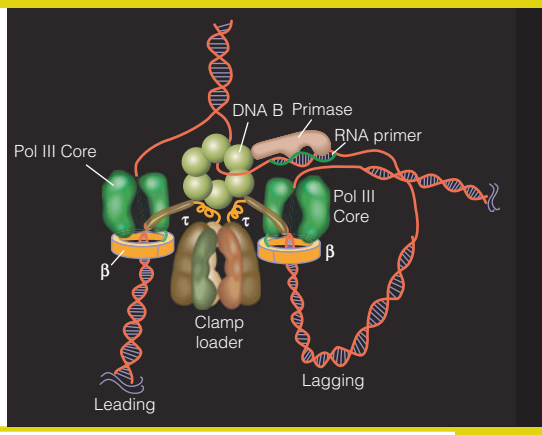


CHAPTER 28

Information Decoding: Translation and Post-translational Protein Processing

We turn now to the most complex process in biological information transfer—the decoding of genetic messages in the four-letter nucleic acid language to amino acid sequences in proteins, expressed in the 20-letter amino acid language. In DNA replication, transcription, and reverse transcription, information transfer is guided strictly by Watson–Crick base pairing between the template nucleic acid and the product, whether it be DNA or RNA. By contrast, when a messenger RNA sequence directs the synthesis of a specific protein, base sequence complementarity is still crucially involved, but a more complex overall process converts information encoded in a nucleotide sequence to information expressed as a specific amino acid sequence.

In terms of the number of components involved—ribosomal RNAs and proteins, transfer RNAs, amino acid activating enzymes, and soluble protein factors—and the number of different proteins in each cell—protein synthesis may well be the most complex of all metabolic processes, and it certainly involves the dominant fraction of a cell's metabolic effort. In a logarithmically growing bacterial cell, as much as 90% of the total metabolic effort may be devoted to protein biosynthesis, with the metabolic machinery for translation accounting for 35% of the cell's dry weight.

When we think of protein synthesis, we consider not only translation, which yields a specific amino acid sequence, but also post-translational processing and traffic, with each protein properly modified and transported to its ultimate intracellular or extracellular destination. We have already seen, for example, that protein processing involves cleavage, as in the conversion of preproinsulin to insulin, and modification of individual amino acids, as in the hydroxylation of proline residues in collagen synthesis or the phosphorylation of specific amino acid residues. We must also consider protein trafficking—how mature, or maturing, proteins are moved to their ultimate destinations. whether inside or outside the cell.

An Overview of Translation

The earliest evidence for the role of ribosomes in protein synthesis came from experiments in which radiolabeled amino acids were injected into rats, followed by isolation of the liver and fractionation of a liver homogenate. The label was shown to be incorporated earliest into ribosomes, either free or membrane-bound in the endoplasmic reticulum. This and related experiments established that the ribosome is the site of protein synthesis. Experiments with cell-free systems soon established the requirements for amino acid activation and for the involvement of small, stable RNAs (transfer RNAs).

In Chapters 4 and 5 we painted an introductory picture of translation. In Figure 4.23 (page 112) we showed how translation involves movement of a ribosome along an mRNA molecule, three nucleotides at a time, with each trinucleotide sequence in mRNA pairing with one specific transfer RNA charged with an amino acid, and with the polypeptide chain growing stepwise, one amino acid per step, from the N-terminus to the C-terminus. Figure 28.1 shows a somewhat

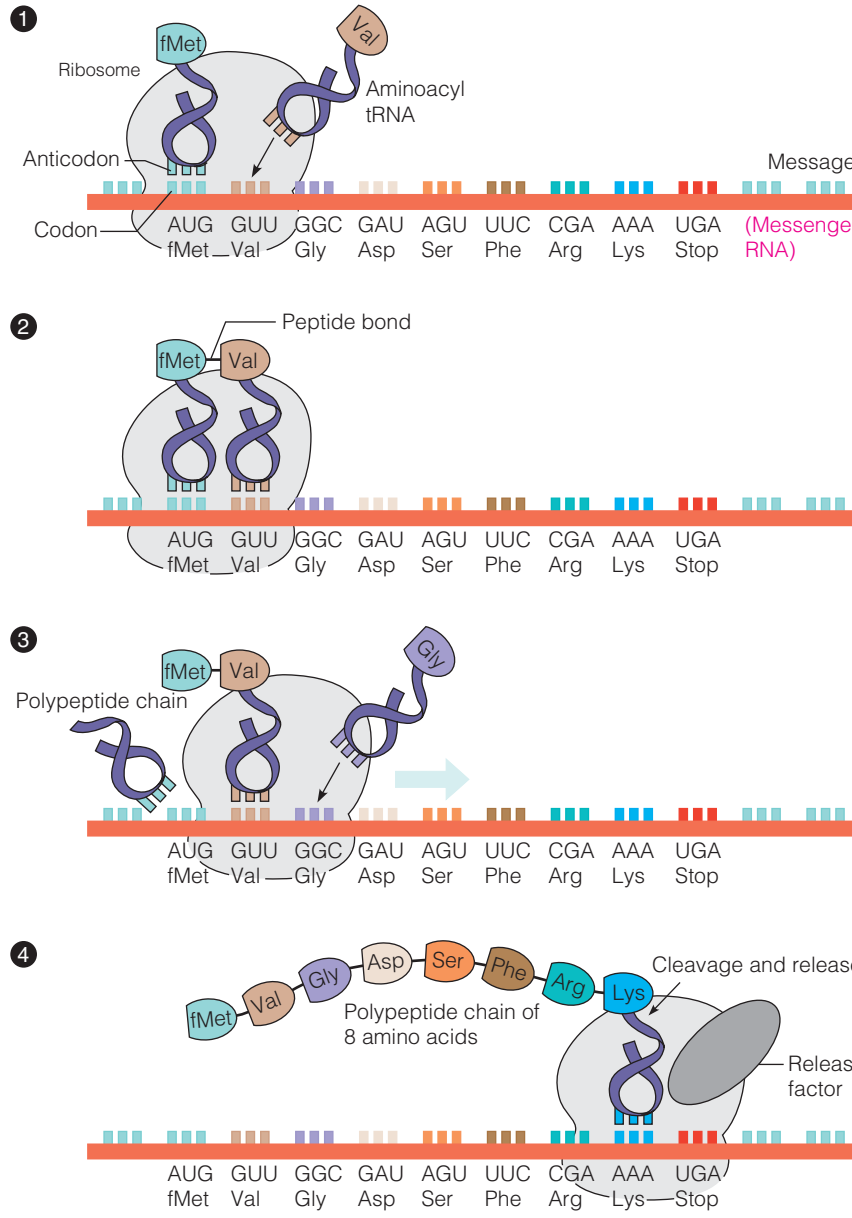


FIGURE 28.1

Translation of an RNA message into a protein. As the ribosome moves along the message, it accepts specific aminoacyl tRNAs in succession, selecting them by matching the trinucleotide anticodon on the tRNA to the trinucleotide codon on the RNA message **(step 1)**. The amino acid (in this example, the second one of the chain, Val) accepts the growing polypeptide chain (in this example the previously bound fMet) **(step 2)**, and the ribosome moves on to the next codon to repeat the process, while releasing the deacylated transfer RNA that held the growing peptide in the previous cycle (the tRNA for fMet, **step 3**). The preceding steps are repeated, adding more amino acids to the chain, until a stop signal is read **(step 4)**, whereupon a protein release factor causes both the polypeptide and the mRNA to be released. The polypeptide shown here is unrealistically short, to illustrate both initiation and termination.

more detailed picture of this process and serves as an overview of translation. In Figure 5.18 (page 152) we showed the genetic code, that is, the correspondence between each of the 64 possible trinucleotide sequences and the 20 amino acids encoded by these sequences. In this chapter we expand upon both the process of protein synthesis and the elucidation and nature of the genetic code.

In 1958, several years before elucidation of the genetic code or the demonstration of messenger RNA, Francis Crick predicted the existence of *adaptor molecules*, each of which would function in translating the genetic message by binding to a specific amino acid and linking it to a molecular code word in the translation machinery. These adaptor molecules turned out to be transfer RNAs. As discussed in Chapter 4, each transfer RNA, or tRNA, molecule is 75–80 nucleotides in length (although some are as large as 93 nucleotides), folded by intramolecular hydrogen bonding into a three-loop structure. Each tRNA molecule is designed to bind one of the 20 amino acids, through the specificity of an amino acid–activating enzyme, more properly called an *aminoacyl-tRNA synthetase*. As shown in Figure 28.2, amino acid activation proceeds at the expense of ATP hydrolysis and results in esterification of the amino acid carboxyl group with the 3′ hydroxyl of the 3′ terminal tRNA nucleotide, yielding an aminoacyl-tRNA.

Transfer RNAs are the adaptor molecules that match amino acid to codon.

Each tRNA contains, in a region known as the **anticodon loop**, a trinucleotide sequence called the **anticodon** that is complementary to the appropriate trinucleotide codon in the message. Thus, the whole set of tRNAs contained in a cell composes a kind of molecular dictionary for the translation—it defines the correspondences between words in the four-letter nucleic acid language (gene sequence) and words in the 20-letter amino acid language (protein amino acid sequence).

The messenger RNA is bound to a ribosome, as shown in Figure 28.1. The aminoacyl tRNAs also bind here, one by one, matching their anticodons to the codons on the message, as shown in Figure 28.1, step 1. The growing peptide chain is transferred from the tRNA to which it is bound to the incoming aminoacyl-tRNA (step 2). The first tRNA is then released, and the ribosome moves one codon length along the message, allowing the next tRNA to come into place, carrying *its* amino acid (step 3). Again, expenditure of energy from high-energy phosphate hydrolysis is required at each step in the movement. As the ribosome moves along the messenger RNA, it eventually encounters a "stop" codon. At this point, the polypeptide chain is released. Step 4 shows a completed, although short, protein. In every cell, of every kind of organism, this remarkable machinery translates the information coded in thousands of different genes into thousands of different proteins. The cellular apparatus that binds all of these components and catalyzes the formation of peptide is the ribosome, a particle composed of both RNA and proteins. A ribosome can bind to mRNA and "read" it, as it moves along the RNA, accepting the charged tRNAs in the order dictated by the message and

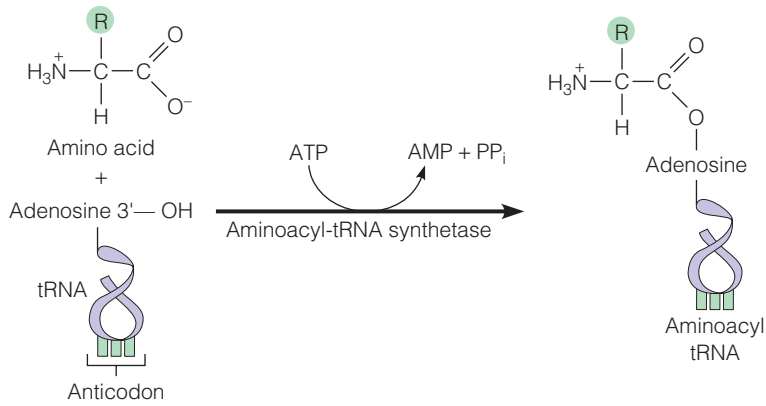


FIGURE 28.2

Activation of amino acids for incorporation into proteins. A specific enzyme, aminoacyl-tRNA synthetase, recognizes both a particular amino acid and a tRNA carrying the corresponding anticodon. This synthetase catalyzes the formation of an aminoacyl tRNA, with accompanying hydrolysis of one ATP to AMP.

incorporating their amino acid residues one by one and in proper order within the growing polypeptide chain.

Messenger RNA is read 5′ → 3′. Polypeptide synthesis begins at the N-terminus.

The mRNA message is always read in the 5′→3′ direction, and the polypeptide chain is synthesized starting with its N-terminal residue. The direction of polypeptide synthesis was established in 1961 in a classic experiment. Howard Dintzis gave reticulocytes (hemoglobin-producing cells) a short pulse of [^{3}H] leucine and isolated the completed hemoglobin molecules at various times after the pulse. After cleaving these molecules into peptides with trypsin, he compared the radioactivity of peptides from various points in the chain. Immediately after the pulse label was added, radioactivity was seen only in chains that were undergoing synthesis before the pulse began and were just being completed during the pulse; the label at this time was found only in the C-terminal peptides. At longer times after the pulse, radioactivity was found to be incorporated into parts of the polypeptide closer and closer to the N-terminus, as synthesis of new protein molecules continued to be initiated. Dintzis therefore concluded that amino acids are added to a polypeptide chain starting at the N-terminus and working toward the C-terminus.

The simple picture of translation we have presented so far leaves a host of questions unanswered. How are tRNA and amino acid matched? How does the ribosome attach to the mRNA and move along it? How does it catalyze peptide bond formation? How does it start and stop translation correctly? How does it avoid making mistakes? Where does the energy for all of this activity come from? To answer such questions, we must dissect the whole process of translation, with careful examination of each of its parts. First, let us consider the genetic code in more detail.

The Genetic Code

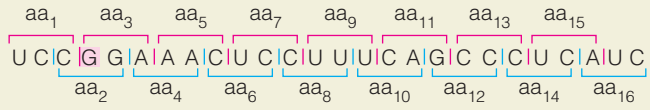
We introduced the genetic code in Chapters 5 and 7. Here, we describe the key experiments that led to deciphering of the code, and we discuss some features of the code, including whether it is universal throughout biological systems.

By the late 1950s it was generally accepted that a protein's amino acid sequence is encoded by the sequence of bases in a nucleic acid template. A triplet code seemed most likely, with three nucleotides specifying one amino acid. Clearly, a doublet code wouldn't work because there are only 16 possible dinucleotide sequences (4×4), and we need at least 20 code words if each amino acid is to have its own code word. So a triplet code seemed the simplest; 64 possible trinucleotides ($4 \times 4 \times 4$) made it likely that some amino acids would have more than one code word.

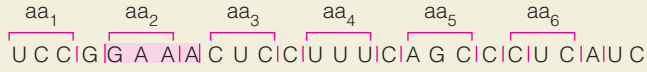
Genetic experiments supported the idea of a triplet code and also a code that is *nonoverlapping* and *unpunctuated*. Figure 28.3 illustrates what we mean by these terms and suggests reasons why overlapping and punctuated codes were rejected.

How the Code Was Deciphered

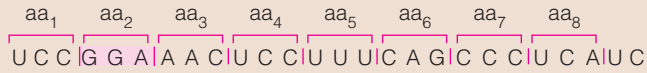
Biochemical elucidation of the code began in 1961 by Marshall Nirenberg and Heinrich Matthaei, with their use of artificial RNA templates for in vitro protein synthesis. Recall from Chapter 27 (page 1129) that the enzyme polynucleotide phosphorylase will catalyze the nontemplate-dependent synthesis, from a mixture of ribonucleoside diphosphates, of a random-sequence RNA whose nucleotide composition matches that of the medium. Nirenberg and Matthaei polymerized UDP with the enzyme to synthesize polyU, a polyribonucleotide containing only UMP residues. When this artificial RNA was placed in a cell-free system containing a bacterial extract, ATP, GTP, and the 20 canonical amino acids (i.e., those commonly found in proteins), the product was a polypeptide containing only phenylalanine. Thus, the genetic code word for phenylalanine was shown to be a



(a) Overlapping code. There will be statistical regularities between adjacent amino acid residues. Point mutations (magenta) will be able to change two amino acid residues.



