step of elongation in bacterial cells (Fig. 27–23) makes an important contribution to the rate and fidelity of the overall biosynthetic process. Both the EF-Tu–GTP and EF-Tu–GDP complexes exist for a few milliseconds before they dissociate. These two intervals provide opportunities for the codon-anticodon interactions to be proofread. Incorrect aminoacyl-tRNAs normally dissociate from the A site during one of these periods. If the GTP analog guanosine 5′-*O*-(3-thiotriphosphate) (GTPγS) is used in place of GTP, hydrolysis is slowed, improving the fidelity (by increasing the proofreading intervals) but reducing the rate of protein synthesis.

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



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. Note that the proofreading mechanism on the ribosome establishes only that the proper codon-anticodon pairing has taken place. The identity of the amino acid attached to a tRNA is not checked on the ribosome. If a tRNA is successfully aminoacylated with the wrong amino acid (as can be done experimentally), this incorrect amino acid is efficiently incorporated into a protein in response to whatever codon is normally recognized by the tRNA.

Stage 4: Termination of Polypeptide Synthesis Requires a Special Signal

Elongation continues until the ribosome adds the last amino acid coded by the mRNA. **Termination,** the fourth stage of polypeptide synthesis, is signaled by the presence of one of three termination codons in the mRNA (UAA, UAG, UGA), immediately following the final coded amino acid. Mutations in a tRNA anticodon that allow an amino acid to be inserted at a termination codon are generally deleterious to the cell (Box 27–4).

In bacteria, once a termination codon occupies the ribosomal A site, three **termination factors,** or **release factors**—the proteins RF-1, RF-2, and RF-3—contribute to (1) 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) dissociation of the 70S ribosome into its 30S and 50S subunits, ready to start a new cycle of polypeptide synthesis (Fig. 27–26). RF-1 recognizes the termination codons UAG and UAA, and RF-2 recognizes UGA and UAA. Either RF-1 or RF-2 (depending on which codon is present) binds at a termination codon and induces peptidyl transferase to transfer the growing polypeptide to a water molecule rather than to another amino acid. The release factors have domains thought to mimic the structure of tRNA, as shown for the elongation factor EF-G in Figure 27–25b. The specific function of RF-3 has not been firmly established, although it is thought to release the ribosomal subunit. In eukaryotes, a single release factor, eRF, recognizes all three termination codons.

Energy Cost of Fidelity in Protein Synthesis Synthesis of a protein true to the information specified in its mRNA requires energy. Formation of each aminoacyl-tRNA uses two high-energy phosphate groups. An additional ATP is consumed each time an incorrectly activated amino acid is hydrolyzed by the deacylation activity of an aminoacyl-tRNA synthetase, as part of its proofreading activity. A GTP is cleaved to GDP and P_i during the first elongation step, and another during the translocation step. Thus, on average, the energy derived from the hydrolysis of more than four NTPs to NDPs is required for the formation of each peptide bond of a polypeptide.

This represents an exceedingly large thermodynamic "push" in the direction of synthesis: at least 4×30.5 kJ/mol = 122 kJ/mol of phosphodiester bond energy to generate a peptide bond, which has a standard free energy of hydrolysis of only about -21 kJ/mol. The net free-energy change during peptide bond synthesis is thus -101 kJ/mol. Proteins are information-containing polymers. The biochemical goal is not simply the formation of a peptide bond but the formation of a peptide bond between two *specified* amino acids. Each of the high-energy phosphate compounds expended in this process plays a critical role in maintaining proper alignment between each new codon in the mRNA and its associated amino acid at the growing end of the polypeptide. This energy permits very high fidelity in the biological translation of the genetic message of mRNA into the amino acid sequence of proteins.

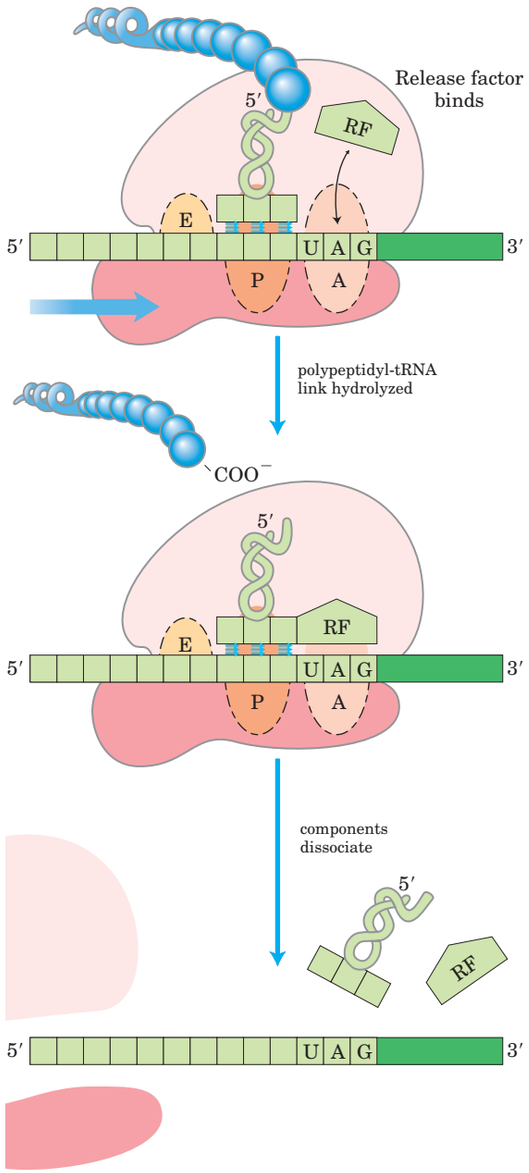


FIGURE 27–26 Termination of protein synthesis in bacteria. Termination occurs in response to a termination codon in the A site. First, a release factor, RF (RF-1 or RF-2, depending on which termination codon is present), binds to the A site. This leads to hydrolysis of the ester linkage between the nascent 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.

Rapid Translation of a Single Message by Polysomes Large clusters of 10 to 100 ribosomes that are very active in protein synthesis can be isolated from both eukaryotic and bacterial cells. Electron micrographs show a fiber between adjacent ribosomes in the cluster, which is called a **polysome** (Fig. 27–27). The connecting strand is a single molecule of mRNA that is being translated simultaneously by many closely spaced ribosomes, allowing the highly efficient use of the mRNA.

In bacteria, transcription and translation are tightly coupled. Messenger RNAs are synthesized and translated in the same 5′→3′ direction. Ribosomes begin translating the 5′ end of the mRNA before transcription is complete (Fig. 27–28). The situation is quite different in eukaryotic cells, where newly transcribed mRNAs must leave the nucleus before they can be translated.

Bacterial mRNAs generally exist for just a few minutes (p. 1020) before they are degraded by nucleases. 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 a rapid cessation of synthesis when the protein is no longer needed.

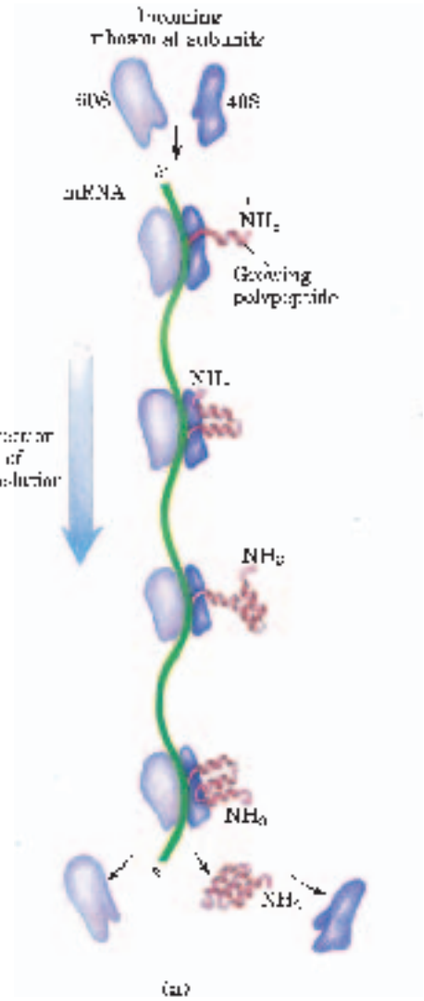
Stage 5: Newly Synthesized Polypeptide Chains Undergo Folding and Processing

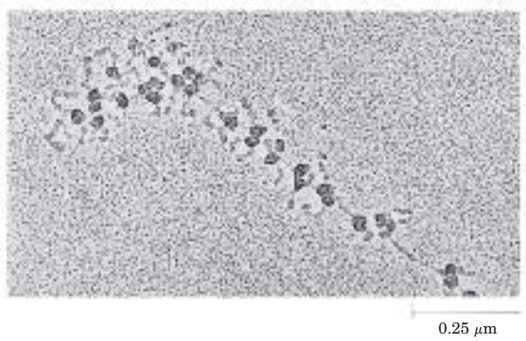
In the final stage of protein synthesis, the nascent polypeptide chain is folded and processed into its biologically active form. During or after its synthesis, the polypeptide progressively assumes its native conformation, with the formation of appropriate hydrogen bonds and van der Waals, ionic, and hydrophobic interactions. 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, both prokaryotic and eukaryotic, do not attain their final biologically active conformation until they have been altered by one or more processing reactions called **posttranslational modifications.**

Amino-Terminal and Carboxyl-Terminal Modifications The first residue inserted in all polypeptides is *N*-formylmethionine (in bacteria) or methionine (in eukaryotes). However, the formyl group, the amino-terminal Met residue, and often additional amino-terminal (and, in some cases, carboxyl-terminal) residues may be removed enzymatically in formation of the final functional protein. In as many as 50% of eukaryotic proteins, the amino group of the amino-terminal residue is *N*-acetylated after translation. Carboxyl-terminal residues are also sometimes modified.

Loss of Signal Sequences As we shall see in Section 27.3, 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.

Modification of Individual Amino Acids The hydroxyl groups of certain Ser, Thr, and Tyr residues of some proteins are enzymatically phosphorylated by ATP (Fig.





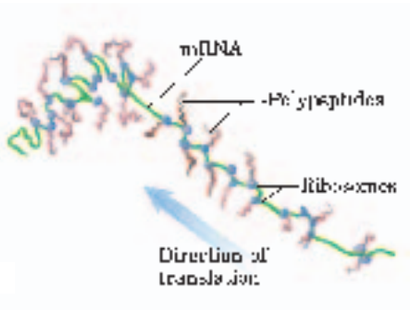


FIGURE 27–27 **Polysome. (a)** Four ribosomes translating a eukaryotic mRNA molecule simultaneously, moving from the 5′ end to the 3′ end and synthesizing a polypeptide from the amino terminus to the carboxyl terminus. **(b)** Electron micrograph and explanatory diagram of a polysome from the silk gland of a silkworm larva. The mRNA is being translated by many ribosomes simultaneously. The nascent polypeptides become longer as the ribosomes move toward the 3′ end of the mRNA. The final product of this process is silk fibroin.

27–29a); 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 that bind Ca^{2+}. Calcium, phosphate, and amino acids are all valuable to suckling young, so casein efficiently provides three essential nutrients. And as we have seen in numerous instances, phosphorylation-dephosphorylation cycles regulate the activity of many enzymes and regulatory proteins.

Extra carboxyl groups may be added to Glu residues of some proteins. For example, the blood-clotting protein prothrombin contains a number of γ-carboxyglutamate residues (Fig. 27–29b) in its amino-terminal region, introduced by an enzyme that requires vitamin K. These carboxyl groups bind Ca^{2+}, which is required to initiate the clotting mechanism.

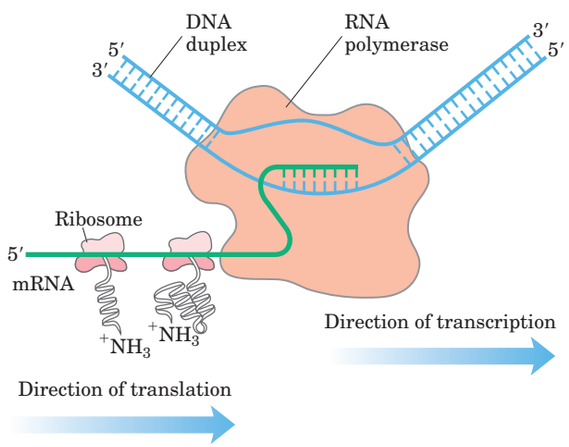


FIGURE 27–28 **Coupling of transcription and translation in bacteria.** The mRNA is translated by ribosomes while it is still being transcribed from DNA by RNA polymerase. This is possible because the mRNA in bacteria does not have to be transported from a nucleus to the cytoplasm before encountering ribosomes. In this schematic diagram the ribosomes are depicted as smaller than the RNA polymerase. In reality the ribosomes (M_r 2.7×10^6) are an order of magnitude larger than the RNA polymerase (M_r 3.9×10^5).

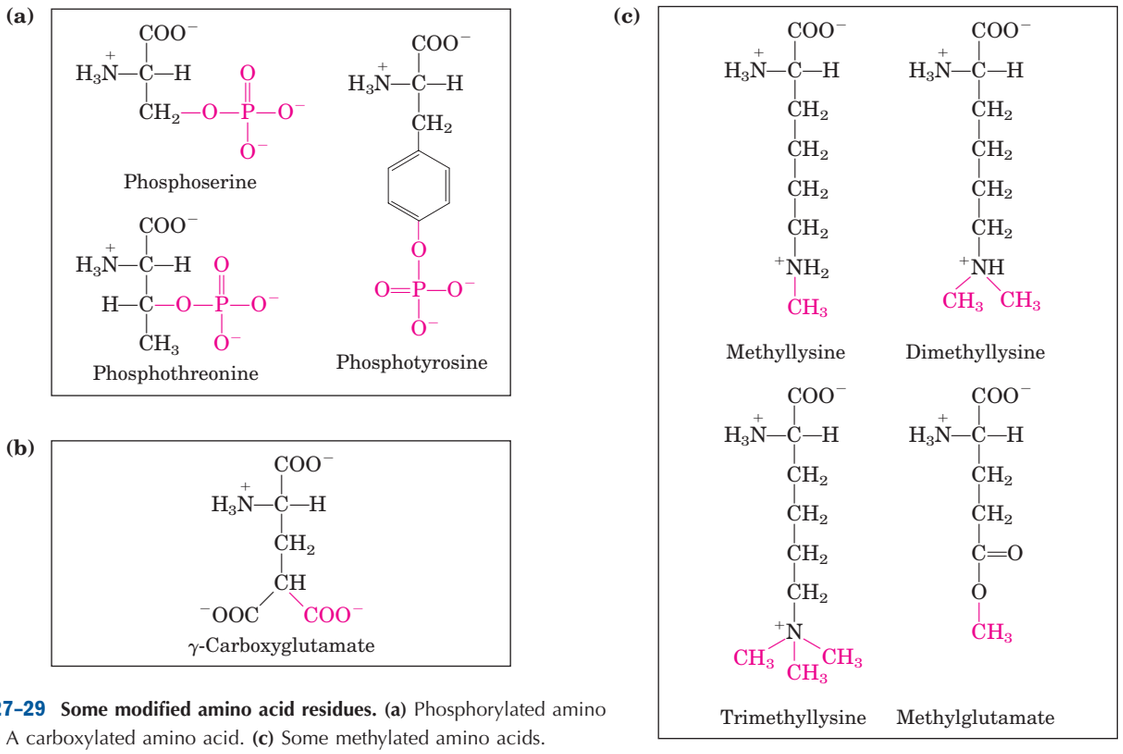


FIGURE 27–29 **Some modified amino acid residues.** **(a)** Phosphorylated amino acids. **(b)** A carboxylated amino acid. **(c)** Some methylated amino acids.

Monomethyl- and dimethyllysine residues (Fig. 27–29c) occur in some muscle proteins and in cytochrome *c*. The calmodulin of most species contains one trimethyllysine residue at a specific position. In other proteins, the carboxyl groups of some Glu residues undergo methylation, removing their negative charge.

Attachment of Carbohydrate Side Chains The carbohydrate side chains of glycoproteins are attached covalently during or after synthesis of the polypeptide. 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. 7–31). Many proteins that function extracellularly, as well as the lubricating proteoglycans that coat mucous membranes, contain oligosaccharide side chains (see Fig. 7–29).

Addition of Isoprenyl Groups A number of eukaryotic proteins are modified by the addition of groups derived from isoprene (isoprenyl groups). A thioether bond is formed between the isoprenyl group and a Cys residue of the protein (see Fig. 11–14). The isoprenyl groups are derived from pyrophosphorylated intermediates of the cholesterol biosynthetic pathway (see Fig. 21–33), such as farnesyl pyrophosphate (Fig. 27–30). Proteins modified in this way include the Ras proteins, products of the *ras* oncogenes and proto-oncogenes, and G proteins (both discussed in Chapter 12), and lamins, proteins found in the nuclear matrix. The isoprenyl group helps to anchor the protein in a membrane. The transforming (carcinogenic) activity of the *ras* oncogene is lost when isoprenylation of the Ras protein is blocked, a finding that has stimulated 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. Two examples are the biotin

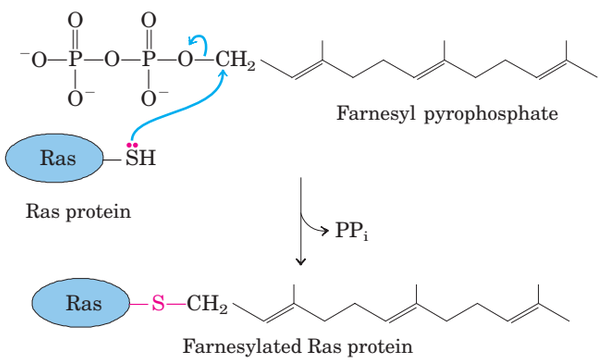


FIGURE 27–30 **Farnesylation of a Cys residue.** The thioether linkage is shown in red. The Ras protein is the product of the *ras* oncogene.

BOX 27–4 WORKING IN BIOCHEMISTRY

Induced Variation in the Genetic Code: Nonsense Suppression

When a mutation introduces a termination codon in the interior of a gene, translation is prematurely halted and the incomplete polypeptide is usually inactive. These are called nonsense mutations. The gene can be restored to normal function if a second mutation either (1) converts the misplaced termination codon to a codon specifying an amino acid or (2) suppresses the effects of the termination codon. Such restorative mutations are called **nonsense suppressors;** they generally involve mutations in tRNA genes to produce altered (suppressor) tRNAs that can recognize the termination codon and insert an amino acid at that position. Most known suppressor tRNAs have single base substitutions in their anticodons.

Suppressor tRNAs constitute an experimentally induced variation in the genetic code to allow the reading of what are usually termination codons, much like the naturally occurring code variations described in Box 27–2. Nonsense suppression does not completely disrupt normal information transfer in a cell, because the cell usually has several copies of each tRNA gene; 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 a "minor" tRNA, leaving the major tRNA to read its codon normally.

For example, *E. coli* has three identical genes for $tRNA^{Tyr}$, each producing a tRNA with the anticodon (5′)GUA. One of these genes is expressed at relatively high levels and thus its product represents the major $tRNA^{Tyr}$ species; the other two genes are 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 carried out inefficiently, but it can produce enough full-length protein from a gene with a nonsense mutation to allow the cell to survive. The major $tRNA^{Tyr}$ continues to translate the genetic code normally for the majority of proteins.

The mutation that leads to creation of a suppressor tRNA does not always occur in the anticodon. The suppression of UGA nonsense codons generally involves the $tRNA^{Trp}$ that normally recognizes UGG. The alteration that allows it to read UGA (and insert Trp residues at these positions) is a G to A change at position 24 (in an arm of the tRNA somewhat removed from the anticodon); this tRNA can now recognize *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 27–2).

Suppression should lead to many abnormally long proteins, but this does not always occur. We understand only a few details of the molecular events in translation termination and nonsense suppression.

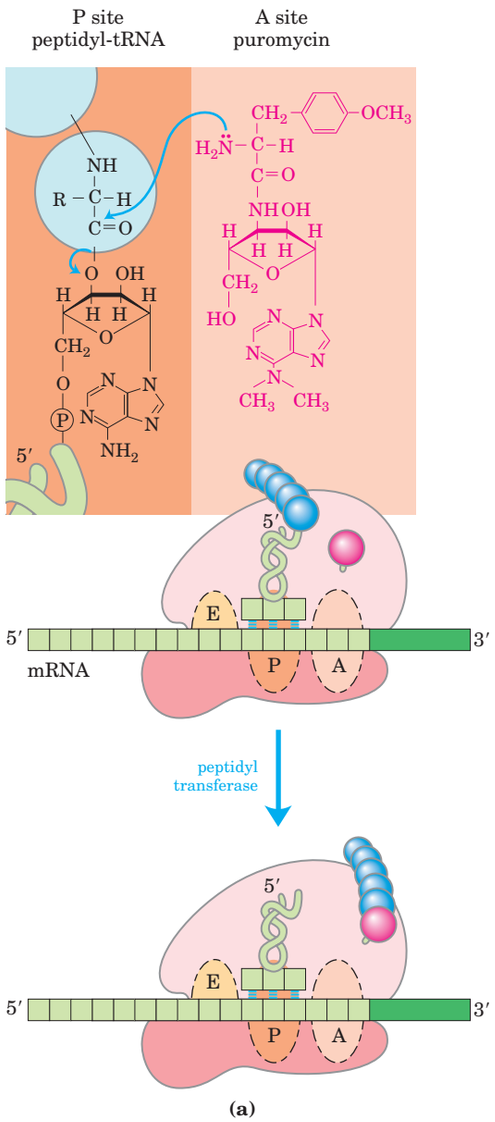
molecule of acetyl-CoA carboxylase and the heme group of hemoglobin or cytochrome *c*.