(b) Punctuated code. Deletions of four nucleotides (or multiples thereof) will restore the reading frame.



(c) Unpunctuated code. Deletions of three nucleotides (or multiples thereof) will restore the reading frame. This is the actual form of the code.

FIGURE 28.3

Three conceivable kinds of genetic codes. Early research on the nature of the code quickly showed that a nonoverlapping, unpunctuated code **(c)** fit all experimental observations.

specific sequence of UMP residues—three if we are really dealing with a triplet code. In short order, polyC was shown to encode only proline and polyA only lysine.

Establishing code words for the other 17 amino acids was trickier. For example, consider the enzymatic synthesis of a polyribonucleotide from a nucleotide mixture containing CDP and ADP in a 5:1 molar ratio. Because the base composition of the polymer reflects the substrate ratio, the polymer contains eight trinucleotide codons, with CCC 125 times more abundant than AAA ($5 \times 5 \times 5$). Codons with two A and one C (2A1C — AAC, ACA, CAA) are five-fold more abundant than AAA, and those with one A and two C (2C1A — CCA, CAC, ACC) are 25-fold more abundant. When this polymer was used by Nirenberg and Matthaei, it stimulated the incorporation of proline, histidine, threonine, glutamine, asparagine, and lysine in molar ratios of 100, 23.4, 20, 3.3, 3.3, and 1, respectively. The data were best explained by assuming that the polymer contained two codons for Pro (CCC and 2C1A) and two for Thr (2A1C and 2C1A). The codon for lysine was already known to be AAA. Codons assigned for Asp and Gln were both 2A1C and, for His, 1A2C. This and other experiments with random-sequence polymers were able to establish the nucleotide *composition* of most codons, but not their *sequence*.

Two approaches led to the identification of codon sequences. First, H. Gobind Khorana synthesized polyribonucleotides of regular repeating sequence. For example, the polymer UCUCUCUC··· was shown to direct synthesis of an alternating copolymer, Ser-Leu-Ser-Leu-Ser-Leu···. If the code is triplet and nonoverlapping, this means that UCU is the codon for either serine or leucine and CUC encodes the other amino acid. Because serine was known to have a 2U1C codon and leucine a 2C1U codon, this established UCU as a serine codon and CUC as a leucine codon. When a trinucleotide was used as the repeating unit, a different result was seen, as shown in Figure 28.4. The polymer AAGAAGAAG··· directed synthesis of three homopolypeptides—polylys, polyarg, and polyglu. This experiment didn't give codon sequences, but it did establish the triplet and nonoverlapping nature of the code. Here, the nature of the product was set by the initial **reading frame**—the trinucleotide sequence chosen for the first amino acid incorporation event. If GAA was selected, for example, every subsequent codon would also be GAA, making all amino acids in the product identical. The experiment did establish GAA, AGA, and

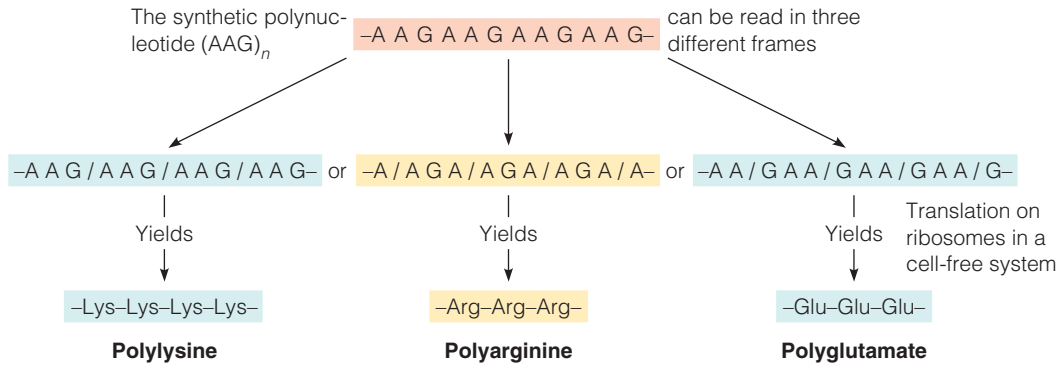


FIGURE 28.4

Use of synthetic polynucleotides with repeating sequences to decipher the code. This example shows how polypeptides derived from the $(AAG)_n$ polymer were used to confirm the triplet code and help identify codons. The polymer $(AAG)_n$ can yield three different polypeptides, depending on which reading frame is employed.

AAG as codons for the three amino acids, but could not lead directly to assignment of each amino acid to one codon.

Experiments of this kind identified many code words, but in 1964 Philip Leder and Marshall Nirenberg developed a new and rapid method for codon assignment that made it possible to complete the deciphering of the code. Leder and Nirenberg found that synthetic trinucleotides would bind to ribosomes and specify the binding of specific tRNAs. For example, UUU and UUC stimulated binding of phenylalanine tRNAs to ribosomes, and CCC and CCU stimulated binding of proline tRNA. Such experiments provided unequivocal evidence for the *redundancy* of the code because several different codons were found to correspond to a single amino acid. By the combined use of these techniques, the entire genetic code was established within a few years after demonstration of polyU-directed Phe incorporation.

Features of the Code

In the genetic code, as shown in Figure 28.5, 61 of the 64 trinucleotides are "sense" codons, that is, they code for one amino acid. The remaining three are normally "nonsense" codons in that they do not code for an amino acid (with some exceptions, see Table 28.1 and discussion on page 1180). When a ribosome encounters a nonsense codon (UAG, UAA, or UGA) in the correct reading frame, there is no aminoacyl-tRNA in the cell containing a matching anticodon, and translation ceases. As we see later, these codons are used as part of the normal machinery for terminating translation of a message. The code is *degenerate* (or redundant) in the sense that most amino acids have more than one codon and *unambiguous*, in the sense that a particular trinucleotide encodes one and only one amino acid. There are some exceptions to this generalization, as summarized in Table 28.1 and discussed later (see page 1180). In other words, the genetic code is almost, but not quite, universal.

The code is almost, but not quite, universal.

Biological Validity of the Code

As described above, the assignment of genetic code words to amino acids was carried out strictly through the use of in vitro systems—amino acid incorporation directed by synthetic templates and assays of aminoacyl-tRNA binding to ribosomes. How could we be assured that these codon assignments are valid for translation of messages in living cells? Some of the validation came from amino acid sequence analysis of mutant human hemoglobins (Chapter 7). Most of the amino acid sequence changes could be accounted for by substitution mutations involving a single base, the most frequent spontaneous mutation. For example, the Glu → Val substitution seen with sickle-cell hemoglobin could be accounted for by changing a GAA Glu codon to a GUA Val codon, or GAG to GUG.

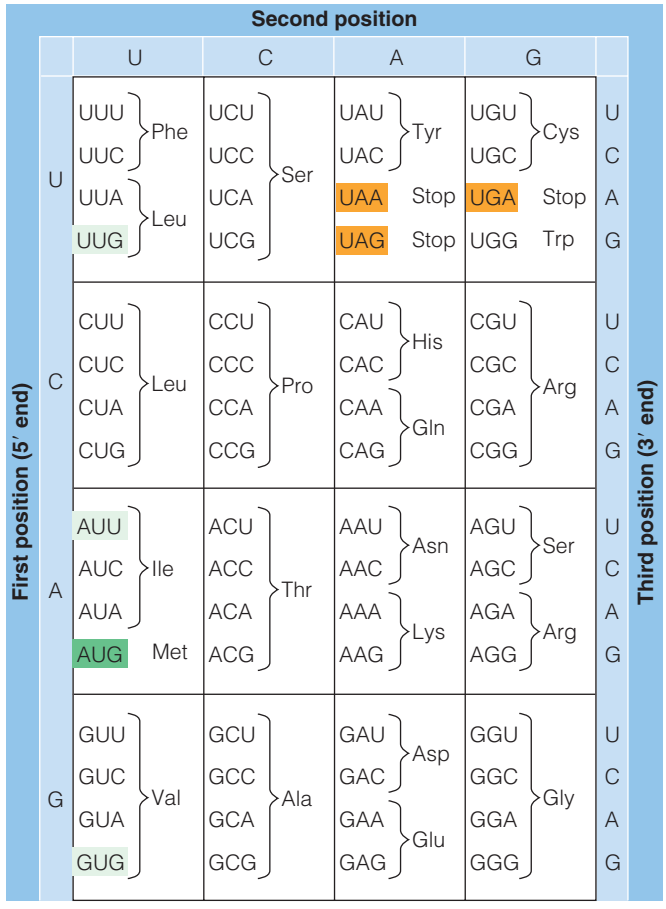


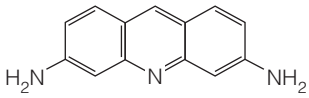
FIGURE 28.5

The genetic code (as written in RNA). We show here the genetic code as used in most organisms. Chain termination, or "stop," codons are shown in orange, and the usual start codon, AUG, is dark green. Other, rarely used, start codons are shown in light green. When AUG is used as a start codon, it codes for *N*-formylmethionine (fMet) in prokaryotes or methionine (Met) in eukaryotes; see page 1182. Otherwise, it codes for Met. Exceptions to these codon assignments are given in Table 28.1.

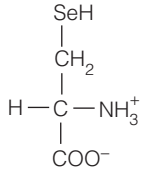
TABLE 28.1 Modifications of the genetic code

Codon	Usual Use	Alternate Use	Where Alternate Use Occurs
AGA AGG	Arg	Stop, Ser	Some animal mitochondria, some protozoans
AUA	Ile	Met	Mitochondria
CGG	Arg	Trp	Plant mitochondria
CUU CUC CUA CUG	Leu	Thr	Yeast mitochondria
AUU GUG UUG	Ile Val Leu	Start (*N*-fMet)	Some prokaryotes[a]
UAA	Stop	Glu	Some protozoans
UAG	Stop	Pyrrolysine Glu	Various archaea Some protozoans
UGA	Stop	Trp Selenocysteine Selenocysteine and Cys	Mitochondria, mycoplasmas Widespread[a] *Euplotes*

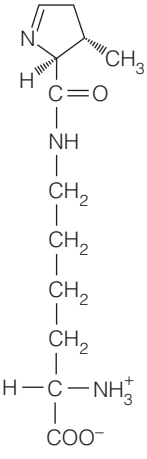
[a]Depends on context of message, other factors.



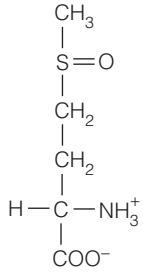
Proflavin



Selenocysteine



Pyrrolysine



Methionine sulfoxide

The code is redundant. Several codons may correspond to a single amino acid, sometimes via wobble in the 5′ anticodon position.

Other important validation experiments were carried out by George Streisinger, using the T4 bacteriophage lysozyme system. Lysozyme is a phage-coded enzyme responsible for rupture of the host bacterium after a cycle of phage reproduction. Lysozyme mutants are easy to detect because they can produce phage but cannot lyse the host cell. Mutations were induced by treating phage-infected cells with **proflavin**, a large planar molecule that can fit, or *intercalate*, between successive base pairs in DNA and induce *frameshift* mutations, that is, additions or deletions of a single base, which alter the reading frame (Figure 7.26, page 257). In one such frameshift mutation, wild-type function was restored by a second mutation. Sequence analysis of the double-mutant lysozyme showed five changes from the wild-type sequence. The data were consistent only with the assumption that the double mutant was created by a single-base insertion and a single-base deletion, which restored the reading frame (Figure 28.6). All of the codons used experimentally to infer the nucleic acid sequence (long before DNA sequencing had been developed) were consistent with the codon assignments as determined in vitro.

Deviations from the Genetic Code

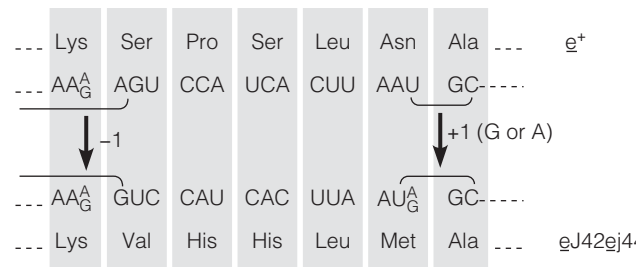
Why has the genetic code remained almost unchanged over so vast an evolutionary span? Perhaps it is simply because even small codon changes could be devastating. A single codon change could alter the sequence of nearly every protein made by the organism. Some of these changes would almost certainly have lethal effects. Therefore, codon changes have been opposed by intense selective pressure during evolution. They represent changes in the most basic rules of the game.

Yet significant deviations do occur, most notably, differences in the mitochondrial code and coding for the "21st and 22nd" amino acids, namely, selenocysteine and pyrrolysine (Chapter 5, page 144). A significant change in the mitochondrial code, as shown in Table 28.1, is the change in AUA from an isoleucine to a methionine codon. It has been argued that this represents an adaptation within mitochondria to oxidative stress. Methionine, whether free or as a residue in a protein, is readily oxidized, but just as readily reduced by methionine sulfoxide reductase. Hence, it is argued that it is advantageous for a mitochondrion to have increased methionine abundance in its proteins, to absorb reactive oxygen species that otherwise would be attacking less resilient targets. Indeed, mitochondrial proteins do have methionine in higher abundance than do proteins from other cell compartments.

Selenocysteine (21st amino acid) and pyrrolysine (22nd amino acid) are translated differently. Both use codons that are otherwise used in translation termination—UGA for selenocysteine (Sec) and UAG for pyrrolysine (Pyl). A special transfer RNA, $tRNA^{Sec}$, is a substrate for a seryl-$tRNA^{Sec}$ synthetase, which charges serine directly, to give Ser-$tRNA^{Sec}$. (Note the convention; Ser refers to the amino acid bound and superscript Sec denotes the amino acid corresponding to the anticodon on that tRNA molecule.) The tRNA-linked Ser is then converted to Sec by a two-step process beginning with phosphorylation of the serine hydroxyl group. The resultant Sec-$tRNA^{Sec}$ responds to a UGA codon. For a particular

FIGURE 28.6

Validation of the genetic code by amino acid sequence analysis of T4 phage lysozyme mutants. *e* is the gene for lysozyme, a portion of whose amino acid sequence is shown. One of two proflavin-induced mutations, either eJ42 or eJ44, disrupted the reading frame by deleting one base pair, and a second mutation restored the wild-type reading frame by inserting a base pair, but altering the amino acid sequence between the two mutant sites. The mRNA sequences encoding these five altered amino acids are inferred from the genetic code and the known action of the mutagen.



Adapted with permission from Eric Terzaghi from *Proceedings of the National Academy of Sciences of the United States of America* 56:500-507, E. Terzaghi, Y. Okada, G. Streisinger, J. Emrich, M. Inouye, and A. Tsugita, Change of a sequence of amino acids in phage T4 lysozyme by acridine-induced mutations, 1966.

UGA to be translated as Sec rather than for termination, that UGA must have a special Sec Insertion Sequence (SECIS) available, usually in the 3′-untranslated region (3′UTR). Although selenocysteine is rather rare among proteins, the human proteome has been shown to contain 25 selenoproteins. As mentioned earlier (Chapter 15), some of these proteins participate in oxidant protection.

Pyrrolysine, by contrast, has a much narrower distribution, having been found so far in only about 1% of all sequenced genomes, mostly methanogenic archaea. Pyrrolysine is converted directly to a pyrrolysyl-tRNA by its own amino acyl-tRNA synthetase. The resultant Pyl-tRNA^{Pyl} has an anticodon that pairs with UAG, normally used in chain termination. So far it is not clear whether any of the UAGs in these genomes are read as stop codons or whether all encode Pyl.