Proteolytic Processing Many proteins are initially synthesized as large, inactive precursor polypeptides that are proteolytically trimmed to form their smaller, active forms. Examples include proinsulin, some viral proteins, and proteases such as chymotrypsinogen and trypsinogen (see Fig. 6–33).

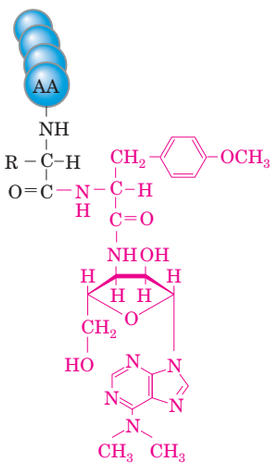
Formation of Disulfide Cross-Links After folding into their native conformations, some proteins form intrachain or interchain disulfide bridges between Cys residues. In eukaryotes, disulfide bonds are common in proteins to be exported from cells. The cross-links formed in this way help to protect the native conformation of the protein molecule from denaturation in the extracellular environment, which can differ greatly from intracellular conditions and is generally oxidizing.

Protein Synthesis Is Inhibited by Many Antibiotics and Toxins

Protein synthesis is a central function in cellular physiology and is the primary target of many naturally occurring antibiotics and toxins. Except as noted, these antibiotics inhibit protein synthesis in bacteria. The differences between bacterial and eukaryotic protein synthesis, though in some cases subtle, are sufficient that most of the compounds discussed below are relatively harmless to eukaryotic cells. Natural selection has favored the evolution of compounds that exploit minor differences in order to affect bacterial systems selectively, such that these biochemical weapons are synthesized by some microorganisms and are extremely toxic to others. Because nearly every step in protein synthesis can be specifically inhibited by one antibiotic or another, antibiotics have become valuable tools in the study of protein biosynthesis.



(a)

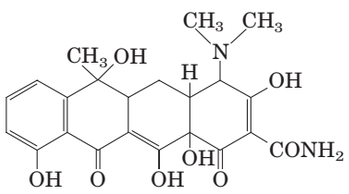


(b)

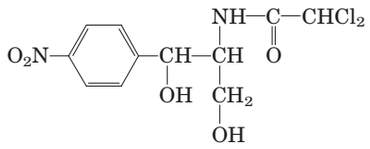
FIGURE 27–31 **Disruption of peptide bond formation by puromycin.** **(a)** The antibiotic puromycin resembles the aminoacyl end of a charged tRNA, and it can bind to the ribosomal A site and participate in peptide bond formation. The product of this reaction, instead of being translocated to the P site, dissociates from the ribosome, causing premature chain termination. **(b)** Peptidyl puromycin.

Puromycin, made by the mold *Streptomyces alboniger,* is one of the best-understood inhibitory antibiotics. Its structure is very similar to the 3′ end of an aminoacyl-tRNA, enabling it to bind to the ribosomal A site and participate in peptide bond formation, producing peptidyl-puromycin (Fig. 27–31). However, because puromycin resembles only the 3′ end of the tRNA, it does not engage in translocation and dissociates from the ribosome shortly after it is linked to the carboxyl terminus of the peptide. This prematurely terminates polypeptide synthesis.

Tetracyclines inhibit protein synthesis in bacteria by blocking the A site on the ribosome, preventing the binding of aminoacyl-tRNAs. **Chloramphenicol** inhibits protein synthesis by bacterial (and mitochondrial

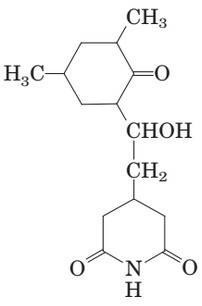


Tetracycline

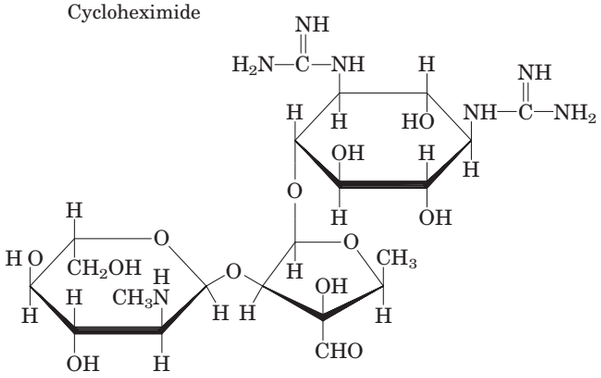


Chloramphenicol

and chloroplast) ribosomes by blocking peptidyl transfer; it 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.



Cycloheximide



Streptomycin

Several other inhibitors of protein synthesis are notable because of their toxicity to humans and other mammals. **Diphtheria toxin** (M_r 58,330) catalyzes the ADP-ribosylation of a diphthamide (a modified histidine) residue of eukaryotic elongation factor eEF2, thereby inactivating it. **Ricin** (M_r 29,895), an extremely toxic protein of the castor bean, inactivates the 60S subunit of eukaryotic ribosomes by depurinating a specific adenosine in 23S rRNA.

SUMMARY 27.2 Protein Synthesis

- Protein synthesis occurs on the ribosomes, which consist of protein and rRNA. Bacteria have 70S ribosomes, with a large (50S) and a small (30S) subunit. Eukaryotic ribosomes are significantly larger (80S) and contain more proteins.
- Transfer RNAs have 73 to 93 nucleotide residues, some of which have modified bases. Each tRNA has 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 D arm; some tRNAs have a fifth arm. The anticodon is responsible for the specificity of interaction between the aminoacyl-tRNA and the complementary mRNA codon.
- The growth of polypeptides on ribosomes begins with the amino-terminal amino acid and proceeds by successive additions of new residues to the carboxyl-terminal end.
- Protein synthesis occurs in five stages.

  1. 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 on the accuracy of this reaction, and some of these enzymes carry out proofreading steps at separate active sites. In bacteria, the initiating aminoacyl-tRNA in all proteins is *N*-formylmethionyl-$tRNA^{fMet}$.

  2. Initiation of protein synthesis involves formation of a complex between the 30S ribosomal subunit, mRNA, GTP, fMet-$tRNA^{fMet}$, three initiation factors, and the 50S subunit; GTP is hydrolyzed to GDP and P_i.

  3. In the elongation steps, GTP and elongation factors are required for binding the incoming aminoacyl-tRNA to the A 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 A site to the P site, a process requiring hydrolysis of GTP. Deacylated tRNAs dissociate from the ribosomal E site.

  4. After many such elongation cycles, synthesis of the polypeptide is terminated with the aid of release factors. At least four high-energy phosphate equivalents (from ATP and GTP) are required to generate each peptide bond, an energy investment required to guarantee fidelity of translation.

  5. Polypeptides fold into their active, three-dimensional forms. Many proteins are further processed by posttranslational modification reactions.
- Many well-studied antibiotics and toxins inhibit some aspect of protein synthesis.

27.3 Protein Targeting and Degradation

The eukaryotic cell is made up of many structures, compartments, and organelles, each with specific functions that require distinct sets of proteins and enzymes. These proteins (with the exception of those produced in mitochondria and plastids) are synthesized on ribosomes in the cytosol, so how are they directed to their final cellular destinations?

We are now beginning to understand this complex and fascinating process. Proteins destined for secretion, integration in the plasma membrane, or inclusion in lysosomes generally share the first few steps of a pathway that begins in the endoplasmic reticulum. Proteins destined for mitochondria, chloroplasts, or the nucleus use three separate mechanisms. And proteins destined for the cytosol simply remain where they are synthesized.

The most important element in many of these targeting pathways is a short sequence of amino acids called a **signal sequence,** whose function was first postulated by Günter Blobel and colleagues in 1970. The signal sequence directs a protein to its appropriate location in the cell and, for many proteins, is removed during transport or after the protein has reached its final destination. In proteins slated for transport into mitochondria, chloroplasts, or the ER, the signal sequence is at the amino terminus of a newly synthesized polypeptide. In many cases, the targeting capacity of particular signal sequences has been confirmed by fusing the signal sequence from one protein to a second protein and showing that the signal directs the second protein to the location where the first protein is normally found. The selective degradation of proteins no longer needed by the cell also relies largely on a set of molecular signals embedded in each protein's structure.

In this concluding section we examine protein targeting and degradation, emphasizing the underlying signals and molecular regulation that are so crucial to cellular metabolism. Except where noted, the focus is now on eukaryotic cells.

Günter Blobel

George Palade

Posttranslational Modification of Many Eukaryotic Proteins Begins in the Endoplasmic Reticulum

Perhaps the best-characterized targeting system begins in the ER. Most lysosomal, membrane, or secreted proteins have an amino-terminal signal sequence (Fig. 27–32) that marks them for translocation into the lumen of the ER; hundreds of such signal sequences have been determined. The carboxyl terminus of the signal sequence is defined by a cleavage site, where protease action removes the sequence after the protein is imported into the ER. Signal sequences vary in length from 13 to 36 amino acid residues, but all have the following features: (1) about 10 to 15 hydrophobic amino acid residues; (2) one or more positively charged 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, typically having amino acid residues with short side chains (especially Ala) at 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 helps to direct the ribosome to the ER, as illustrated by

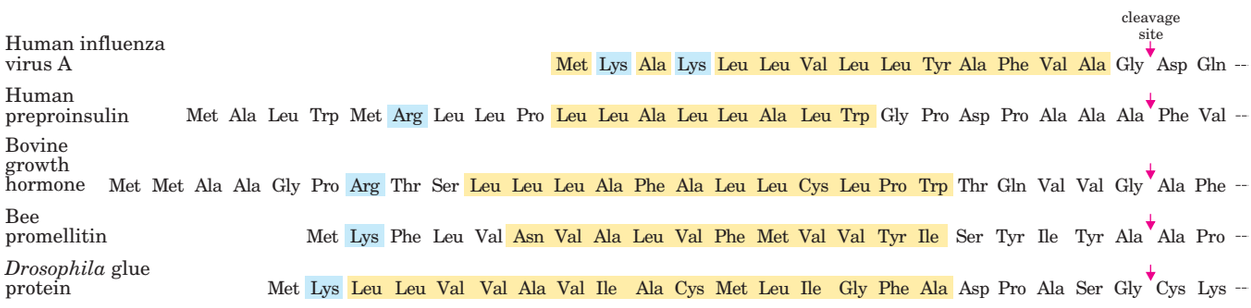


FIGURE 27–32 Translocation into the ER directed by amino-terminal signal sequences of some eukaryotic proteins. The hydrophobic core (yellow) is preceded by one or more basic residues (blue). Note the polar and short-side-chain residues immediately preceding (to the left of, as shown here) the cleavage sites (indicated by red arrows).