Finally, although we say that the code is unambiguous, that concept may need revision in light of recent work with ciliated protozoa of the genus *Euplotes*. This organism uses UGA as one of three cysteine codons (the others are UGU and UGC). At least one gene in *E. crassus* has UGA triplets encoding both cysteine and selenocysteine. Sequence context is obviously key to ensuring correct insertional specificity. Because this organism also uses UGA as a tryptophan codon in its mitochondria, UGA is busy indeed.

The Wobble Hypothesis

If you examine the codon table shown in Figure 28.5, you will note that, in general, each amino acid is characterized by the first two codon letters. For example, all four Pro codons start with CC, and all four Val codons start with GU. Thus, redundancy is usually expressed in the third letter—ACU, ACC, ACA, and ACG all code for threonine. Soon after the code was deciphered, it was recognized that a single tRNA may recognize several different codons. The multiple recognition always involves the 3′ residue of the codon and therefore the 5′ residue of the anticodon.

In 1966, Francis Crick proposed that the 5′ base of the anticodon was capable of "wobble" in its position during translation, allowing it to make alternative (non-Watson–Crick) hydrogen-bonding arrangements with several different codon bases. An example is shown in Figure 28.7. G in the 5′ anticodon position can pair with either C or U in the codon, depending on the relative orientation of the pair. Considering both base-pairing possibilities and the observed selectivity of tRNAs, Crick proposed the set of "wobble rules" given in Table 28.2. This hypothesis nicely explains the frequently observed degeneracy in the 3′ site of the codon. The rather uncommon nucleoside, *inosine* (I, Chapter 22), is found in a number of anticodons, where it shows the ability to pair with A, U, or C.

Not all cases of multiple codon use involve translation of a single tRNA using wobble. As an example, consider the six leucine codons. Four of the six begin with CU and in principle could be translated by two different tRNAs, using wobble. However, the remaining two codons, UUA and UUG, will require a different anticodon, such as 3′-AAU-5′, which could translate both codons. In fact, *E. coli* contains five different leucine tRNAs and multiple **isoaccepting tRNAs**—tRNAs accepting and translating the same amino acid are common.

Codon Bias

Redundancy of the genetic code means that several nucleotide triplets can encode the same amino acid—leucine, for example, with six codons. In principle, a silent mutation, such as CUA ⟶ CUG, should have no biological consequences because both triplets encode leucine. Yet we find that use of degenerate codons by certain organisms is highly selective. In an extreme case, about half of all 64 codons are used either negligibly or not at all by the bacterium *Thermus thermophilus*. Although we don't know the evolutionary mechanisms leading to such

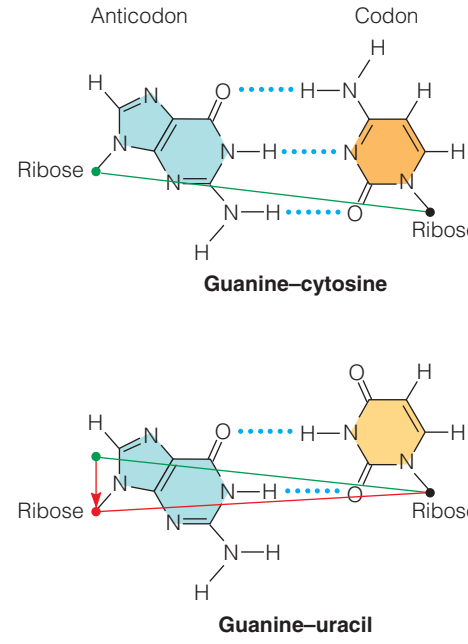


FIGURE 28.7

The wobble hypothesis. As an example, we show how the anticodon base G can pair with either C or U in a codon. Movement ("wobble") of the base in the 5′ anticodon position is necessary for this capability (see arrow).

TABLE 28.2 Base-pairing capabilities in wobble pairs

Base at 5′ Position in Anticodon		Base at 3′ Position in Codon
G	pairs with	C or U
C	pairs with	G
A	pairs with	U
U	pairs with	A or G
I	pairs with	A, U, or C

asymmetry in codon selection, knowledge of codon bias has practical significance, for those wishing to express recombinant eukaryotic proteins in bacteria. Consider *E. coli*, a frequently used host for recombinant gene expression. Of the six arginine codons, two, AGA and AGG, are rarely used in *E. coli*, meaning that each of these triplets represents fewer than 1% of the arginine codons in the entire genome. Related to this is that the intracellular concentration of the tRNA with the anticodon, 3′-TCI-5′, which can translate these two rare codons, is quite low. This means that a recombinant gene with more abundant representation of these codons will be poorly expressed after transfer into *E. coli*. This situation can be remedied by site-directed mutagenesis of the recombinant genes to change these rare codons to arginine codons that are more abundant, and more efficiently translated, in the *E. coli* genome. An alternative approach is to engineer the *E. coli* host for overexpression of the rare tRNA so that the codons AGA and AGG can be efficiently translated.

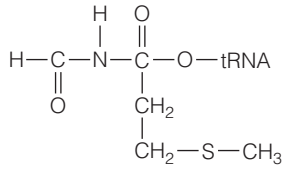
An advantage of a code design that has synonymous codons with similar structures is that many mutations involving single-base changes are silent because a codon change such as the CUA → CUG mentioned above doesn't change the sense of the genetic message. Not only are many single-base changes silent, but many more are conservative, in the sense that a mutation may substitute a structurally similar amino acid that can be tolerated by the protein with no loss of function. For example, each of six leucine codons can be converted to a codon for the closely related valine by a single-base change. This suggests that the code has evolved to maximize genetic stability.

Stopping and Starting

Because the messenger RNA is invariably longer than the open reading frame that is to be translated, specific start and stop signals are required to begin and end translation. In almost all organisms, UAA, UAG, and UGA are used for stop signals and do not code for any amino acid (with the exceptions discussed above). A stop signal indicates that translation is to terminate and the polypeptide product is to be released by the ribosome. Clearly, three stop signals are more than is absolutely necessary, so it is not surprising to find that these codons are also used for designating amino acids in mitochondria and in other special cases (see Table 28.1).

Prokaryotic messengers contain translational start and stop signals, as well as a sequence that aligns the mRNA on the ribosome.

Although nature has been generous in designating stop signals, it has been stingy in apportioning starts. The start signal commonly used in translation is AUG, which also serves as the single methionine codon. How does the ribosome know how to interpret this triplet properly so as to distinguish between internal Met sites and start sites? The answer is that the 5′ end of any message contains specific sequences to ensure that it is correctly attached to the ribosome (see page 1183). As the message begins to be read, the *first* AUG encountered is interpreted as a start signal, and translation begins. Although prokaryotic and eukaryotic cells handle this situation somewhat differently, the consequence is that *N*-formylmethionine (in prokaryotes) or methionine (in eukaryotes) is usually the first amino acid incorporated into a polypeptide chain. Therefore, all proteins start with *N*-fMet or Met, at least when they are first synthesized. However, in most cases this residue is either deformylated or removed as translation proceeds. Any AUG encountered after the start is treated as a signal to incorporate methionine within the sequence at that point. Very occasionally, GUG (normally valine), UUG (normally leucine), or AUU (normally isoleucine) serves as a prokaryotic start codon when located near the 5′ end of a message (see Table 28.1). When they do, however, they code for *N*-formylmethionine in the first position. In other positions these triplets are read as normal codons.



N-formylmethionyl-tRNA

The Major Participants in Translation: mRNA, tRNA, and Ribosomes

mRNAs

As we indicated in Chapter 27, eukaryotic messenger RNAs are quite different from prokaryotic mRNAs. Prokaryotic mRNAs are more complex because many or most are *polycistronic*; they encode two or more polypeptide chains. This means that the mRNA sequence must be punctuated so that translation of the RNA corresponding to each gene is controlled by its own initiation and termination signals. Eukaryotic messages almost always encode just one protein, but the mRNA structure is the result of post-transcriptional processing far more extensive than that seen in prokaryotic systems.

As a good example of a prokaryotic mRNA, consider that produced by transcription of the *E. coli lac* operon, which was introduced in Chapter 27 and receives further attention in Chapter 29. This group of three linked genes—*lacZ*, *lacY*, and *lacA*—controls the utilization of lactose and related sugars by bacteria. As shown in Figure 28.8, these three genes are expressed as a single mRNA molecule some 5300 nucleotides in length. Within this mRNA are three **open reading frames**, corresponding to the *lacZ*, *Y*, and *A* genes. An open reading frame is a sequence within a messenger RNA, bounded by start and stop codons, that can be continuously translated. Each open reading frame has its own start and stop signals, and you can see that these signals vary considerably. There is extra, untranslated RNA between the reading frames and at the ends. The regions 5′ to each start signal contain sequences rich in A and G, which help to align the mRNA on the ribosome so that translation can begin at the proper points and in the correct reading frame. Such attachment sequences, found on all prokaryotic mRNAs, are called *Shine–Dalgarno sequences*, after J. Shine and L. Dalgarno, who first described them. A Shine–Dalgarno sequence can base-pair with a sequence contained in the ribosomal RNA, as shown in Table 28.3, to produce a proper alignment for starting translation. The different attachment sequences appear to have different affinities for ribosomes. For example, the three genes of the *lac* operon (Figure 28.8) are not translated to equal extents—*lacZ* is translated much more frequently than *lacY* or *lacA*.

Shine–Dalgarno sequences help align ribosomes on mRNAs to properly start translation.

FIGURE 28.8

The *lac* operon mRNA. The mRNA for the *E. coli lac* operon is about 5300 nucleotides long and contains the open reading frames for the *lacZ*, *lacY*, and *lacA* genes, each flanked appropriately by start, stop, and Shine–Dalgarno (SD) sequences.

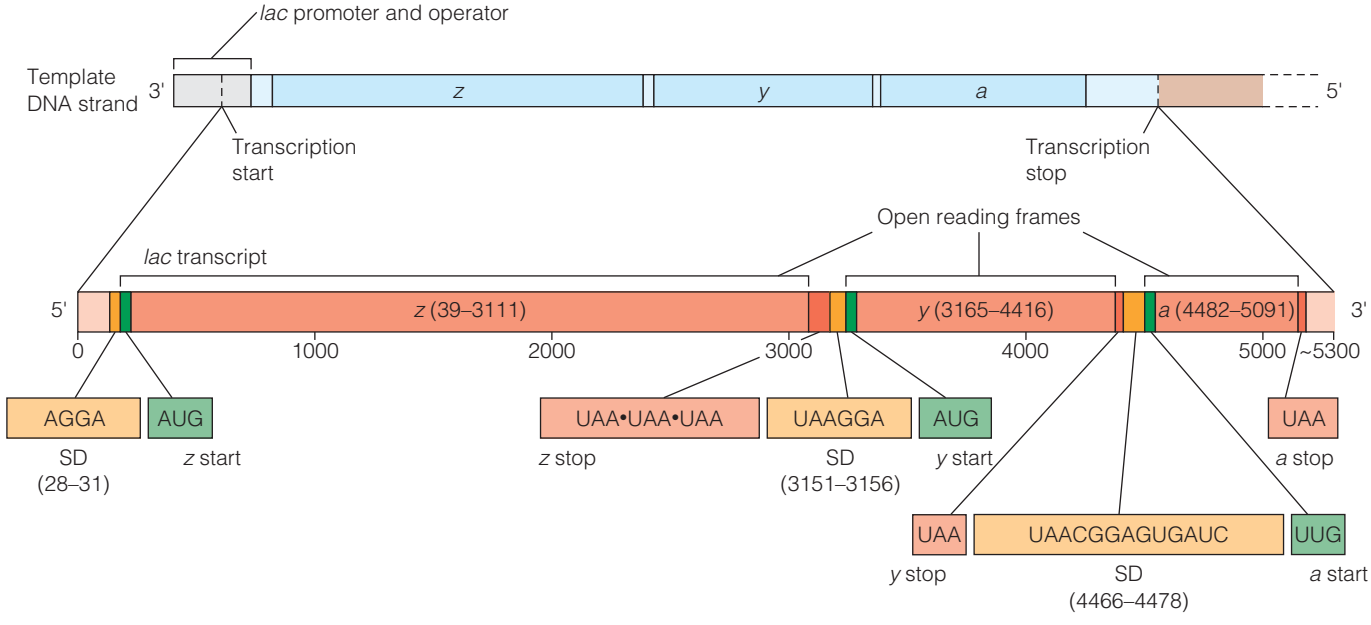


TABLE 28.3 Representative Shine–Dalgarno sequences

Message for	Shine–Dalgarno Sequence
Ribosomal protein L10	SD sequence Start 5′ AGGAGCAAAGCUA AUG 3′ mRNA 3′ AUUCCUCCA 5′ Complementary 3′ end of 16S ribosomal RNA
E. coli lac z	5′ AGGAAACAGCU AUG 3′ 3′ AUUCCUCCA 5′
λ phage Cro	5′ UAAGGAGGUUGU AUG 3′ 3′ AUUCCUCCA 5′

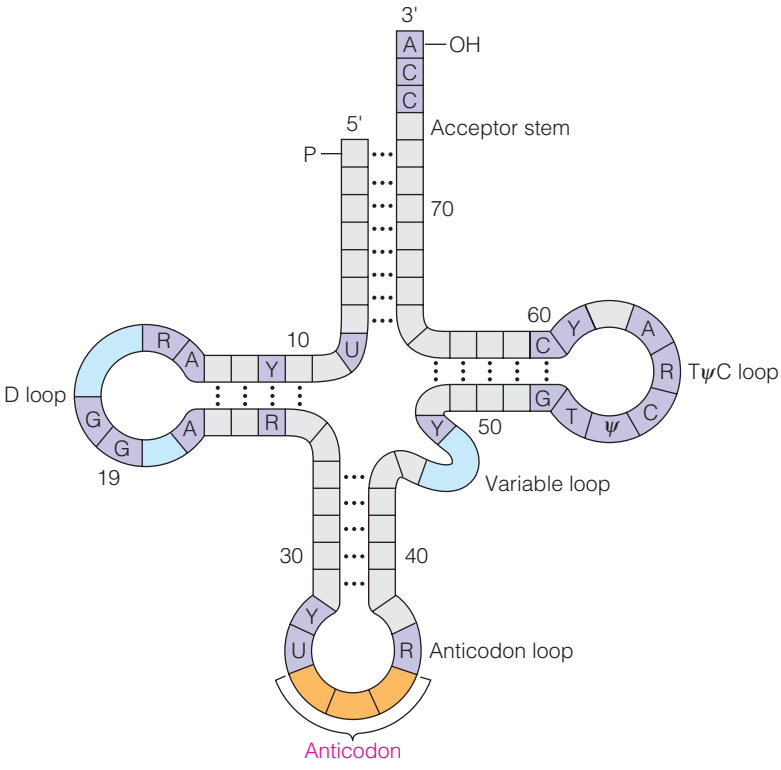
The mRNA produced from the *lac* operon has all the basic elements necessary for its function: sequences to align it properly on the ribosome and sequences that start and stop translation at the proper points. Many mRNAs also have possibilities for forming three-dimensional secondary and tertiary structures, which can participate in regulating the relative production of the various protein products. We shall return to this point in Chapter 29.