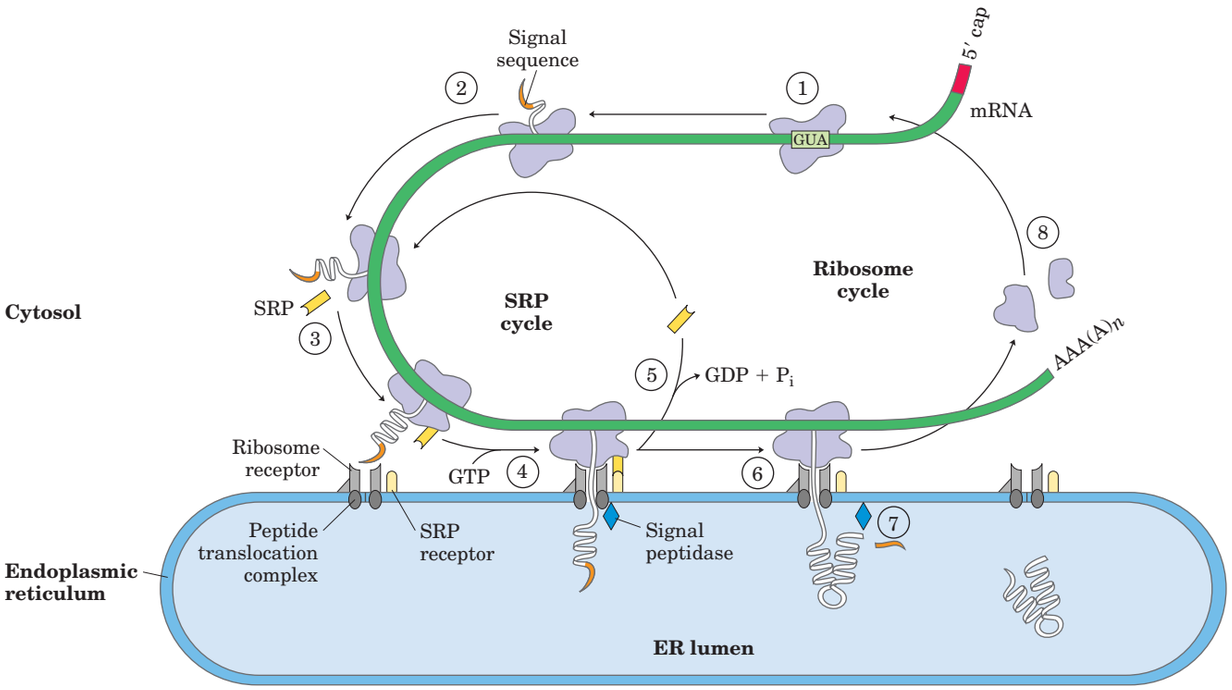
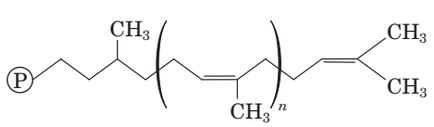


FIGURE 27–33 Directing eukaryotic proteins with the appropriate signals to the endoplasmic reticulum. This process involves the SRP cycle and translocation and cleavage of the nascent polypeptide. The steps are described in the text. SRP is a rod-shaped complex containing a 300 nucleotide RNA (7SL-RNA) and six different proteins (combined M_r 325,000). One protein subunit of SRP binds directly to the signal sequence, inhibiting elongation by sterically blocking the entry of aminoacyl-tRNAs and inhibiting peptidyl transferase. Another protein subunit binds and hydrolyzes GTP. The SRP receptor is a heterodimer of α (M_r 69,000) and β (M_r 30,000) subunits, both of which bind and hydrolyze multiple GTP molecules during this process.

steps ① through ⑧ in Figure 27–33. ① The targeting pathway begins with initiation of protein synthesis on free ribosomes. ② The signal sequence appears early in the synthetic process, because it is at the amino terminus, which as we have seen is synthesized first. ③ As it emerges from the ribosome, the signal sequence—and the ribosome itself—are bound by the large **signal recognition particle (SRP);** SRP then binds GTP and halts elongation of the polypeptide when it is about 70 amino acids long and the signal sequence has completely emerged from the ribosome. ④ The GTP-bound SRP now directs the ribosome (still bound to the mRNA) and the incomplete polypeptide to GTP-bound SRP receptors in the cytosolic face of the ER; the nascent polypeptide is delivered to a **peptide translocation complex** in the ER, which may interact directly with the ribosome. ⑤ SRP dissociates from the ribosome, accompanied by hydrolysis of GTP in both SRP and the SRP receptor. ⑥ Elongation of the polypeptide now resumes, with the ATP-driven translocation complex feeding the growing polypeptide into the ER lumen until the complete protein has been synthesized. ⑦ The signal sequence is removed by a signal peptidase within the ER lumen; ⑧ the ribosome dissociates and is recycled.

Glycosylation Plays a Key Role in Protein Targeting

In the ER lumen, newly synthesized proteins are further modified in several ways. Following the removal of signal sequences, polypeptides are folded, disulfide bonds formed, and many proteins glycosylated to form glycoproteins. In many glycoproteins the linkage to their oligosaccharides is through Asn residues. These *N*-linked oligosaccharides are diverse (Chapter 7), but the pathways by which they form have a common first step. A 14 residue core oligosaccharide is built up in a stepwise fashion, then transferred from a dolichol phosphate donor molecule to certain Asn residues in the protein (Fig. 27–34). The transferase is on the lumenal face of the ER and thus cannot catalyze glycosylation of cytosolic proteins. After transfer, the core oligosaccharide is trimmed and elaborated in different ways on different



Dolichol phosphate
(n = 9–22)

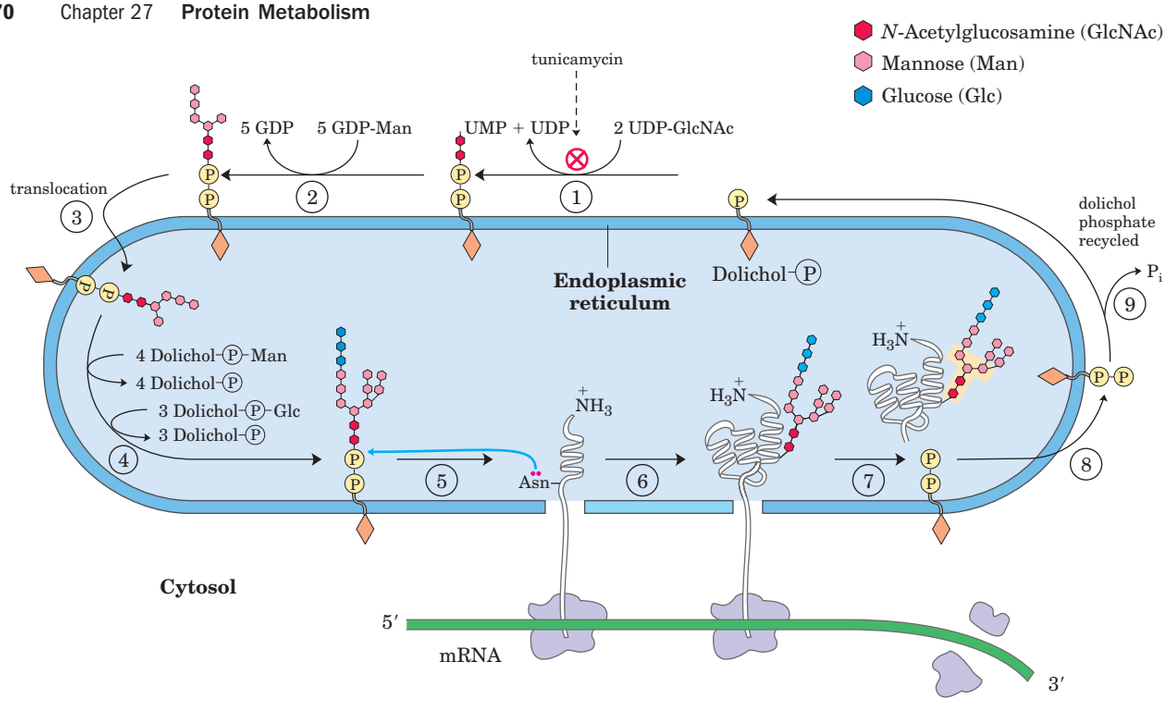
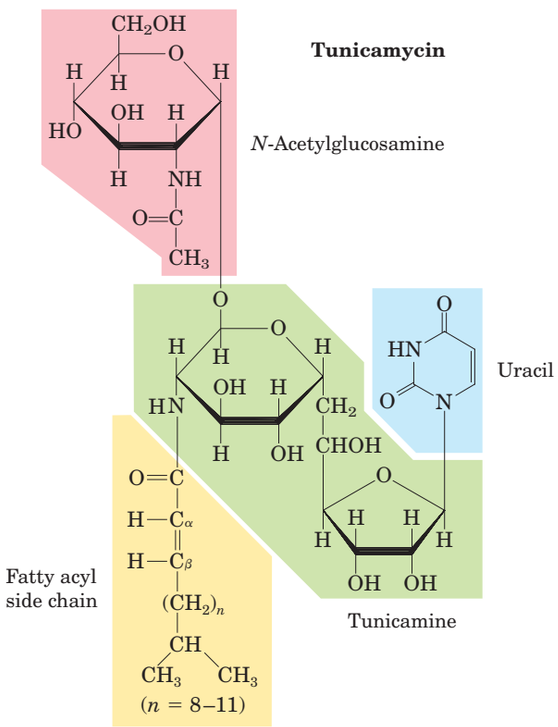


FIGURE 27–34 Synthesis of the core oligosaccharide of glycoproteins. The core oligosaccharide is built up by the successive addition of monosaccharide units. (1), (2) The first steps occur on the cytosolic face of the ER. (3) Translocation moves the incomplete oligosaccharide across the membrane (mechanism not shown), and (4) completion of the core oligosaccharide occurs within the lumen of the ER. The precursors that contribute additional mannose and glucose residues to the growing oligosaccharide in the lumen are dolichol phosphate derivatives. In the first step in the construction of the *N*-linked oligosaccharide moiety of a glycoprotein, (5), (6) the core oligosaccharide is transferred from dolichol phosphate to an Asn residue of the protein within the ER lumen. The core oligosaccharide is then further modified in the ER and the Golgi complex in pathways that differ for different proteins. The five sugar residues shown surrounded by a beige screen (after step (7)) are retained in the final structure of all *N*-linked oligosaccharides. (8) The released dolichol pyrophosphate is again translocated so that the pyrophosphate is on the cytosolic face of the ER, then (9) a phosphate is hydrolytically removed to regenerate dolichol phosphate.

proteins, but all *N*-linked oligosaccharides retain a pentasaccharide core derived from the original 14 residue oligosaccharide. Several antibiotics act by interfering with one or more steps in this process and have aided in elucidating the steps of protein glycosylation. The best-characterized is **tunicamycin,** which mimics the structure of UDP-*N*-acetylglucosamine and blocks the first step of the process (Fig. 27–34, step (1)). A few proteins are *O*-glycosylated in the ER, but most *O*-glycosylation occurs in the Golgi complex or in the cytosol (for proteins that do not enter the ER).