Transfer RNA

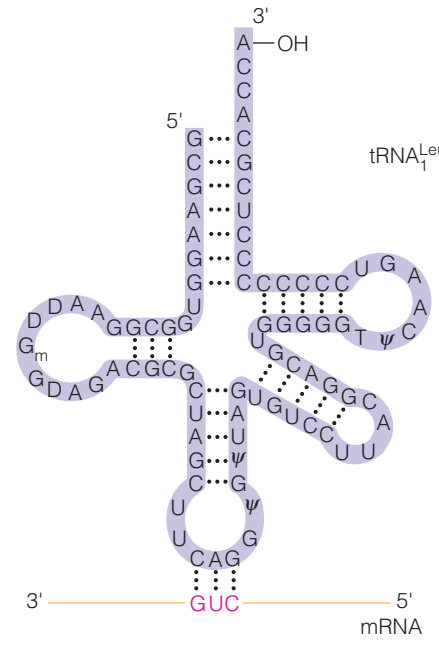
Any cell, prokaryotic or eukaryotic, contains a battery of different types of tRNA molecules sufficient to incorporate all 20 amino acids into protein. This does not mean that there need be as many tRNA types as there are codons, for, as noted earlier, some tRNAs can recognize more than one codon, when the difference is in the third, or wobble, position. *E. coli*, for example, has about 40 different tRNAs—plenty to code for all amino acids, but not as many as the 61 amino acid codons. As noted on page 1180, the tRNA specific to a given amino acid is designated by writing the amino acid as a superscript, for example, $tRNA^{Ala}$.

Transfer RNA was the first natural polynucleotide sequence to be determined, in a pioneering study of yeast $tRNA^{Ala}$ by Robert Holley in 1965. Since then, thousands of tRNAs have been sequenced. All have the general structure shown schematically in Figure 28.9a and have similar sequences of about 70 to 80 nucleotides or more. There is, however, considerable variation in detail, as shown in the examples in Figure 28.9b and c. Furthermore, the tRNAs are unique among RNA molecules in their high content of unusual and modified bases, three of which are shown in Figure 28.10. Biosynthesis of the modified bases always occurs *post-transcriptionally*, as mentioned in Chapter 27. For example, an isomerase converts a uridine residue (1-ribosyluracil) to the unusual C-glycoside pseudouridine (5-ribosyluracil), and *S*-adenosylmethionine-dependent methyltransferases are responsible for converting standard bases to their methylated derivatives.

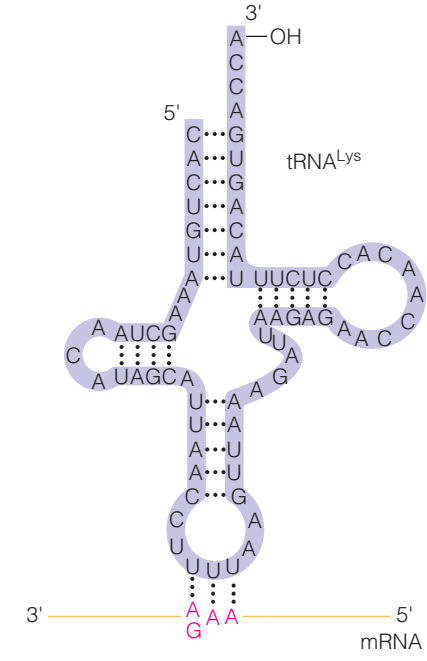
Cloverleaf models of the kind shown in Figure 28.9 are useful for showing the general pattern of hydrogen bonding and denoting the functional parts of the tRNA. The *anticodon triplet* in the loop at the bottom is complementary to the mRNA codon and will make base pairs with it. Because the codon and anticodon, when paired, constitute a short stretch of double-stranded RNA, their directions must be antiparallel. In Figure 28.9 we have written the tRNA molecules with their 5′ ends to the left. Therefore, the messenger RNA, when shown in such figures, is written with its 5′ end to the right, opposite to the normal convention.



(a)



(b)



(c)

FIGURE 28.9

Structure of tRNAs. (a) Generalized tRNA structure. The positions of invariant and rarely varied bases are shown in purple. Regions in the D loop and the variable loop that can contain different numbers of nucleotides are shown in blue. The anticodon is shown in orange. **(b)** A leucine tRNA from *E. coli*. **(c)** A human mitochondrial tRNA for lysine. Code for bases: Y = pyrimidine, R = purine, ψ = pseudouridine, T = ribothymidine, and D = dihydrouridine (see Figure 28.10).

The *acceptor stem* at the top of the cloverleaf figure is where the amino acid will be attached, at the 3′ terminus of the tRNA. This stem always has the sequence 5′···CCA—OH 3′. Other common features of tRNA molecules are the *D loop* and the TψC *loop*, regions that contain a substantial fraction of invariant positions and frequently contain modified or unusual bases as well. The so-called *variable loop* is indeed variable, both in nucleotide composition and in length, as Figure 28.9 demonstrates.

Although cloverleaf models are convenient for depicting the primary structure and some elements of secondary structure, they are not good three-dimensional representations of tRNA molecules. X-ray diffraction studies of tRNA molecules

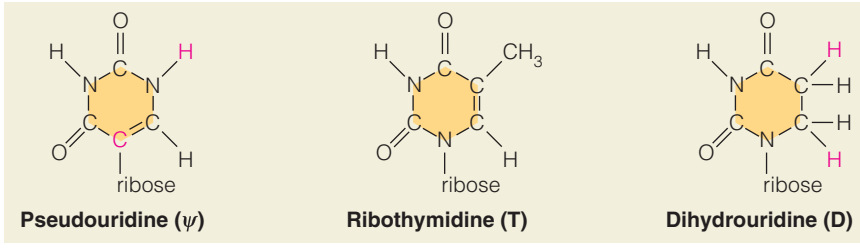


FIGURE 28.10

A sampling of the modified and unusual bases found in tRNAs.

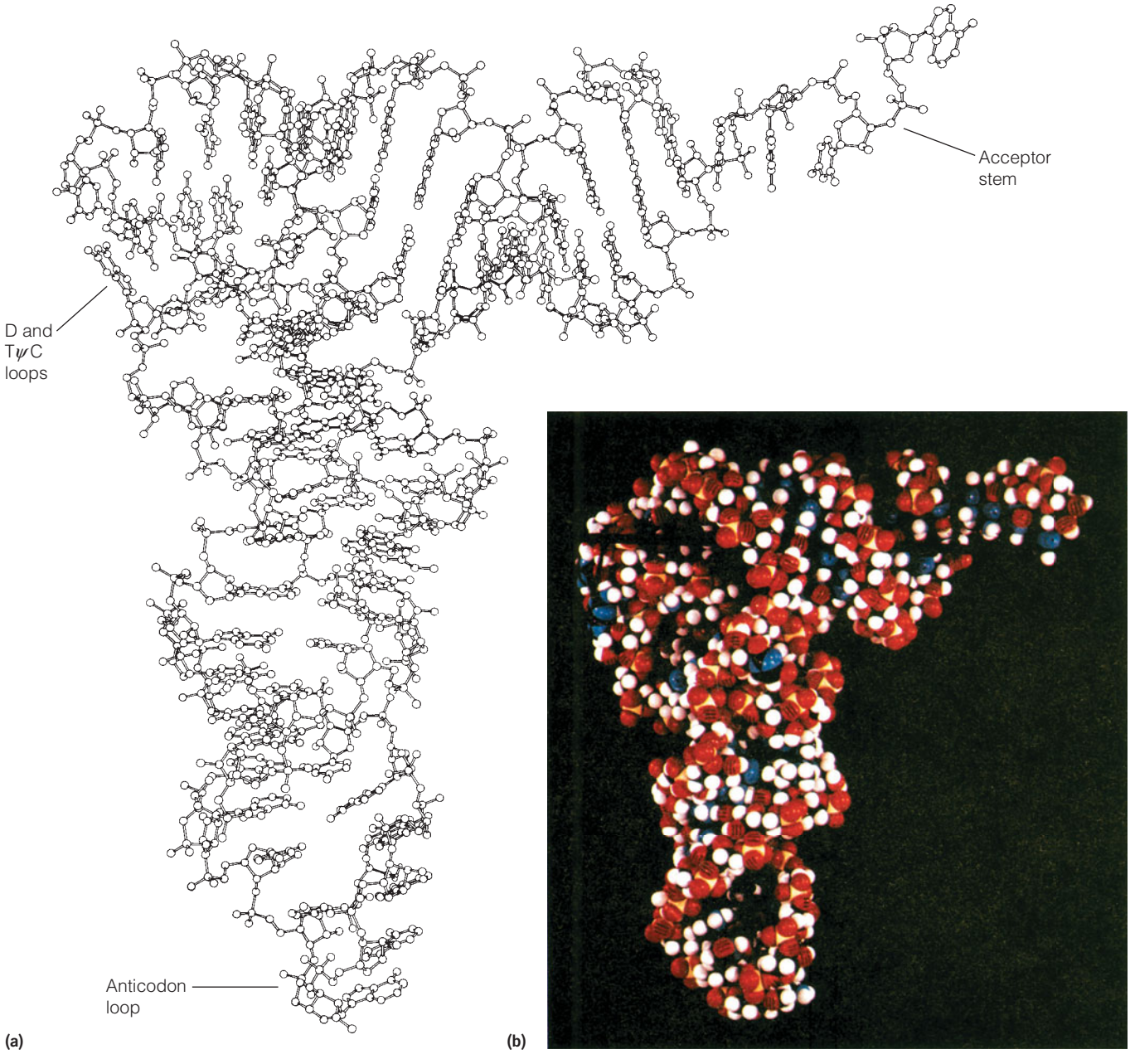


FIGURE 28.11

Model of yeast phenylalanine tRNA derived from X-ray diffraction studies. The anticodon is at the bottom, the 3′ acceptor stem at the upper right. **(a)** Drawing showing all atomic positions. **(b)** Space-filling model.

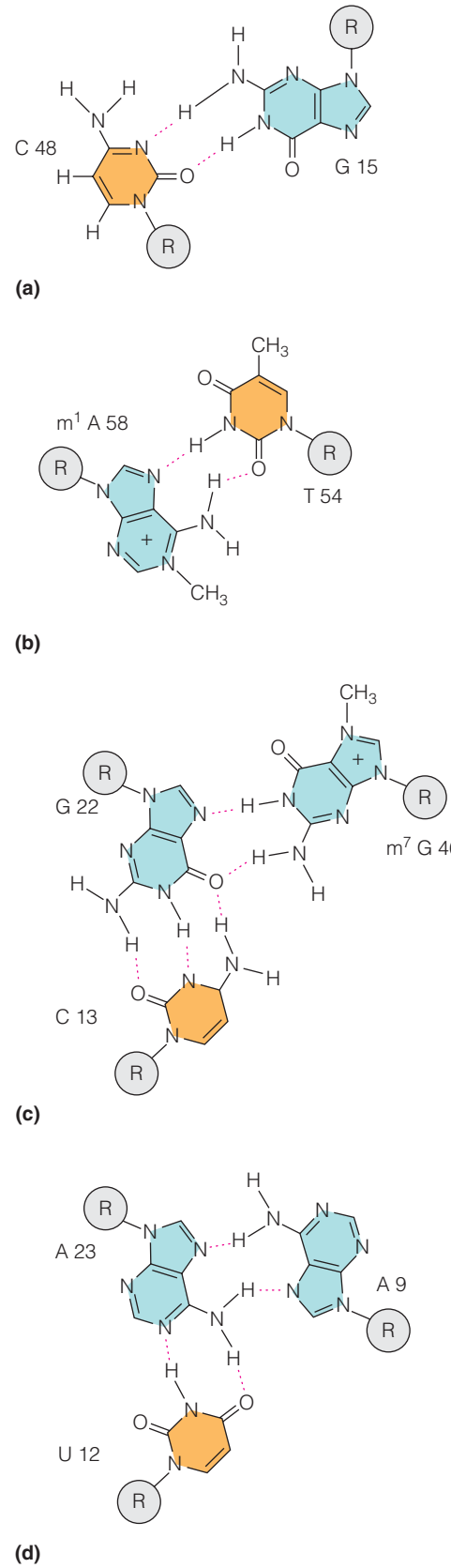
From *Science* 185:435–440, S. H. Kim, F. L. Suddath, G. J. Quigley, A. McPherson, J. L. Sussman, A. H. J. Wang, N. C. Seeman, and A. Rich, Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. © 1974. Reprinted with permission from AAAS and Sung-Hou Kim.

All tRNAs share a general common structure that includes an anticodon loop, which pairs with codons, and an acceptor stem, to which the amino acid is attached.

have revealed that the real molecular shape is more complex, as you can see in Figures 28.11 and 4.20 (page 109). As these figures show, a tRNA molecule looks rather like a hand-held drill or soldering gun. The anticodon loop is at the bottom of the grip, and the acceptor stem is at the working tip. The D loop and the TψC loop are folded inward in a complex fashion near the top of the grip, to provide a maximum of hydrogen bonding and base stacking. Some of the hydrogen-bonding patterns required to produce this folding are rather unusual (Figure 28.12). The three-dimensional shapes of the tRNAs are highly conserved even though the primary structures vary. A likely explanation is that such conservation is necessary so that each tRNA can fit equally well onto the ribosome and carry out its function.

FIGURE 28.12

Unusual base pairings in tRNA. All are from the yeast $tRNA^{Phe}$ shown in Figure 28.11. **(a, b)** Some unusual pair matches. **(c, d)** Some examples of triple interactions. R represents the ribosyl residue of the RNA chain. The bases prefixed by m are methylated at the carbon atom corresponding to the superscript. Numbers following the letters designating bases show the position in the sequence.



Coupling of tRNAs to Amino Acids and Formation of Aminoacylated tRNAs: The First Step in Protein Synthesis

Amino acids are attached to tRNAs by a covalent bond between the carboxylate of the amino acid and a ribose 3′ hydroxyl group of the invariant 3′ terminal adenosine residue on the tRNA. Pairing of the correct amino acid residues and the tRNAs is accomplished by a set of enzymes called **aminoacyl-tRNA synthetases** (abbreviated AARS). In *E. coli* there are 21 synthetases, each of which recognizes one amino acid and one or more tRNAs. Lysine is unique in having two synthetases. The reaction linking the two molecules, shown in Figure 28.13, proceeds in two steps. First, the amino acid, which is bound to the synthetase, is activated by ATP to form an **aminoacyl adenylate**. While still bound to the enzyme, this intermediate reacts with one of the correct tRNAs to form the covalent bond and release AMP.

Because all of the synthetases perform essentially the same function, one might expect them to represent minor variations on a common theme. This is, however, not the case; there are two general classes of aminoacyl-tRNA synthetases (I and II). Their active sites are completely different, and the two classes bind their cognate tRNAs from opposite sides. Furthermore, the class I enzymes tend to function as monomers, whereas the class II enzymes function as dimers or tetramers. Moreover, the enzymes differ mechanistically. Class II enzymes link the aminoacyl moiety in the aminoacyl adenylate intermediate directly to the 3′ hydroxyl in the tRNA acceptor, while class I enzymes synthesize first a 2′-aminoacyl-tRNA intermediate, which then undergoes intramolecular transesterification, giving the 3′-aminoacyl-tRNA product.

The reasons for these extreme differences are unknown, but they may reflect the utilization of some amino acids in proteins before others in the very early evolution of protein synthesis. A recent observation that may bear on this question is that some members of some classes of organisms (Gram-positive bacteria and archaea, for example), as well as some organelles, use an indirect transamidation route for charging some tRNAs. For example, $tRNA^{Gln}$ is charged first with Glu, which is then replaced by Gln.

$$\text{Glu} + \text{tRNA}^{\text{Gln}} + \text{ATP} \xrightleftharpoons{\text{glutamyl-tRNA synthetase}} \text{Glu-tRNA}^{\text{Gln}} + \text{AMP} + \text{PP}_i$$

$$\text{Gln} + \text{Glu-tRNA}^{\text{Gln}} + \text{ATP} \xrightleftharpoons{\text{Glu-tRNA}^{\text{Glu}}\ \text{amidotransferase}} \text{Gln-tRNA}^{\text{Gln}} + \text{ADP} + \text{P}_i + \text{Glu}$$

Thus, these organisms do not require (although they may have) a $Gln\text{-}tRNA^{Gln}$ synthetase, as do Gram-negative bacteria and eukaryotes. It has been suggested that glutamine was one of the last amino acids to be added to the protein repertoire and that it was initially incorporated by this route.

In higher organisms nine aminoacyl-tRNA synthetases are organized into a high-molecular-weight complex, along with three accessory proteins. The biological function of this complex is unknown, but it is assumed to contribute toward coordinating amino acid synthesis with protein synthesis.

You might expect that the synthetase would identify the correct tRNA on the basis of its anticodon, but many studies indicate that the identification process is more complex, and various nucleotides act as *identity elements.* In 1988, Ya-Ming Hou and Paul Schimmel showed that changing a single base pair (between residues 3 and 70 in the acceptor stem) of $tRNA^{Cys}$ or $tRNA^{Phe}$ to the G-U pair found in $tRNA^{Ala}$ caused the alanine synthetase to accept the $tRNA^{Cys}$ or $tRNA^{Phe}$ and couple it to alanine. Other tRNAs appear to be recognized by their synthetases at many different locations (see Figure 28.14). No simple rule has emerged, although it is

Amino acids are coupled to their appropriate tRNAs by aminoacyl-tRNA synthetases.

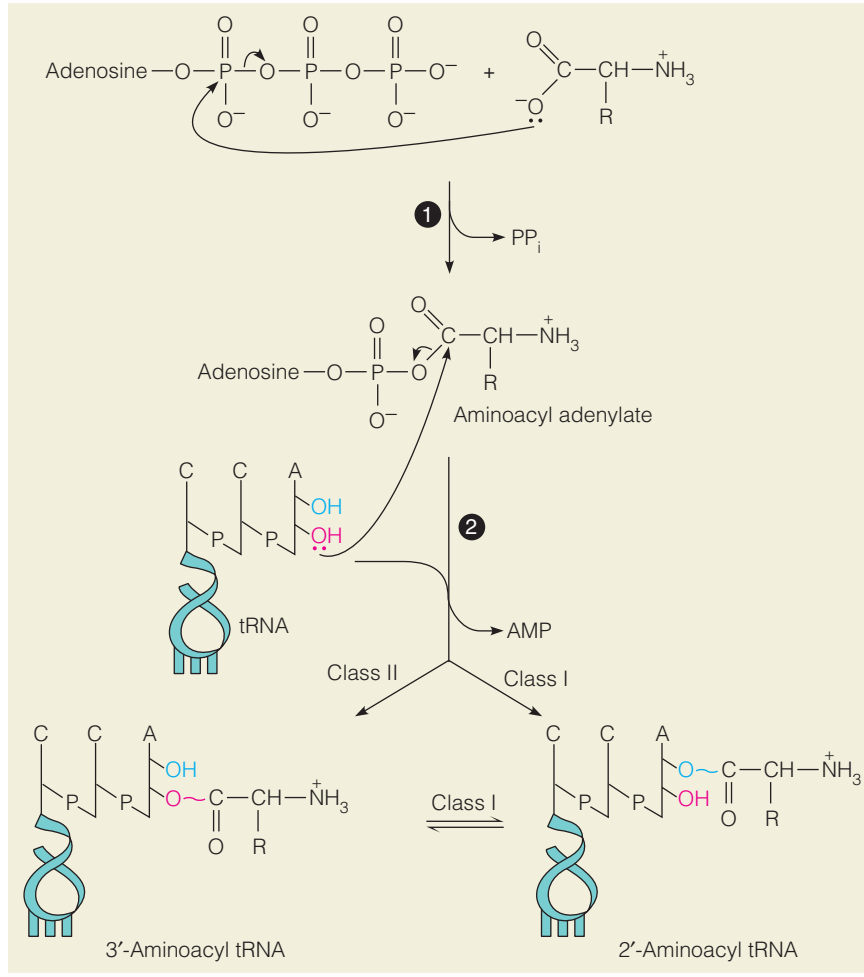


FIGURE 28.13

Formation of aminoacyl tRNAs by aminoacyl tRNA synthetase. In step 1 the amino acid is accepted by the synthetase and is adenylylated, with the aminoacyl adenylate remaining bound to the enzyme. In step 2 the proper tRNA is accepted by the synthetase, and the amino acid residue is transferred to the 3′ OH of the 3′-terminal residue of the tRNA (class II enzymes) or to the 2′ hydroxyl, followed by isomerization to the 3′ aminoacyl-tRNA (class I enzymes). For class I enzymes the 2′ hydroxyl of the 3′-terminal AMP residue is the nucleophile for reaction 2.

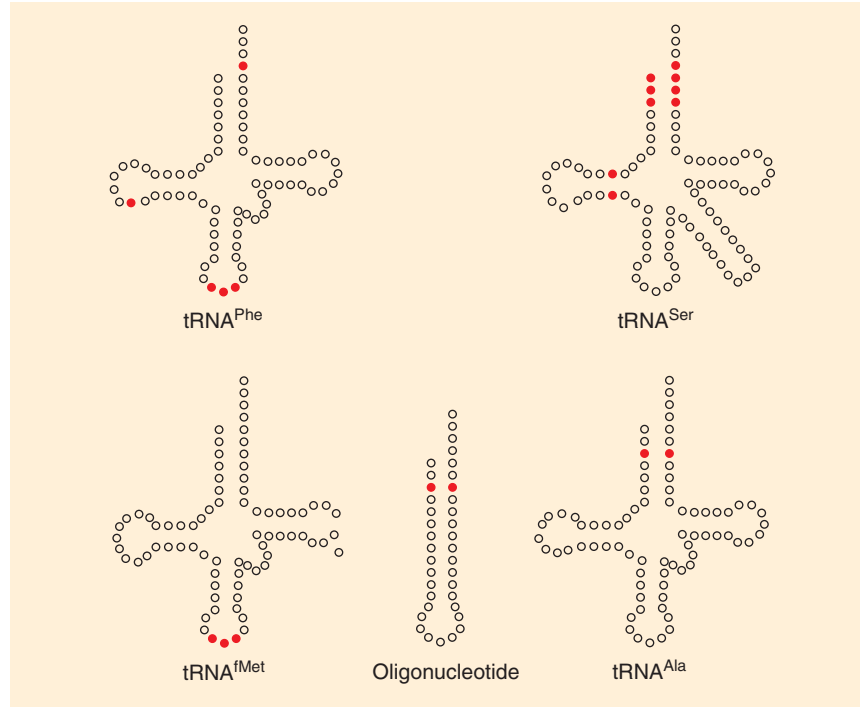


FIGURE 28.14

Major "identity elements" in some tRNAs. Red circles represent the positions that have been shown to identify the tRNA to its cognate synthetase. Shown also is a synthetic polynucleotide containing the G-U alanine identity element (in red), which is a good substrate for alanyl-tRNA synthetase.

From *Science* 240:1591–1592, L. Schulman and J. Abelson, Recent excitement in understanding transfer RNA identity. © 1988. Reprinted with permission from AAAS. Adapted with permission from John Abelson.

clear that identity elements are clustered in the anticodon loop and the acceptor stem. A dramatic illustration of the importance of identity elements is shown by the fact that the yeast $tRNA^{Ala}$ shown in the figure can be trimmed to just a single-hairpin molecule, as shown, and, so long as a critical G-U base pair (shown in red) is present, the molecule can be efficiently and accurately aminoacylated.

Aminoacyl-tRNA synthetases contribute toward the fidelity of translation by a process akin to proofreading by DNA polymerases. In the instant between formation of an enzyme-bound aminoacyl adenylate and its conversion to aminoacyl-tRNA, the enzyme can sense the improper fit of the amino acid side chain and hydrolyze the intermediate before the amino acid can be linked to tRNA. Moreover, even if the wrong aminoacyl-tRNA is synthesized, the enzyme has a short time in which it can identify the mischarged amino acid as incorrect and hydrolyze it before it can be released to participate in translation. In these ways, aminoacyl-tRNAs contribute toward an overall error frequency for protein synthesis of about 10^{-4}, less accurate than DNA replication, to be sure, but with the consequences of error being much lower because the error is not propagated to the next generations. Most of the work on AARS proofreading has been done with isoleucyl-tRNA synthetase and its ability to mischarge valine, which differs from isoleucine by only one methylene group. Despite this small structural difference, the mischarging frequency is only 3×10^{-4}.

Insight into the recognition of tRNAs by their synthetases has been provided by crystallographic analysis of the complexes formed. Figure 28.15 shows the structure of a class I synthetase-tRNA complex, *E. coli* glutaminyl-tRNA. As shown in the figure, the tRNA lies across the protein, making a number of specific contacts, including crucial ones in the anticodon region and in the acceptor stem. Both of these regions are distorted in the complex, with the acceptor stem being elongated and inserted into the active site pocket. This pocket is formed by a common protein structural motif called the *dinucleotide fold,* which frequently acts as a nucleotide-binding region. In this case it also binds the ATP required for acylation. It provides a binding site for glutamine as well. Thus, all three participants in the reactions are grouped close together.

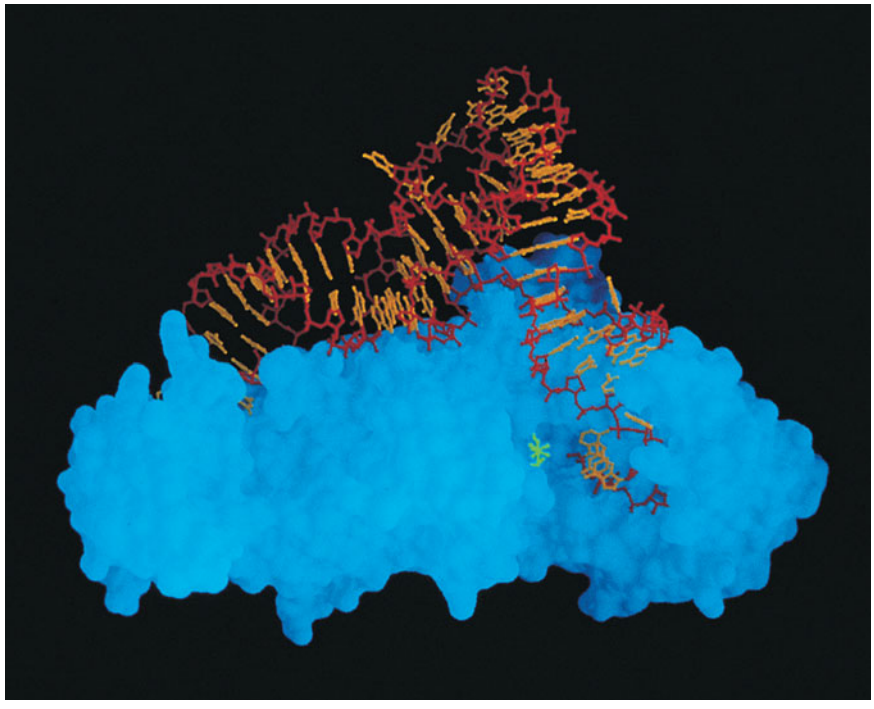


FIGURE 28.15

A model of the *E. coli* glutaminyl tRNA synthetase coupled with its tRNA and ATP. The tRNA is represented by a detailed atomic model, the protein by its solvent-accessible surface (blue). The ATP (green) and the 3′ acceptor stem of the tRNA fit into a deep cleft in the synthetase. This cleft will also accommodate the amino acid. This is a monomeric class I synthetase. PDB ID 1GSG.

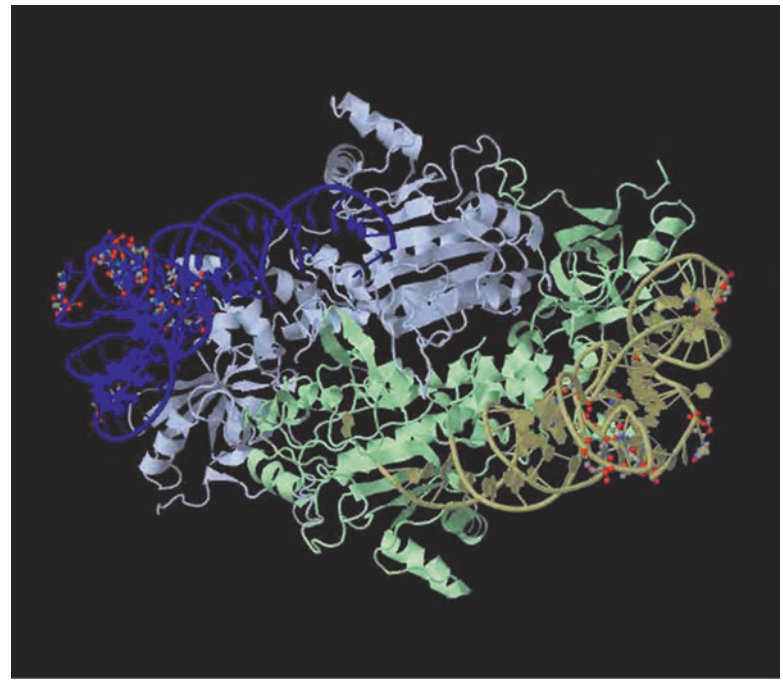


FIGURE 28.16

Yeast aspartyl tRNA synthetase complexed with two molecules of $tRNA^{Asp}$. This is a dimeric class II synthetase. Protein subunits are in white and pale green. tRNA molecules are in blue and gold. PDB ID 1ASY.

Similar interactions are seen with a class II synthetase. Figure 28.16 shows the dimeric yeast aspartyl tRNA synthetase complexed with two molecules of $tRNA^{asp}$. Note that the tRNA is bound to the enzyme in opposite orientation from that seen with class I synthetases. Only one of the two tRNA molecules is bound in a catalytically productive conformation.

One other feature of aminoacyl-tRNA synthetases deserves comment. In higher eukaryotes most of these enzymes are "moonlighting proteins"—proteins that evolved initially to play their well-known function in protein synthesis, but which in further evolution acquired additional functions. In humans, aminoacyl-tRNA synthetases are involved in functions as diverse as autoimmunity, control of apoptosis, regulation of ribosomal RNA synthesis, vascular development, and coordination of the DNA damage response. In all cases studied the catalytic machinery for aminoacyl-tRNA synthesis has remained undisturbed, and evolutionary modifications to convey additional functions occur elsewhere on the protein molecule.

The Ribosome and Its Associated Factors

We have now described two of the participants that must be brought together to carry out protein biosynthesis—the mRNA and the set of tRNAs charged with the appropriate amino acids. The actors are in the wings, and all that is needed is a proper director and a stage on which the events can unfold. Both are provided by the ribosome, and the typical cell requires many. An *E. coli* cell, for example, contains as many as 20,000 ribosomes, accounting for about 25% of the dried cell mass. Thus, a cell devotes a large part of its energy to producing ribosomes and to using them in protein synthesis.