Suitably modified proteins can now be moved to a variety of intracellular destinations. Proteins travel from the ER to the Golgi complex in transport vesicles (Fig. 27–35). In the Golgi complex, oligosaccharides are *O*-linked to some proteins, and *N*-linked oligosaccharides are further modified. By mechanisms not yet fully understood, the Golgi complex also sorts proteins and sends them to their final destinations. The processes that segregate proteins targeted for secretion from those targeted for the plasma membrane or lysosomes must distinguish among these proteins on the basis of structural features other than signal sequences, which were removed in the ER lumen.



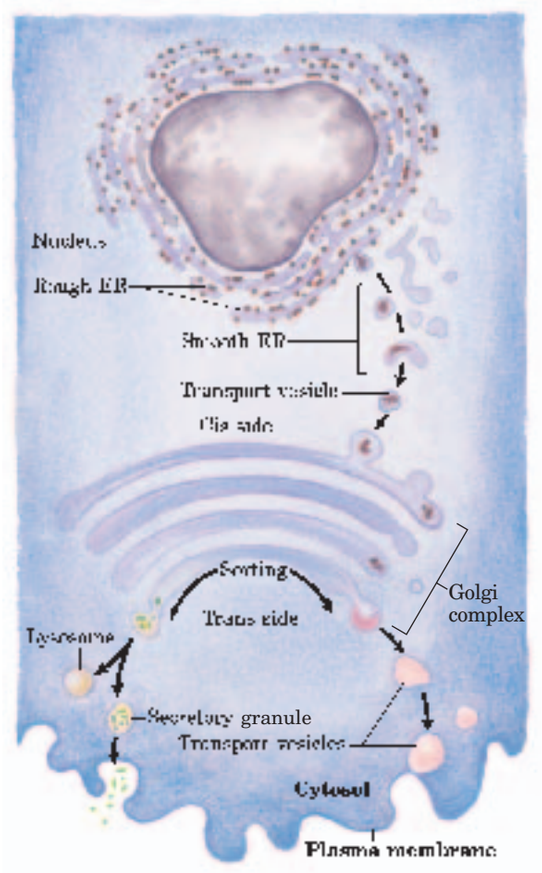


FIGURE 27–35 Pathway taken by proteins destined for lysosomes, the plasma membrane, or secretion. Proteins are moved from the ER to the cis side of the Golgi complex in transport vesicles. Sorting occurs primarily in the trans side of the Golgi complex.

This sorting process is best understood in the case of hydrolases destined for transport to lysosomes. On arrival of a hydrolase (a glycoprotein) in the Golgi complex, an as yet undetermined feature (sometimes called a signal patch) of the three-dimensional structure of the hydrolase is recognized by a phosphotransferase, which phosphorylates certain mannose residues in the oligosaccharide (Fig. 27–36). The presence of one or more mannose 6-phosphate residues in its *N*-linked oligosaccharide is the structural signal that targets the protein to lysosomes. A receptor protein in the membrane of the Golgi complex recognizes the mannose 6-phosphate signal and binds the hydrolase 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 complex dissociates in a process facilitated by the lower pH in the vesicle and by phosphatase-catalyzed removal of phosphate groups from the mannose 6-phosphate residues. The receptor is then recycled 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. 27–34, step ①), hydrolases that should be targeted for lysosomes are instead secreted, confirming that the *N*-linked oligosaccharide plays a key role in targeting these enzymes to lysosomes.

The pathways that target proteins to mitochondria and chloroplasts also rely on amino-terminal signal sequences. Although mitochondria and chloroplasts contain DNA, most of their proteins are encoded by nuclear DNA and must be targeted to the appropriate organelle. Unlike other targeting pathways, however, the mitochondrial and chloroplast pathways begin only *after* a precursor protein has been completely synthesized and released from the ribosome. Precursor proteins destined for mitochondria or chloroplasts are bound by cytosolic chaperone proteins and delivered to receptors on the exterior surface of the target organelle. Specialized translocation mechanisms then transport the protein to its final destination in the organelle, after which the signal sequence is removed.

Signal Sequences for Nuclear Transport Are Not Cleaved

Molecular communication between the nucleus and the cytosol requires the movement of macromolecules through nuclear pores. RNA molecules synthesized in the nucleus are exported to the cytosol. Ribosomal proteins synthesized on cytosolic ribosomes are imported into the nucleus and assembled into 60S and 40S ribosomal subunits in the nucleolus; completed subunits are then exported back to the cytosol. A variety of nuclear proteins (RNA and DNA polymerases, histones, topoisomerases, proteins that regulate gene expression, and so forth) are synthesized in the cytosol and imported into the nucleus. This traffic is modulated by a complex system of molecular signals and transport proteins that is gradually being elucidated.

In most multicellular eukaryotes, the nuclear envelope breaks down at each cell division, and once division is completed and the nuclear envelope reestablished, the dispersed nuclear proteins must be reimported. To allow this repeated nuclear importation, the signal sequence that targets a protein to the nucleus—the nuclear localization sequence, NLS—is not removed after the protein arrives at its destination. An NLS, unlike other signal sequences, may be located almost anywhere along the primary sequence of the protein. NLSs can vary considerably, but many consist of four to eight amino acid residues and include several consecutive basic (Arg or Lys) residues.

Nuclear importation is mediated by a number of proteins that cycle between the cytosol and the nucleus (Fig. 27–37), including importin α and β and a small GTPase known as Ran. A heterodimer of importin α and β functions as a soluble receptor for proteins targeted to the nucleus, with the α subunit binding NLS-bearing

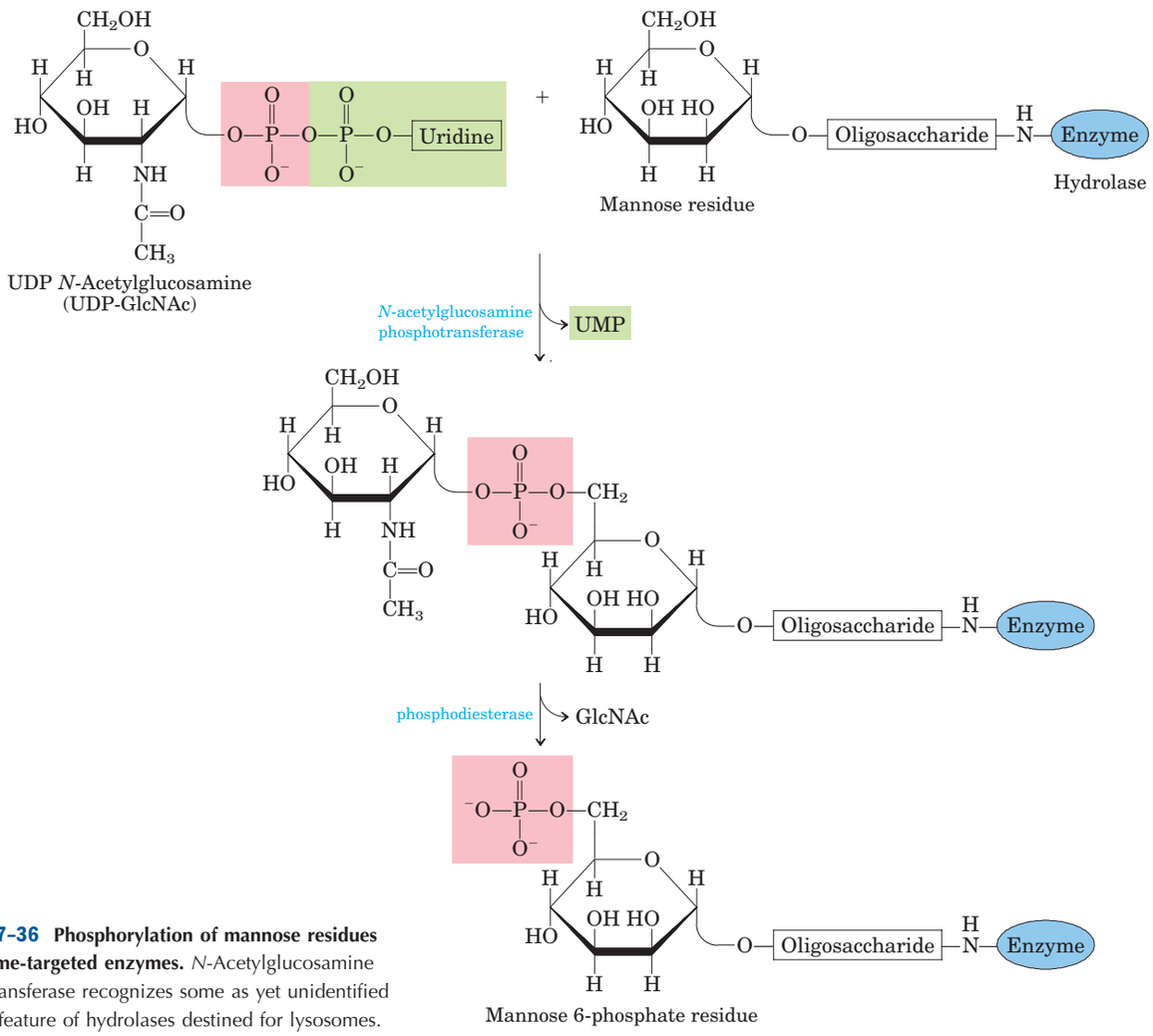


FIGURE 27–36 Phosphorylation of mannose residues on lysosome-targeted enzymes. *N*-Acetylglucosamine phosphotransferase recognizes some as yet unidentified structural feature of hydrolases destined for lysosomes.