Soluble Protein Factors in Translation

Before describing ribosomes in detail, however, we mention one more set of participants, whose functions will be described in detail later. These are the soluble proteins that participate in the three stages of translation—initiation factors, elongation factors, and release factors. Table 28.4 introduces these factors as initially studied in bacteria, as well as their eukaryotic counterparts. We shall refer back to the information in this table as we discuss mechanisms in translation.

TABLE 28.4 Soluble protein factors in translation

Function	Factor (Bacteria)	Factor (Eukaryotes)	Role in Translation
Initiation	IF1	eIF1, eIF1A	Promotes dissociation of preexisting 70S ribosome
	IF2	eIF2, eIF2B	Helps attach initiator tRNA
	IF3	eIF3, eIF4C	Similar to IF1; prepares mRNA for ribosome binding
		eIF4A, eIF4B, eIF4F	Same as eIF1, eIF1A
		eIF5	Helps dissociate eIF2, eIF3, eIF4C
		eIF6	Helps dissociate 60S subunit from inactive ribosomes
Elongation	EF-Tu	eEF1α	Helps deliver aminoacyl-tRNA to ribosomes
	EF-Ts	eEF1$\beta\gamma$	Helps recharge EF-Tu with GTP
	EF-G	eEF2	Facilitates translocation
Termination	RF1	eRF	Release factor (UAA,UAG)
	RF2		Release factor (UAA, UGA)
	RF3		A GTPase that promotes release

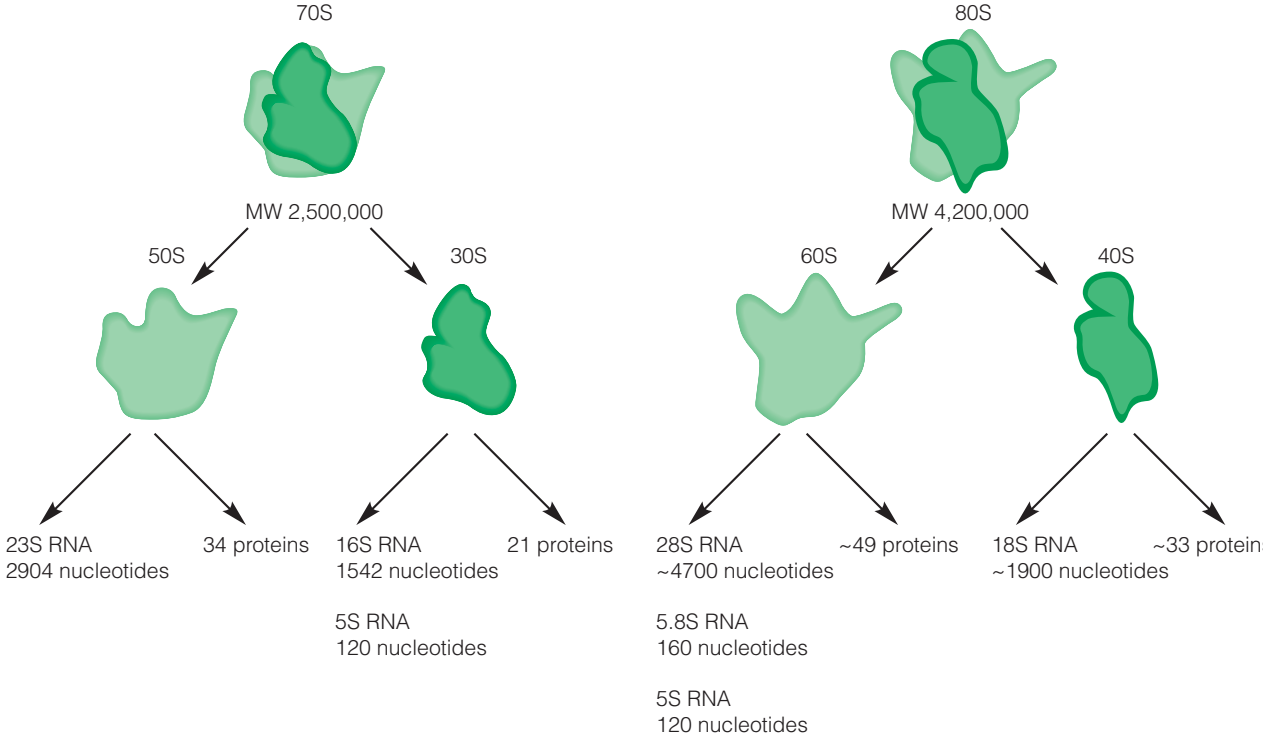
Components of Ribosomes

The ribosome is a large ribonucleoprotein particle containing 60–70% RNA and 30–40% protein. Ribosomes and their subunits are characterized in terms of their sedimentation coefficients in ultracentrifugation. Thus, the individual bacterial ribosome is called a 70S particle and has a molecular mass of about 2.5×10^6 Da. Eukaryotic ribosomes are somewhat larger, with a sedimentation coefficient of 80S and a molecular mass of 4.2×10^6 Da. When isolated ribosomes are placed in a buffer containing low Mg^{2+} ion, they dissociate into two smaller subunits. As shown in Figure 28.17, bacterial 70S ribosomes dissociate into 30S and 50S subunits. We shall see later that dissociation and reassociation of these subunits are crucially important during translation. Figure 28.17 also shows the number of RNA and protein components in each subunit. Note that the 50S bacterial subunit contains two rRNA molecules (5S and 23S) and 34 different proteins, while the

FIGURE 28.17

Components of bacterial and eukaryotic ribosomes. Bacterial (to the left) and eukaryotic (to the right) ribosomes are assembled along the same structural plan, with eukaryotic ribosomes being somewhat larger and more complex. The shapes of the ribosomal subunits were determined by electron microscopy.

Modified from *Molecular Biology of the Cell*, 4th ed., B. Alberts et al. Garland Science, New York, 2002.



30S subunit contains just one rRNA (16S) and 21 proteins—all different from those in the 50S subunit. Proteins from the small subunit are called S1, S2, S3···S21, while those from the large subunit are called L1, L2, L3···L34. All proteins are present in one copy per ribosome, except for L12, which is present in four copies. Eukaryotic ribosomes are significantly larger, with larger rRNAs and more proteins. We shall be discussing primarily bacterial ribosomes, whose structures and functions are known in much greater detail.

Once the complexity of the ribosome was revealed, particularly the large numbers of proteins in each subunit, it seemed a daunting task to determine the structure of the particle and to understand the function of each protein. However, Peter Traub and Masayasu Nomura learned as early as 1968 that they could reassemble 30S ribosomal subunits from the separated RNA and protein components. The product, when combined with 50S subunits, was active for in vitro protein synthesis. An obligatory order of assembly was seen, with some proteins being incorporated only after binding of certain other proteins. As expected if transcription and translation are coupled, the proteins bound earliest in the pathway are those linked to the 5′ end. The ability to assemble ribosomes in vitro allowed analysis of the function of individual ribosomal proteins because ribosomal subunits could be assembled with specific proteins missing, followed by functional analysis of these deliberately altered particles.

Despite their complexity, ribosomal subunits can be assembled in vitro.

More recent analysis shows that the pathway of ribosome assembly in vivo differs from the pathway in vitro pathway in significant ways. In this analysis, carried out by James Williamson and colleagues, assembly intermediates accumulated in *E. coli* treated with the antibiotic neomycin and were analyzed by ^{15}N pulse labeling and mass spectrometry. Additional information was obtained from electron micrographic analysis of individual assembly intermediates. Distinctions between the in vitro and in vivo pathways include the following. First, ribosome assembly in vivo occurs via parallel pathways, with one pathway adding proteins initially at the rRNA 5′ domain and another at the 3′ domain. Second, some proteins bound to central domains are incorporated before proteins binding to the ends of 16S rRNA. Third, the ribosomes are assembled both from newly synthesized proteins and from proteins that originated in previously synthesized intact subunits. Figure 28.18 summarizes principal features of the in vitro and in vivo pathways.

Sequence analyses reveal no significant homologies among the different proteins in a ribosome, but comparison of sequences between corresponding proteins in the ribosomes of different organisms reveals considerable evolutionary conservation. Thus, the ribosome is a complex object that evolved early in the history of life and has remained relatively unchanged. Although the ribosomes of eukaryotes differ significantly from those of prokaryotes, the evolutionary continuity is clear. The sequences of many ribosomal RNAs tell the same story. Indeed, because of their relatively slow evolutionary rates of change, rRNAs are useful as evolutionary yardsticks over vast phylogenetic distances. In fact, it was sequence analysis of 16S rRNAs that led Carl Woese to propose the existence of a third domain of life, the archaea.

Ribosomal RNA Structure

When the sequences of 16S rRNAs were originally determined, they were found to contain many regions of self-complementarity, which are capable of forming double-helical segments. A pattern like that shown in Figure 28.19 may seem so complex as to appear almost arbitrary, but comparison with other, even distantly related, 16S RNA sequences shows that the potentially double-stranded regions are highly conserved. Indeed, the secondary structure seems more highly conserved than is the primary structure, for it is often found that there are compensatory mutations in double-helical regions so as to maintain base pairing. A schematic illustration like that in Figure 28.19 is analogous to the cloverleaf

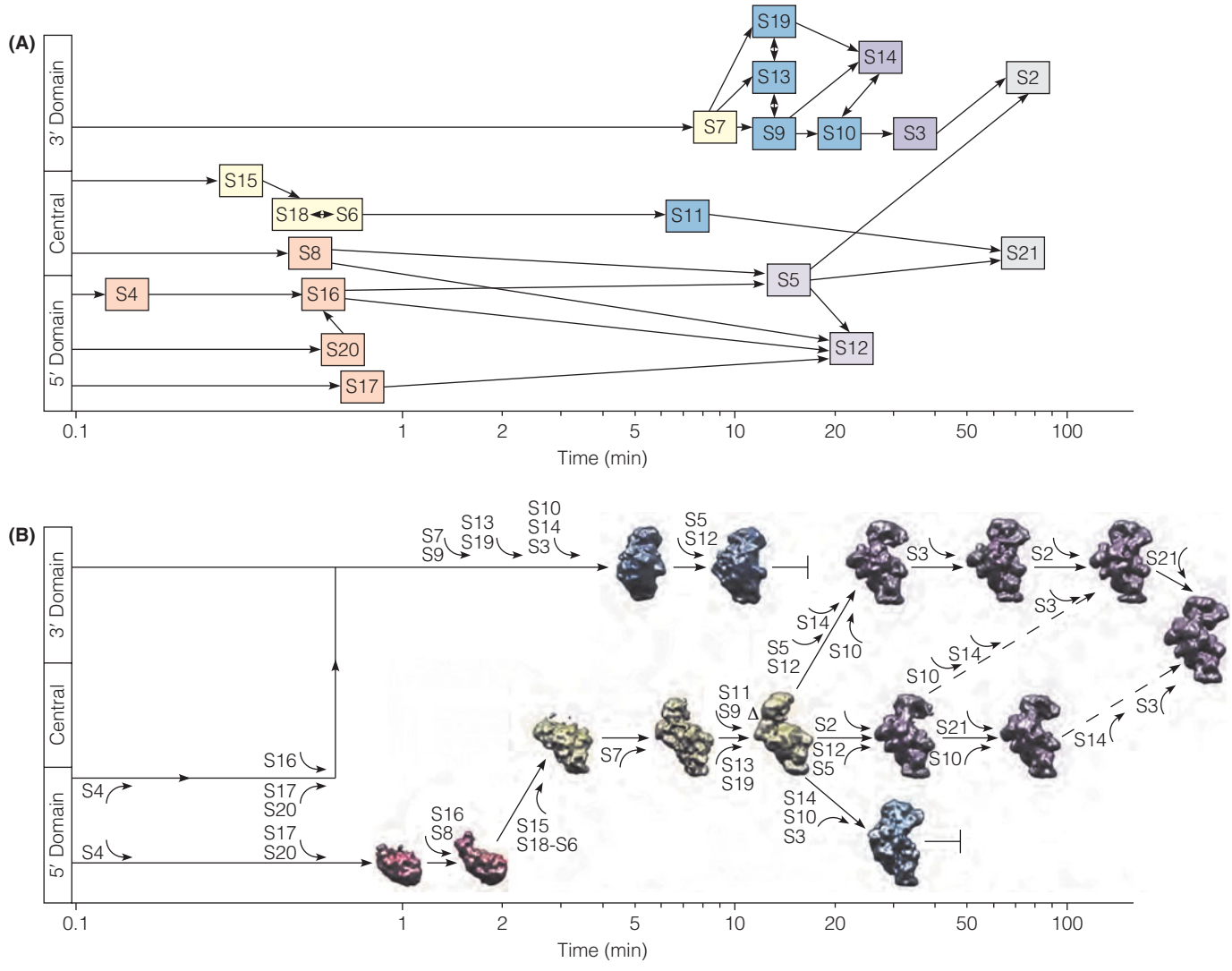


FIGURE 28.18

Assembly map for the 30S subunit. (a) Assembly pathway in vitro, as determined by Traub and Nomura. Arrows indicate the obligatory nature of some protein-binding events. For example, S7 must bind before S9, S13, or S19, but once S7 is bound, any of these three proteins can be added. The earliest protein-binding events occur near the 5′ end of the 16S rRNA, and an intermediate to which 5′- and central-domain proteins are bound must be formed before addition of 3′ domain proteins. **(b)** In vivo assembly map, as determined by Williamson and colleagues. Parallel pathways begin, with proteins added at either the 5′ domain or the 3′ domain of 16S rRNA.

visualization of a tRNA (Figure 28.9). The actual rRNA is folded into a three-dimensional structure, just as is the tRNA. In the case of the ribosomal subunit, however, the structure is further complicated by the presence of ribosomal proteins bound to the RNA. However, it is now clear that the pattern shown in Figure 28.19 faithfully describes the secondary structure of 16S rRNA. 23S rRNA has a comparable secondary structure, actually more complex, reflecting its larger size.

Internal Structure of the Ribosome

Although EM images of intact ribosomes and their subunits were obtained some time ago, high resolution was difficult to achieve because of the necessity of staining or shadowing the particles. Nor could such techniques hope to tell us how the proteins and RNA were positioned inside the ribosome. Nevertheless, the overall

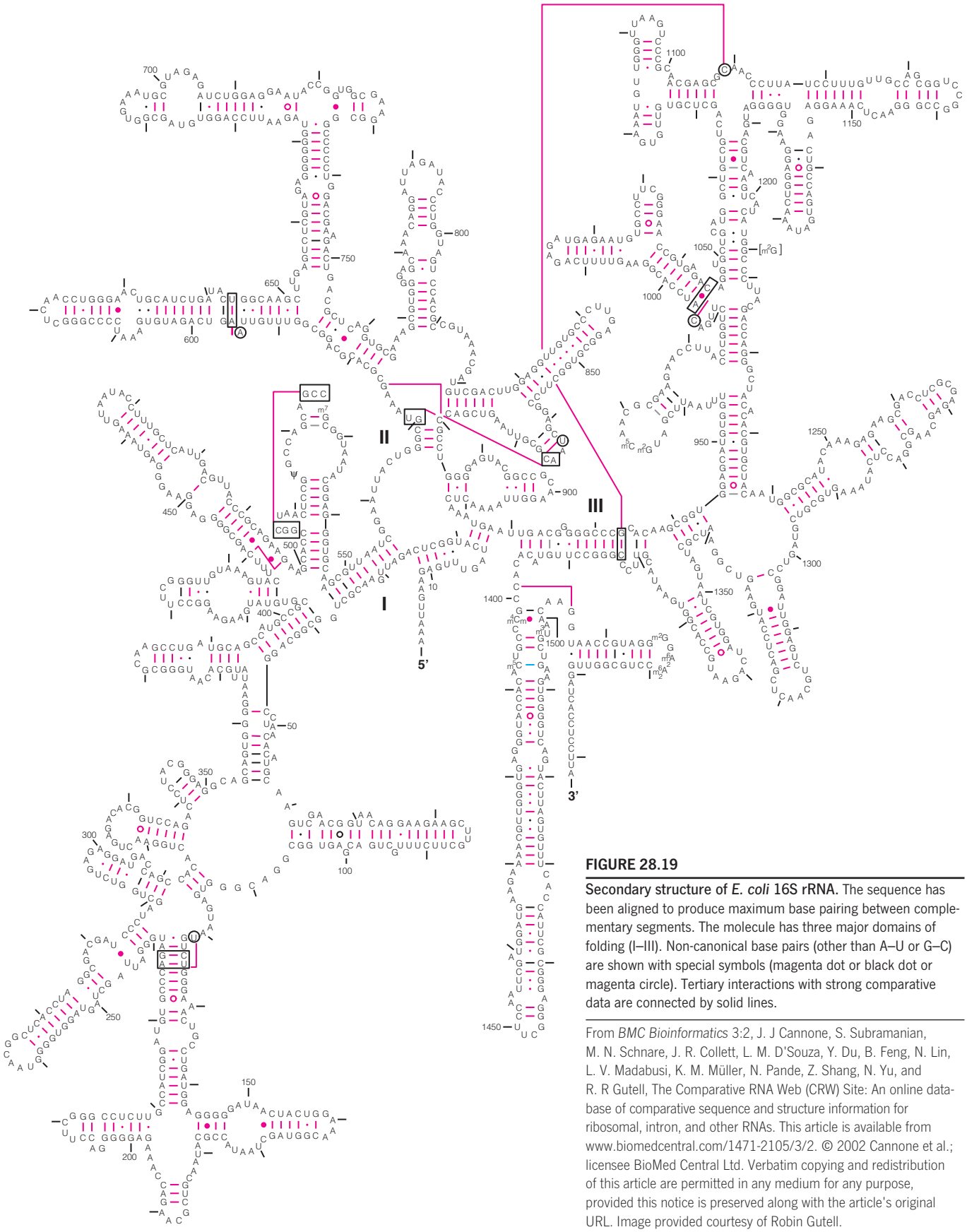


FIGURE 28.19

Secondary structure of *E. coli* 16S rRNA. The sequence has been aligned to produce maximum base pairing between complementary segments. The molecule has three major domains of folding (I–III). Non-canonical base pairs (other than A–U or G–C) are shown with special symbols (magenta dot or black dot or magenta circle). Tertiary interactions with strong comparative data are connected by solid lines.

shape of each subunit was seen as early as 1976 (Figure 28.20). Other techniques such as protein–protein or protein–RNA cross-linking, immunoelectron microscopy, cryoelectron microscopy, sequence analysis of the individual proteins, and neutron scattering gave much information about the placement of individual proteins within the overall structure.

Attempts to crystallize ribosomes began in the 1970s, but the size and complexity of these particles frustrated these early efforts. A key to the success of crystallization attempts, particularly in the laboratory of Ada Yonath, was the use of extremophilic bacteria as the source material. Although the *E. coli* ribosome was the most thoroughly studied to that point, the archaeal thermophile *Thermus thermophilus* and the halophile *Haloarcula marismortui* yielded the best crystals. Because the structure of the ribosome is well conserved evolutionarily, these bacteria made satisfactory models.

Figure 28.21 shows a model of the 70S ribosome based upon the first medium-resolution structures, in the late 1990s. Critical features known already or shown by this model are that the ribosome has three tRNA binding sites, that mRNA binding and decoding occur on the 30S subunit, that aminoacyl-tRNAs fill the gap between 30S and 50S subunits, that the newly synthesized polypeptide chain exits the ribosome through a tunnel in the 50S subunit, and that the peptidyltransferase reaction, which creates peptide bonds, occurs at a site on the 50S subunit.

Within a year after the medium-resolution structure was published, Thomas Steitz's laboratory published a high-resolution structure of the 50S subunit from *H. marismortui*, and shortly afterward Venki Ramakrishnan and colleagues described the 30S subunit from *T. thermophilus*, followed by the complete 70S ribosome from this organism. Figure 28.22 shows the Steitz 50S structure, and Figure 28.23 shows the Ramakrishnan 70S structure. Perhaps the most striking feature of both structures is that the peptidyltransferase site lies far from any protein. This structural work established conclusively that the ribosome is a *ribozyme*. We shall

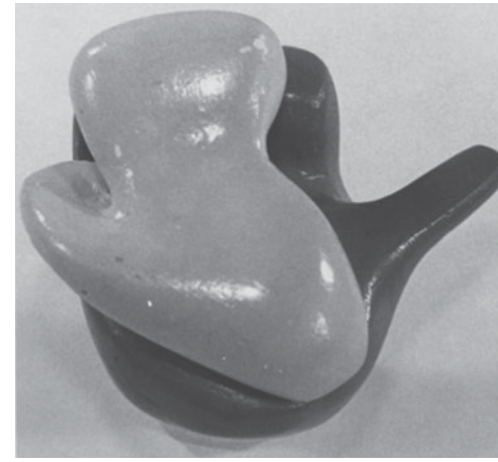


FIGURE 28.20

Images of ribosomal subunits as determined by electron microscopy. The 50S subunit is shown in black and the 30S subunit in light gray.

Reprinted from *Journal of Molecular Biology* 105: 131–159, J. A. Lake, Ribosome structure determined by electron microscopy of *Escherichia coli* small subunits, large subunits and monomeric ribosomes. © 1976, with permission from Elsevier.

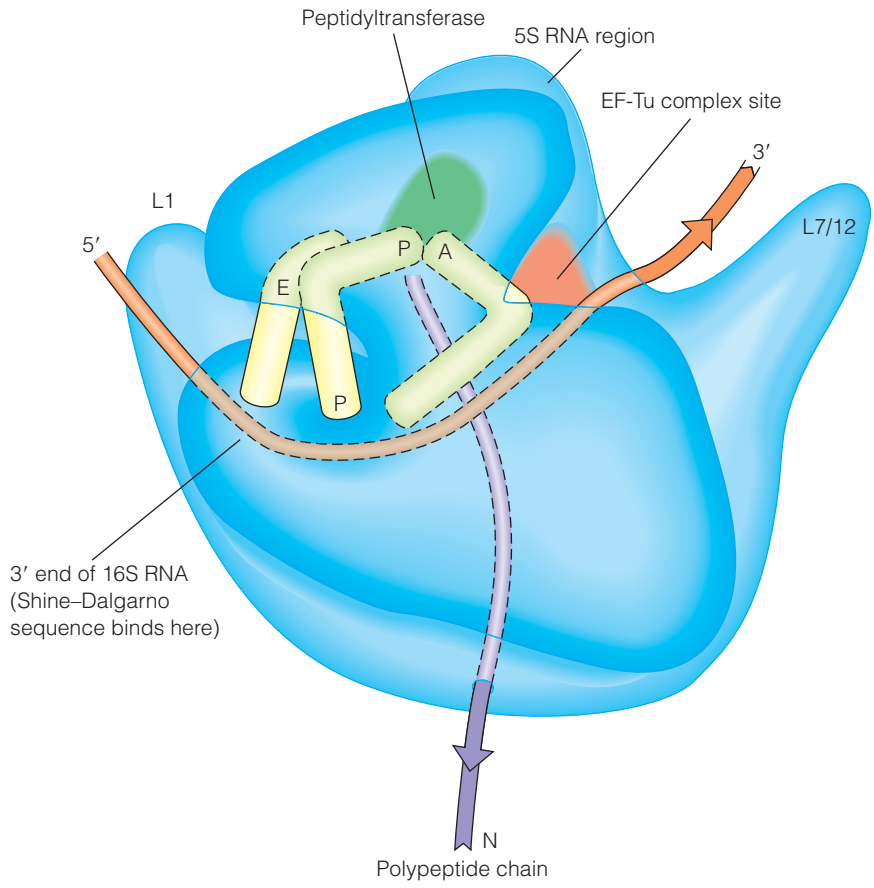


FIGURE 28.21

A model of the 70S ribosome based upon early structural data. This model shows all three tRNA-binding sites occupied simultaneously, which does not normally occur. This view has the 30S subunit in front and the 50S subunit to the rear.

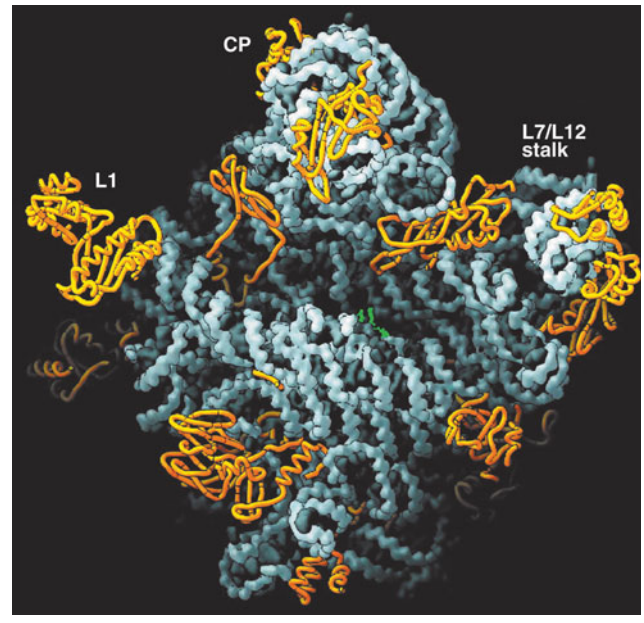


FIGURE 28.22

A high-resolution model of the 50S ribosomal subunit. This view shows the two stalks and central protuberance (CP), seen also in the early electron micrographs. In this image RNA is in gray, and proteins are in gold. The peptidyltransferase site, in green, is identified from the binding of an inhibitor. PDB ID 1FFK.

From *Science* 289:905–920, N. Ban, P. Nissen, J. Hansen, P. B. Moore, and T. A. Steitz, The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. © 2000. Reprinted with permission from AAAS.

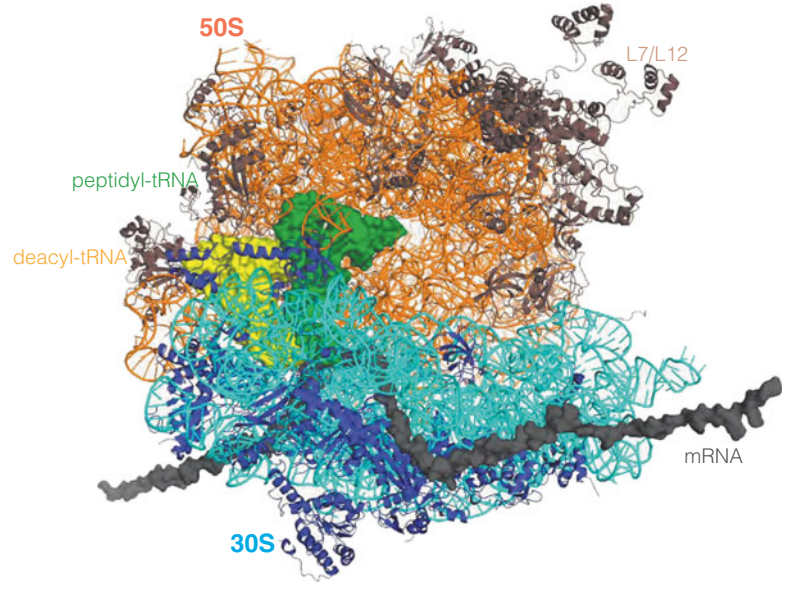


FIGURE 28.23

A model of the 70 ribosome, with mRNA and tRNA bound. The 30S subunit is in light blue-green (RNA) and blue (protein), and the 50S subunit is in orange (RNA) and brown (protein). Two bound tRNAs can be seen; peptidyl-tRNA in green and deacylated tRNA in yellow. mRNA is shown in gray. PDB ID 2j00 (30S-1), 2j01 (50S-1), 2j02 (30S-2), and 2j03 (50S-2).

Courtesy of V. Ramakrishnan (2009) Nobel Prize lecture. © The Nobel Foundation.

return to this reaction later and to the question of how this structural work illuminated ribosome function. Yonath, Steitz, and Ramakrishnan shared the 2009 Nobel Prize in Chemistry for their contributions to ribosome structure and function.

In late 2010 Adam Ben-Shem and colleagues reported a structure for the yeast 80S ribosome at 4.15 Å resolution, and in early 2011 the large subunit of the *Tetrahymena thermophila* ribosome was reported at somewhat higher resolution. The yeast structure, which is shown on the cover of this book, was in a "racheted" state, believed to represent an intermediate in translocation, the movement from one codon to the next. A major distinction from bacterial ribosomes is the greater extent of interaction of ribosomal proteins with each other, rather than with ribosomal RNAs.

Mechanism of Translation

We now have identified all of the major participants in the translation process: a messenger RNA, charged tRNAs, soluble protein factors, and the ribosome, where the actual translation events occur. As with transcription, we can divide translation into three stages: *initiation, elongation,* and *termination.* Here we describe

these steps, primarily as they occur in prokaryotes, where our understanding is most detailed. Significant, though not fundamental, differences in eukaryotic protein synthesis are discussed next.

Each step in translation requires a number of specific proteins that interact with the major participants listed above. These proteins are referred to as *initiation factors* (IFs), *elongation factors* (EFs), and *release factors* (RFs). These factors, together with some of their properties and functions, are listed in Table 28.4.

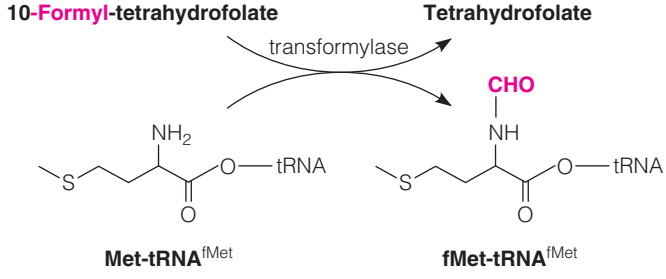
Translation involves three steps—initiation, elongation, and termination—each aided by soluble protein factors.

Initiation

Initiation of translation is schematized in Figure 28.24. Initiation results in formation of a *70S initiation complex,* which consists of a ribosome bound to mRNA and to a charged *initiator tRNA.* In bacteria the initiator tRNA is charged with *N*-formylmethionine (fMet). First the mRNA and tRNA bind to a free 30S subunit, then the 50S subunit is added to form the entire complex. In most cases the initiation codon to which the initiator tRNA binds is AUG, also used as an internal methionine codon. As we indicated previously (page 1184), the initiator AUG is distinguished from internal methionine codons by the presence upstream of a Shine–Dalgarno sequence, which binds to a complementary sequence in 16S rRNA, thereby positioning the initiator AUG.