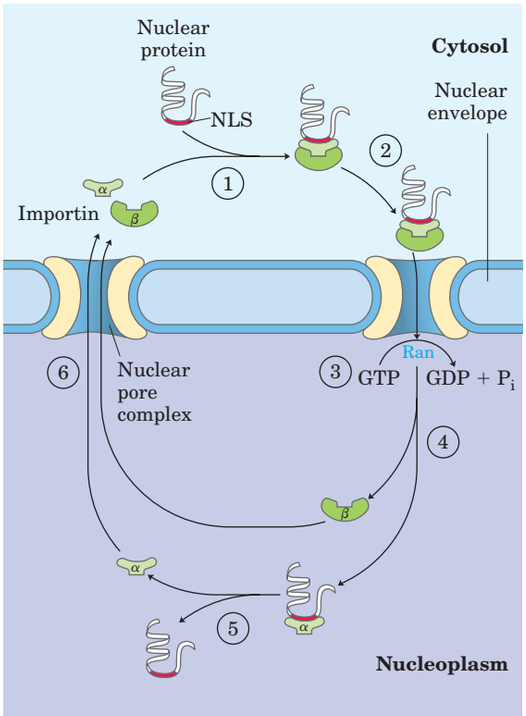
proteins in the cytosol. The complex of the NLS-bearing protein and the importin docks at a nuclear pore and is translocated through the pore by an energy-dependent mechanism that requires the Ran GTPase. The two importin subunits separate during the translocation, and the NLS-bearing protein dissociates from importin α inside the nucleus. Importin α and β are then exported from the nucleus to repeat the process. How importin α remains dissociated from the many NLS-bearing proteins inside the nucleus is not yet clear.

Bacteria Also Use Signal Sequences for Protein Targeting

Bacteria can target proteins to their inner or outer membranes, to the periplasmic space between these membranes, or to the extracellular medium. They use signal sequences at the amino terminus of the proteins (Fig. 27–38), much like those on eukaryotic proteins targeted to the ER, mitochondria, and chloroplasts.

Most proteins exported from *E. coli* make use of the pathway shown in Figure 27–39. Following translation, a protein to be exported may fold only slowly, the amino-terminal signal sequence impeding the folding. The soluble chaperone protein SecB binds to the protein's signal sequence or other features of its incompletely folded structure. The bound protein is then delivered to SecA, a protein associated with the inner surface of the plasma membrane. SecA acts as both a receptor and a translocating ATPase. Released from SecB and bound to SecA, the protein is delivered to a translocation complex in the membrane, made up of SecY, E, and G, and is translocated stepwise through the membrane at the SecYEG complex in lengths of about 20 amino acid residues. Each step is facilitated by the hydrolysis of ATP, catalyzed by SecA.

(a)



(b)

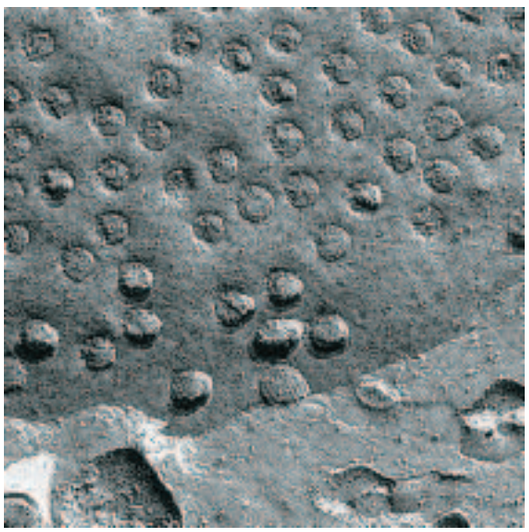


FIGURE 27–37 Targeting of nuclear proteins. (a) (1) A protein with an appropriate nuclear localization signal (NLS) is bound by a complex of importin α and β. (2) The resulting complex binds to a nuclear pore, and (3) translocation is mediated by the Ran GTPase. (4) Inside the nucleus, importin β dissociates from importin α, and (5) importin α then releases the nuclear protein. (6) Importin α and β are transported out of the nucleus and recycled. **(b)** Scanning electron micrograph of the surface of the nuclear envelope, showing numerous nuclear pores.

Inner membrane proteins																																
																														cleavage site		
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	--						

FIGURE 27–38 Signal sequences that target proteins 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 indicated by red arrows. Note that the inner bacterial cell membrane (see Fig. 1–6) is where phage fd coat proteins and DNA are assembled into phage particles. OmpA is outer membrane protein A; LamB is a cell surface receptor protein for bacteriophage lambda.

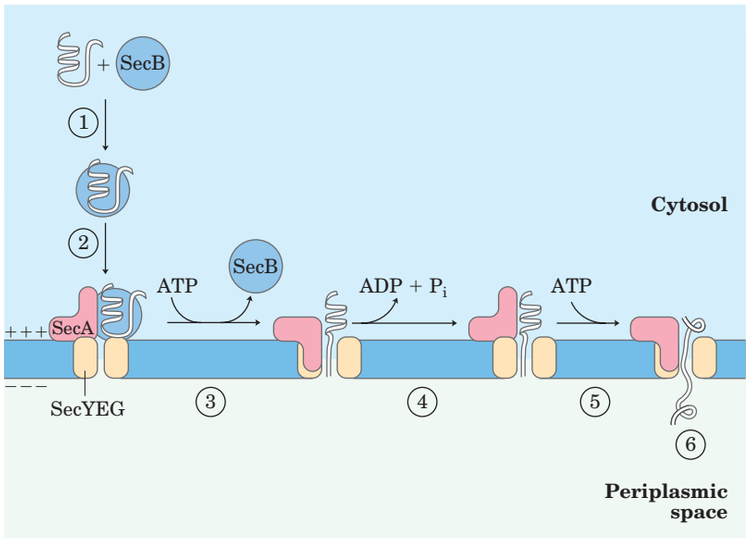


FIGURE 27–39 Model for protein export in bacteria. (1) A newly translated polypeptide binds to the cytosolic chaperone protein SecB, which (2) delivers it to SecA, a protein associated with the translocation complex (SecYEG) in the bacterial cell membrane. (3) SecB is released, and SecA inserts itself into the membrane, forcing about 20 amino acid residues of the protein to be exported through the translocation complex. (4) Hydrolysis of an ATP by SecA provides the energy for a conformational change that causes SecA to withdraw from the membrane, releasing the polypeptide. (5) SecA binds another ATP, and the next stretch of 20 amino acid residues is pushed across the membrane through the translocation complex. Steps (4) and (5) are repeated until (6) the entire protein has passed through and is released to the periplasm. The electrochemical potential across the membrane (denoted by + and −) also provides some of the driving force required for protein translocation.

An exported protein is thus pushed through the membrane by a SecA protein located on the cytoplasmic surface, rather than being pulled through the membrane by a protein on the periplasmic surface. This difference may simply reflect the need for the translocating ATPase to be where the ATP is. The transmembrane electrochemical potential can also provide energy for translocation of the protein, by an as yet unknown mechanism.

Although most exported bacterial proteins use this pathway, some follow an alternative pathway that uses signal recognition and receptor proteins homologous to components of the eukaryotic SRP and SRP receptor (Fig. 27–33).

Cells Import Proteins by Receptor-Mediated Endocytosis

Some proteins are imported into cells from the surrounding medium; examples in eukaryotes include low-density lipoprotein (LDL), the iron-carrying protein transferrin, peptide hormones, and circulating proteins destined for degradation. The proteins bind to receptors in invaginations of the membrane called **coated pits,** which concentrate endocytic receptors in preference to other cell-surface proteins. The pits are coated on their cytosolic side with a lattice of the protein **clathrin,** which forms closed polyhedral structures (Fig. 27–40). The clathrin lattice grows as more recep-

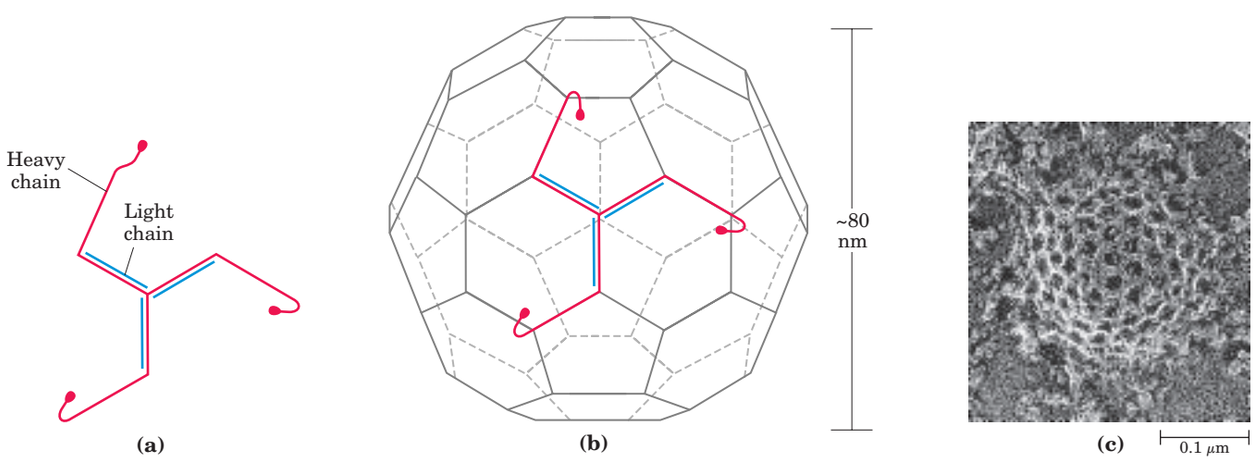


FIGURE 27–40 Clathrin. (a) Three light (L) chains (M_r 35,000) and three heavy (H) chains (M_r 180,000) of the $(HL)_3$ clathrin unit, organized as a three-legged structure called a triskelion. **(b)** Triskelions tend to assemble into polyhedral lattices. **(c)** Electron micrograph of a coated pit on the cytosolic face of the plasma membrane of a fibroblast.

tors are occupied by target proteins, until a complete membrane-bounded endocytic vesicle buds off the plasma membrane and enters the cytoplasm. The clathrin is quickly removed by uncoating enzymes, and the vesicle fuses with an endosome. ATPase activity in the endosomal membranes reduces the pH therein, facilitating dissociation of receptors from their target proteins.

The imported proteins and receptors then go their separate ways, their fates varying with the cell and protein type. Transferrin and its receptor are eventually recycled. Some hormones, growth factors, and immune complexes, after eliciting the appropriate cellular response, are degraded along with their receptors. LDL is degraded after the associated cholesterol has been delivered to its destination, but the LDL receptor is recycled (see Fig. 21–42).

Receptor-mediated endocytosis is exploited by some toxins and viruses to gain entry to cells. Influenza virus (see Fig. 11–24), diphtheria toxin, and cholera toxin all enter cells in this way.

Protein Degradation Is Mediated by Specialized Systems in All Cells

Protein degradation prevents the buildup of abnormal or unwanted proteins and permits the recycling of amino acids. The half-lives of eukaryotic proteins vary from 30 seconds to many days. Most proteins turn over rapidly relative to the lifetime of a cell, although a few (such as hemoglobin) can last for the life of the cell (about 110 days for an erythrocyte). Rapidly degraded proteins include those that are defective because of incorrectly inserted amino acids or because of damage accumulated during normal functioning. And enzymes that act at key regulatory points in metabolic pathways often turn over rapidly.

Defective proteins and those with characteristically short half-lives are generally degraded in both bacterial and eukaryotic cells by selective ATP-dependent cytosolic systems. A second system in vertebrates, operating in lysosomes, recycles the amino acids of membrane proteins, extracellular proteins, and proteins with characteristically long half-lives.

In *E. coli,* many proteins are degraded by an ATP-dependent protease called Lon (the name refers to the "long form" of proteins, observed only when this protease is absent). The protease is activated in the presence of defective proteins or those slated for rapid turnover; two ATP molecules are hydrolyzed for every peptide bond cleaved. The precise role of this ATP hydrolysis is not yet clear. Once a protein has been reduced to small inactive peptides, other ATP-independent proteases complete the degradation process.

The ATP-dependent pathway in eukaryotic cells is quite different, involving the protein **ubiquitin,** which, as its name suggests, occurs throughout the eukaryotic kingdoms. One of the most highly conserved proteins known, ubiquitin (76 amino acid residues) 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 (E1, E2, and E3 in Fig. 27–41).

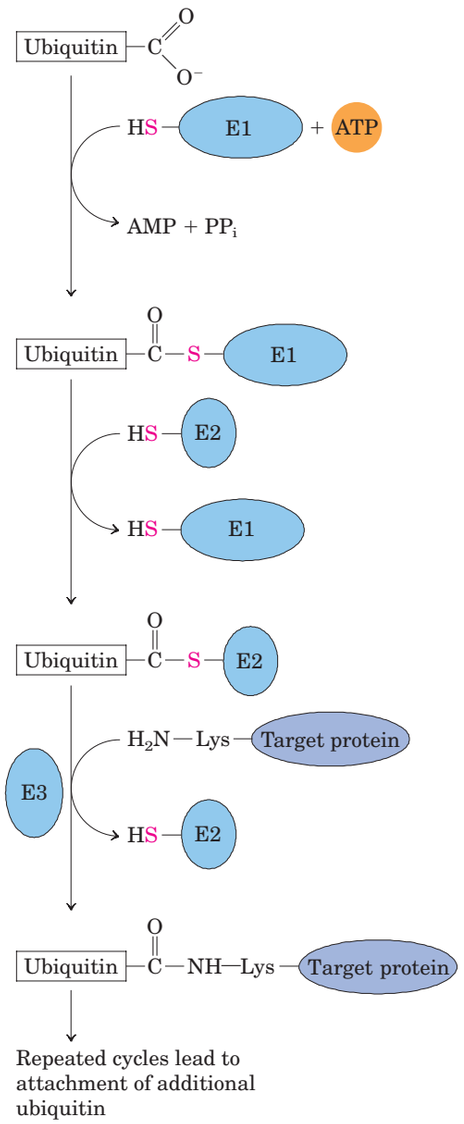
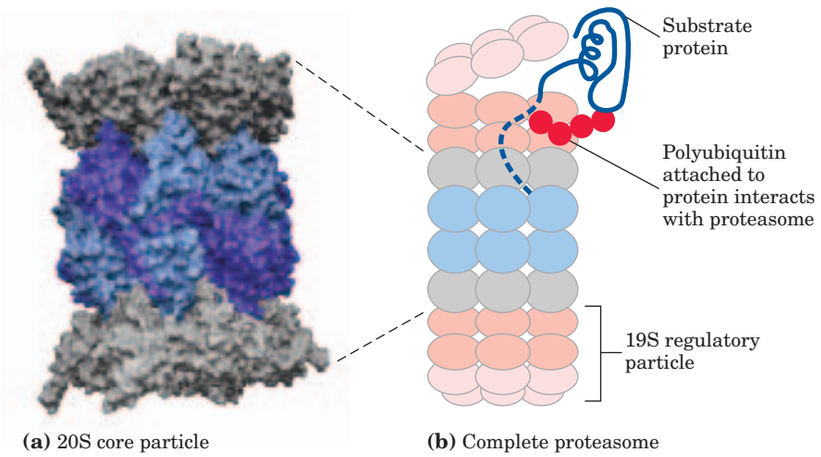


FIGURE 27–41 Three-step cascade pathway by which ubiquitin is attached to a protein. Two different enzyme-ubiquitin intermediates are involved. The free carboxyl group of ubiquitin's carboxyl-terminal Gly residue is ultimately linked through an amide (isopeptide) bond to an ϵ-amino group of a Lys residue of the target protein. Additional cycles produce polyubiquitin, a covalent polymer of ubiquitin subunits that targets the attached protein for destruction in eukaryotes.



(a) 20S core particle **(b)** Complete proteasome

FIGURE 27–42 Three-dimensional structure of the eukaryotic proteasome. The 26S proteasome is highly conserved in all eukaryotes. The two subassemblies are the 20S core particle and the 19S regulatory particle. **(a)** (PDB ID 1IRU) The core particle consists of four rings arranged to form a barrel-like structure. Each of the inner rings has seven different β subunits (light blue), three of which have protease activities (dark blue). The outer rings each have seven different α subunits (gray). **(b)** A regulatory particle forms a cap on each end of the core particle. The core particle is colored as in **(a)**. The base and lid segments of each regulatory particle are presented in different shades of red. The regulatory particle unfolds ubiquitinated proteins (blue) and translocates them into the core particle, as shown.

Ubiquitinated proteins are degraded by a large complex known as the **26S proteasome** (M_r 2.5×10^6) (Fig. 27–42). The proteasome consists of two copies each of at least 32 different subunits, most of which are highly conserved from yeasts to humans. The proteasome contains two main types of subcomplexes, a barrel-like core particle and regulatory particles on either end of the barrel. The 20S core particle consists of four rings; the outer rings are formed from seven α subunits, and the inner rings from seven β subunits. Three of the seven subunits in each β ring have protease activities, each with different substrate specificities. The stacked rings of the core particle form the barrel-like structure within which target proteins are degraded. The 19S regulatory particle on each end of the core particle contains 18 subunits, including some that recognize and bind to ubiquitinated proteins. Six of the subunits are ATPases that probably function in unfolding the ubiquitinated proteins and translocating the unfolded polypeptide into the core particle for degradation.

TABLE 27–9 Relationship between Protein Half-Life and Amino-Terminal Amino Acid Residue

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 the β-galactosidase protein modified so that in each experiment it had a different amino-terminal residue. (See Chapter 9 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.

Although we do not yet understand all the signals that trigger ubiquitination, one simple signal has been found. For many proteins, the identity of the first residue that remains after removal of the amino-terminal Met residue, and any other posttranslational proteolytic processing of the amino-terminal end, has a profound influence on half-life (Table 27–9). These amino-terminal signals have been conserved over billions of years of evolution, and are the same in bacterial protein degradation systems and in the human ubiquitination pathway. More complex signals, such as the destruction box discussed in Chapter 12 (see Fig. 12–44), are also being identified.

Ubiquitin-dependent proteolysis is as important for the regulation of cellular processes as for the elimination of defective proteins. Many proteins required at only one stage of the eukaryotic cell cycle are rapidly degraded by the ubiquitin-dependent pathway after completing their function. The same pathway also processes and presents class I MHC antigens (see Fig. 5–22). Ubiquitin-dependent destruction of cyclin is critical to cell-cycle regulation (see Fig. 12–44). The E2 and E3 components of the ubiquitination cascade pathway

(Fig. 27–41) are in fact two large families of proteins. Different E2 and E3 enzymes exhibit different specificities for target proteins and thus regulate different cellular processes. Some E2 and E3 enzymes are highly localized in certain cellular compartments, reflecting a specialized function.

Not surprisingly, defects in the ubiquitination pathway have been implicated in a wide range of disease states. An inability to degrade certain proteins that activate cell division (the products of oncogenes) can lead to tumor formation, whereas a too-rapid degradation of proteins that act as tumor suppressors can have the same effect. The ineffective or overly rapid degradation of cellular proteins also appears to play a role in a range of other conditions: renal diseases, asthma, neurodegenerative disorders such as Alzheimer's and Parkinson's diseases (associated with the formation of characteristic proteinaceous structures in neurons), cystic fibrosis (caused in some cases by a too-rapid degradation of a chloride ion channel, with resultant loss of function; see Box 11–3), Liddle's syndrome (in which a sodium channel in the kidney is not degraded, leading to excessive Na^+ absorption and early-onset hypertension)—and many other disorders. Drugs designed to inhibit proteasome function are being developed as potential treatments for some of these conditions. In a changing metabolic environment, protein degradation is as important to a cell's survival as is protein synthesis, and much remains to be learned about these interesting pathways. ■

SUMMARY 27.3 Protein Targeting and Degradation

- After synthesis, many proteins are directed to particular locations in the cell. One targeting mechanism involves a peptide signal sequence, generally found at the amino terminus of a newly synthesized protein.
- In eukaryotic cells, one class of signal sequences is recognized by the signal recognition particle (SRP), which binds the signal sequence as soon as it appears on the ribosome and transfers the entire ribosome and incomplete polypeptide to the ER. Polypeptides with these signal sequences are moved into the ER lumen as they are synthesized; once in the lumen they may be modified and moved to the Golgi complex, then sorted and sent to lysosomes, the plasma membrane, or transport vesicles.
- Proteins targeted to mitochondria and chloroplasts in eukaryotic cells, and those destined for export in bacteria, also make use of an amino-terminal signal sequence.
- Proteins targeted to the nucleus have an internal signal sequence that is not cleaved once the protein is successfully targeted.
- Some eukaryotic cells import proteins by receptor-mediated endocytosis.
- All cells eventually degrade proteins, using specialized proteolytic systems. Defective proteins and those slated for rapid turnover are generally degraded by an ATP-dependent system. In eukaryotic cells, the proteins are first tagged by linkage to ubiquitin, a highly conserved protein. Ubiquitin-dependent proteolysis is carried out by proteasomes, also highly conserved, and is critical to the regulation of many cellular processes.

Key Terms

Terms in bold are defined in the glossary.

aminoacyl-tRNA 1035
aminoacyl-tRNA synthetases 1035
translation 1035
codon 1035
reading frame 1036
initiation codon 1038
termination codons 1038
open reading frame (ORF) 1039
anticodon 1039
wobble 1041
initiation 1056
Shine-Dalgarno sequence 1056
aminoacyl (A) site 1056
peptidyl (P) site 1056
exit (E) site 1056
initiation complex 1057
elongation 1058
elongation factors 1058
peptidyl transferase 1058
translocation 1060
termination 1061
release factors 1061
polysome 1062
posttranslational modification 1062
nonsense suppressor 1065
puromycin 1066
tetracyclines 1066
chloramphenicol 1066
cycloheximide 1067
streptomycin 1067
diphtheria toxin 1067
ricin 1067
signal recognition particle (SRP) 1069
signal sequence 1068
tunicamycin 1070
coated pits 1074
clathrin 1074
ubiquitin 1075
proteasome 1076

Further Reading

Genetic Code

Bass, B.L. (2002) RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.* **71,** 817–846.

Blanc, V. & Davidson, N.O. (2003) C-to-U RNA editing: mechanisms leading to genetic diversity *J. Biol. Chem.* **278,** 1395–1398.

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.

Klobutcher, L.A. & Farabaugh, P.J. (2002) Shifty ciliates: frequent programmed translational frameshifting in Euplotids. *Cell* **111,** 763–766.

Knight, R.D., Freeland, S.J., & Landweber, L.F. (2001) Rewiring the keyboard: evolvability of the genetic code. *Nat. Rev. Genet.* **2,** 49–58.

Maas, S., Rich, A., & Nishikura, K. (2003) A-to-I RNA editing: recent news and residual mysteries. *J. Biol. Chem.* **278,** 1391–1394.

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

Stadtman, T.C. (1996) Selenocysteine. *Annu. Rev. Biochem.* **65,** 83–100.

Protein Synthesis

Ban, N., Nissen, P., Hansen, J., Moore, P.B., & Steitz, T.A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 angstrom resolution. *Science* **289,** 905–920.
The first high-resolution structure of a major ribosomal subunit.

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

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 adaptor hypothesis and showing that amino acids are not checked after they are linked to tRNAs.

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.

Giege, R., Sissler, M., & Florentz, C. (1998) Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acid Res.* **26,** 5017–5035.

Gingras, A.-C., Raught, B., & Sonenberg, N. (1999) eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu. Rev. Biochem.* **68,** 913–964.

Gray, N.K. & Wickens, M. (1998) Control of translation initiation in animals. *Annu. Rev. Cell Dev. Biol.* **14,** 399–458.

Green, R. & Noller, J.F. (1997) Ribosomes and translation. *Annu. Rev. Biochem.* **66,** 679–716.

Ibba, M. & Soll, D. (2000) Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.* **69,** 617–650.

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

Moore, P.B. & Steitz, T.A. (2003) The structural basis of large ribosomal subunit function. *Annu. Rev. Biochem.* **72,** 813–850.

Ramakrishnan, V. (2002) Ribosome structure and the mechanism of translation. *Cell* **108,** 557–572.
A good overview, incorporating structural advances.

Rodnina, M.V. & Wintermeyer, W. (2001) Fidelity of aminoacyl-tRNA selection on the ribosome: kinetic and structural mechanisms. *Annu. Rev. Biochem.* **70,** 415–435.

Sprinzl, M. (1994) Elongation factor Tu: a regulatory GTPase with an integrated effector. *Trends Biochem. Sci.* **19,** 245–250.

Woese, C.R., Olsen, G.J., Ibba, M., & Soll, D. (2000) Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol. Mol. Biol. Rev.* **64,** 202–236.

Protein Targeting and Secretion

Görlich, D. & Mattaj, I.W. (1996) Nucleocytoplasmic transport. *Science* **271,** 1513–1518.

Hartmann-Petersen, R., Seeger, M., & Gordon C. (2003) Transferring substrates to the 26S proteasome. *Trends Biochem. Sci.* **28,** 26–31.

Higgins, M.K. & McMahon, H.T. (2002) Snap-shots of clathrin-mediated endocytosis. *Trends Biochem. Sci.* **27,** 257–263.

Neupert, W. (1997) Protein import into mitochondria. *Annu. Rev. Biochem.* **66,** 863–917.

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

Rapoport, T.A., Jungnickel, B., & Kutay, U. (1996) Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes. *Annu. Rev. Biochem.* **65,** 271–303.

Schatz, G. & Dobberstein, B. (1996) Common principles of protein translocation across membranes. *Science* **271,** 1519–1525.

Schekman, R. & Orci, L. (1996) Coat proteins and vesicle budding. *Science* **271,** 1526–1532.

Schmid, S.L. (1997) Clathrin-coated vesicle formation and protein sorting: an integrated process. *Annu. Rev. Biochem.* **66,** 511–548.

Varshavsky, A. (1997) The ubiquitin system. *Trends Biochem. Sci.* **22,** 383–387.

Voges, D., Zwickl, P., & Baumeister, W. (1999) The 26S proteasome: a molecular machine designed for controlled proteolysis. *Annu. Rev. Biochem.* **68,** 1015–1057.

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

Problems

1. Messenger RNA Translation Predict the amino acid sequences of peptides formed by ribosomes in response to the following mRNA sequences, assuming that the reading frame begins with the first three bases in each sequence.

(a) GGUCAGUCGCUCCUGAUU
(b) UUGGAUGCGCCAUAAUUUGCU
(c) CAUGAUGCCUGUUGCUAC
(d) AUGGACGAA

2. How Many Different mRNA Sequences 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 about 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 Sequence 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 it? Give reasons for your answer.

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

(5′)CTTAACACCCCTGACTTCGCGCCGTCG(3′)

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

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

(c) If the complementary (nontemplate) strand of this DNA were transcribed and translated, would 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 two amino acids with only one codon. How does the single codon for methionine specify both the initiating residue and interior Met residues of polypeptides synthesized by *E. coli*?

6. Synthetic mRNAs The genetic code was elucidated with polyribonucleotides synthesized either enzymatically or chemically in the laboratory. Given what we now know about the genetic code, 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. Energy Cost of Protein Biosynthesis Determine the minimum energy cost, in terms of ATP equivalents 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 15). From your data, what is the *extra* energy cost of making a protein, in which all the residues are ordered in a specific sequence, compared with the cost of making a polysaccharide containing the same number of residues but lacking the informational content of the protein?

In addition to the direct energy cost for the synthesis of a protein, there are indirect energy costs—those required for the cell to make the necessary enzymes for protein synthesis. Compare the magnitude of the indirect costs to a eukaryotic cell of the biosynthesis of linear (α1→4) glycogen chains and the biosynthesis of polypeptides, in terms of the enzymatic machinery involved.

8. Predicting Anticodons from Codons Most amino acids have more than one codon and attach to more than one tRNA, each with a different anticodon. Write all possible anticodons for the four codons of 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 has/have a wobbly base pair?

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

9. Effect of Single-Base Changes on Amino Acid Sequence Much important confirmatory evidence on the genetic code has come from assessing changes in the amino acid sequence of mutant proteins after a single base has been changed in the gene that encodes the protein. Which of the following amino acid replacements would be consistent with the genetic code if the replacements were caused by a single base change? Which cannot be the result of a single-base mutation? Why?

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

10. Basis of the Sickle-Cell Mutation Sickle-cell hemoglobin has a Val residue at position 6 of the β-globin chain, instead of the Glu residue found in normal hemoglobin A. Can you predict what change took place in the DNA codon for glutamate to account for replacement of the Glu residue by Val?

11. Importance of the "Second Genetic Code" Some aminoacyl-tRNA synthetases do not recognize and 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.

(a) What features of $tRNA^{Ala}$ are recognized by Ala-tRNA synthetase?

(b) Describe the consequences of a C→G mutation in the third position of the anticodon of $tRNA^{Ala}$.

(c) What other kinds of mutations might have similar effects?

(d) Mutations of these types are never found in natural populations of organisms. Why? (Hint: Consider what might happen both to individual proteins and to the organism as a whole.)

12. 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 use 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 an amino acid attached to an incoming tRNA and added to the growing polypeptide is never checked. A proofreading step that hydrolyzed the previously formed peptide bond after an incorrect amino acid had been inserted into a growing polypeptide (analogous to the proofreading step of DNA polymerases) would be impractical. Why? (Hint: Consider how the link between the growing polypeptide and the mRNA is maintained during elongation; see Figs 27–24 and 27–25.)

13. Predicting the Cellular Location of a Protein The gene for a eukaryotic polypeptide 300 amino acid residues long is altered so that a signal sequence recognized by SRP occurs at the polypeptide's amino terminus and a nuclear localization signal (NLS) occurs internally, beginning at residue 150. Where is the protein likely to be found in the cell?

14. Requirements for Protein Translocation across a Membrane The secreted bacterial protein OmpA has a precursor, ProOmpA, which has the amino-terminal signal sequence required for secretion. If purified ProOmpA is denatured with 8 M urea and the urea is then removed (such as by running the protein solution rapidly through a gel filtration column) the protein can be translocated across isolated bacterial inner membranes in vitro. However, translocation becomes impossible if ProOmpA is first allowed to incubate for a few hours in the absence of urea. Furthermore, the capacity for translocation is maintained for an extended period if ProOmpA is first incubated in the presence of another bacterial protein called trigger factor. Describe the probable function of this factor.

15. Protein-Coding Capacity of a Viral DNA The 5,386 bp genome of bacteriophage ϕX174 includes genes for 10 proteins, designated A to K, with sizes given in the table below. How much DNA would be required to encode these 10 proteins? How can you reconcile the size of the ϕX174 genome with its protein-coding capacity?

Protein	*Number of amino acid residues*	*Protein*	*Number of amino acid residues*
A	455	F	427
B	120	G	175
C	86	H	328
D	152	J	38
E	91	K	56