Binding of mRNA and initiator tRNA also requires binding of the three **initiation factors** (IF1, IF2, and IF3) to a free 30S subunit. The factors IF3 and IF1 promote dissociation of preexisting 70S ribosomes, thereby producing the free 30S subunits needed for initiation (Figure 28.24, step 1). The third factor, IF2, is bound carrying a molecule of GTP; it delivers the charged initiator tRNA in binding to the 30S subunit. IF2 is a G protein, similar to those involved in signal transduction. At about the same time that the IF2–fMet-tRNA^{fMet} complex is bound, the mRNA is bound (step 2). Although the order of these additions is still uncertain, it is clear that IF2–GTP is absolutely required for binding of the first (initiator) tRNA. With binding of the initiator tRNA and the mRNA, formation of the *30S initiation complex* is complete. The initiation complex has high affinity for a 50S subunit and binds one from the available pool (step 3), with concomitant release of IF3.

The initiator tRNA is special. It recognizes and binds to the AUG codon that would normally code for methionine, but it actually carries an *N*-formylmethionine. The formyl group is added *after* charging of the tRNA, by an enzyme *(transformylase)* that recognizes the particular tRNA^{fMet} and transfers a formyl group from 10-formyltetrahydrofolate (see Figure 20.17, page 852). *Only* tRNA^{fMet} is accepted to form the 30S initiation complex; all subsequent charged tRNAs require the fully assembled 70S ribosome. Therefore, most (not all) prokaryotic proteins are synthesized with the same N-terminal residue, *N*-formylmethionine. In almost all cases, the formyl group is removed during chain elongation. For many proteins the methionine itself is also cleaved off later.

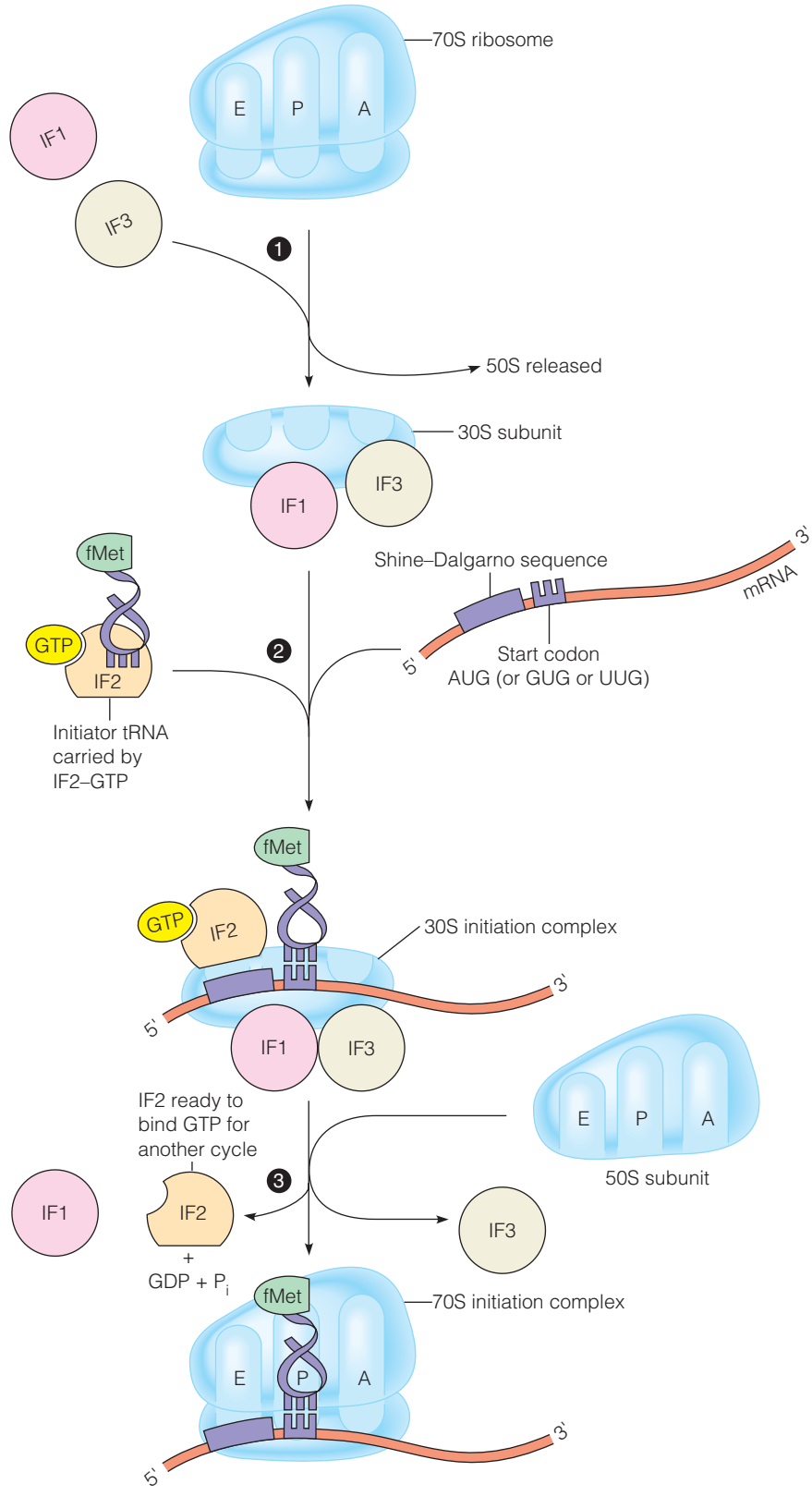


The mRNA attaches to the 30S subunit near the 5′ end of the message, which is appropriate because all messages are translated in the 5′ → 3′ direction. As mentioned above, an AUG initiation codon is recognized by an upstream Shine–Dalgarno sequence, which is complementary to the sequence, 3′···UCCUCC···5′ in 16S rRNA. This will pair with any Shine–Dalgarno sequence (for example, those shown in Table 28.3). This pairing aligns the message correctly for the start of

In initiation, the correct attachment of mRNA to the ribosome is determined by binding of the Shine–Dalgarno sequence to a sequence on the 16S rRNA of the ribosome.

FIGURE 28.24

Initiation of protein biosynthesis in prokaryotes. The ribosome contains three tRNA binding sites, shown here as E, P, and A; these are called the exit, peptidyl-tRNA, and aminoacyl-tRNA binding sites, respectively. The initiator AUG codon is positioned so that fMet-tRNA binds in the P site.



translation. In particular, it places the initiator codon next to the P site, one of three tRNA binding sites in the ribosome (see below).

Translation cannot start until the 50S subunit has bound to the 30S initiation complex. The ribosome has three sites for tRNA binding, called the P (peptidyl) site, the A (aminoacyl) site, and the E (exit) site. The AUG initiator codon, with its bound fMet-tRNA^{fMet}, aligns with the P site. At this point, the GTP molecule carried by IF2 is hydrolyzed, and IF2–GDP, P_i, and IF1 are all released. The 70S initiation complex so formed is ready to accept a second charged tRNA and begin elongation of the protein chain.

The locations of the P, A, and E binding sites for tRNAs were originally established by chemical cross-linking (Figure 28.25), but are now confirmed by X-ray crystallography (Figure 28.23). The anticodon ends of the tRNA molecules contact the 30S subunit, whereas the acceptor ends interact specifically with the 50S subunit. All of the ribosomal proteins contacted lie in the cavity between the 30S and 50S subunits. The tRNA molecules are oriented with their anticodons reaching the mRNA at the bottom of the cavity close to the 30S subunit and their acceptor ends contacting the peptidyltransferase region on the 50S subunit, near the top of the cavity.

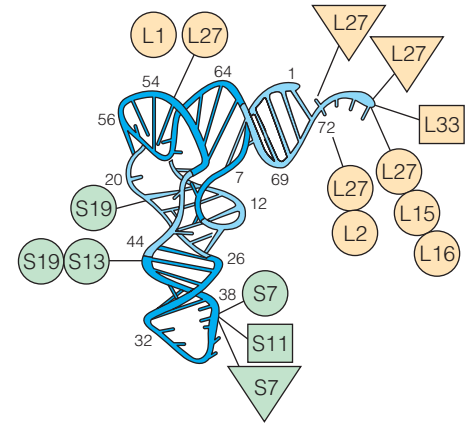


FIGURE 28.25

Environment of tRNAs at the ribosome as determined by cross-linking. Cross-links from defined nucleotide positions in the tRNA to ribosomal proteins are shown. Proteins were differentially cross-linked depending on the location of the tRNA (A site, triangles; P site, circles; E site, squares; S = small subunit, L = large subunit).

Biochimie 76:1235–1246, J. Wower, K. V. Rosen, S. S. Hixson, and R. A. Zimmermann, Recombinant photoreactive tRNA molecules as probes for cross-linking studies. Copyright © 1994 Société française de biochimie et biologie moléculaire/Elsevier Masson SAS. All rights reserved.

Elongation

Growth of the polypeptide chain on the ribosome occurs by a cyclic process. Figure 28.26 illustrates a single round in this cycle. In this particular example, the fifth amino acid from the N-terminus is being linked to the sixth. However, all cycles are the same until a termination signal is reached.

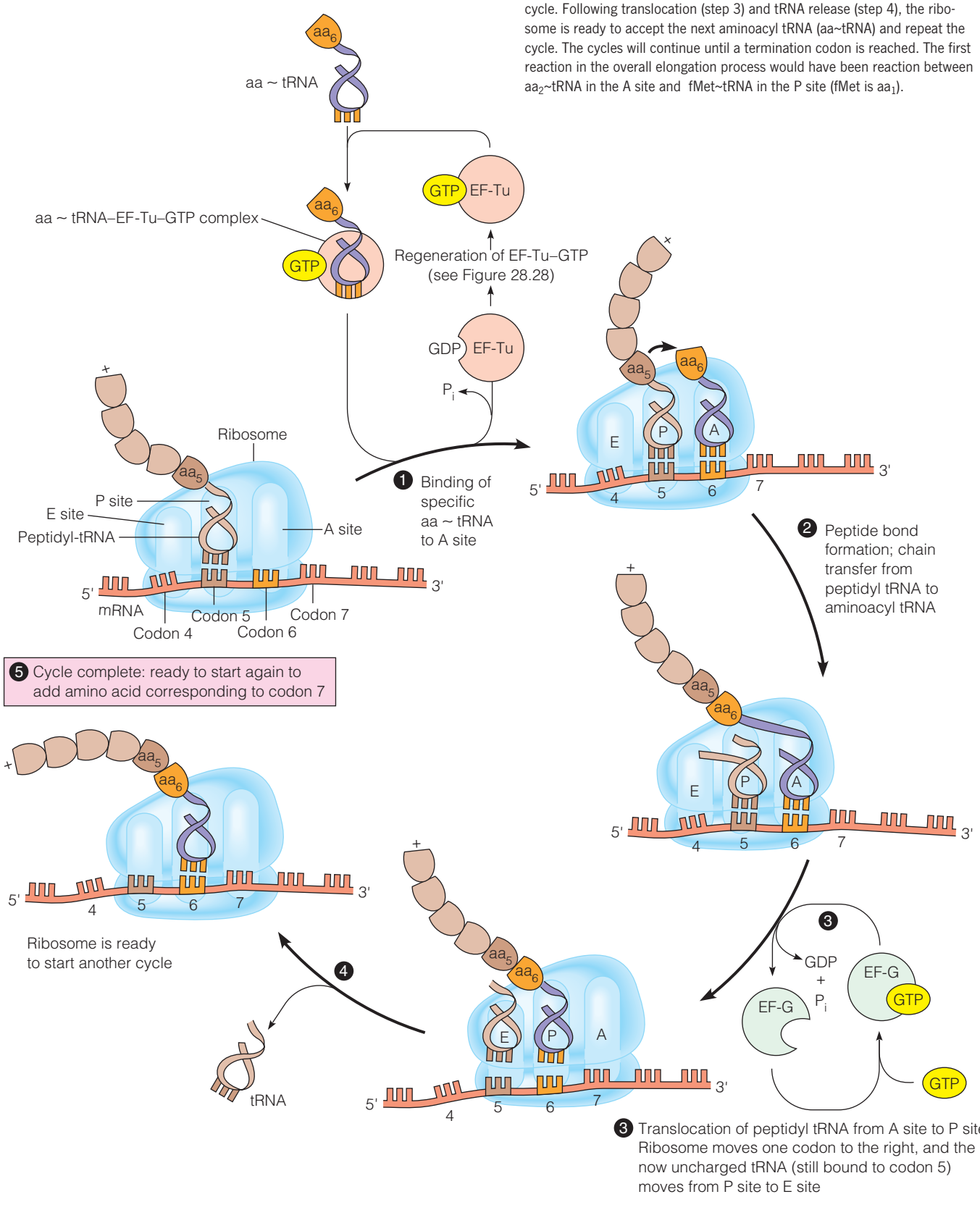
At the beginning of each cycle, the nascent polypeptide chain is attached to a tRNA in the P (peptidyl) site, and the A (aminoacyl) and E (exit) sites are empty. Aligned with the A site is the mRNA codon corresponding to the *next* amino acid to be incorporated. The charged (aminoacylated) tRNA is escorted to the A site in a complex with a protein, the elongation factor EF-Tu, which also carries a molecule of GTP. (Note the parallel to IF2–GTP here.) EF-Tu plays an active role in ensuring that the correct aminoacyl-tRNA is fitted to its codon. As schematized in Figure 28.27, which results from work in the Ramakrishnan laboratory, the aminoacyl-tRNA is distorted in its complex with EF-Tu. Initial binding puts the tRNA anticodon loop into the decoding center on the 30S subunit, with the acceptor stem near the EF-Tu site. Nucleotides in the decoding site probe the major groove of the anticodon loop, specifically in positions 1 and 2. This structural work confirmed the wobble hypothesis, by showing that codon–anticodon fitting is more stringent in the first two positions. GTP hydrolysis by EF-Tu results in conformational changes that move the aminoacyl-tRNA entirely into the A site and cause dissociation of EF-Tu itself. The EF-Tu–GTP complex is then regenerated by the subsidiary cycle shown in Figure 28.28. After the charged tRNA is in place, it is checked both before and after the GTP hydrolysis and rejected if incorrect.

The next, and crucial, step is peptide bond formation (Figure 28.26, step 2). The polypeptide chain that was attached to the tRNA in the P site is now transferred to the amino group of the amino acid carried by the A-site tRNA. This step is catalyzed by *peptidyltransferase,* an integral part of the 50S subunit. As mentioned previously, the structure determination of the 50S subunit established conclusively that catalysis is carried out by the RNA portion of the subunit. This finding was crucial to acceptance of the "RNA world" model for the origin of life, for it indicates how progenitors to living cells could exist in the presence of RNA but the absence of protein.

Analysis of the 50S subunit in the Steitz laboratory showed that a conserved AMP residue (2486 in *H. marismortui* and 2451 in *E. coli*) exists in an environment that makes the purine ring unusually basic, probably resulting from hydrogen bonding to a nearby GMP. This suggests a process in which N3 abstracts a proton from the amino group of the aminoacyl-tRNA, converting the amino group to a better nucleophile that attacks the carboxyl carbon of the C-terminal amino acid, linked to peptidyl-tRNA, as shown in Figure 28.29. The protonated

FIGURE 28.26

Chain elongation in prokaryotic translation. The process is depicted as a cycle. Following translocation (step 3) and tRNA release (step 4), the ribosome is ready to accept the next aminoacyl tRNA (aa~tRNA) and repeat the cycle. The cycles will continue until a termination codon is reached. The first reaction in the overall elongation process would have been reaction between aa_2~tRNA in the A site and fMet~tRNA in the P site (fMet is aa_1).



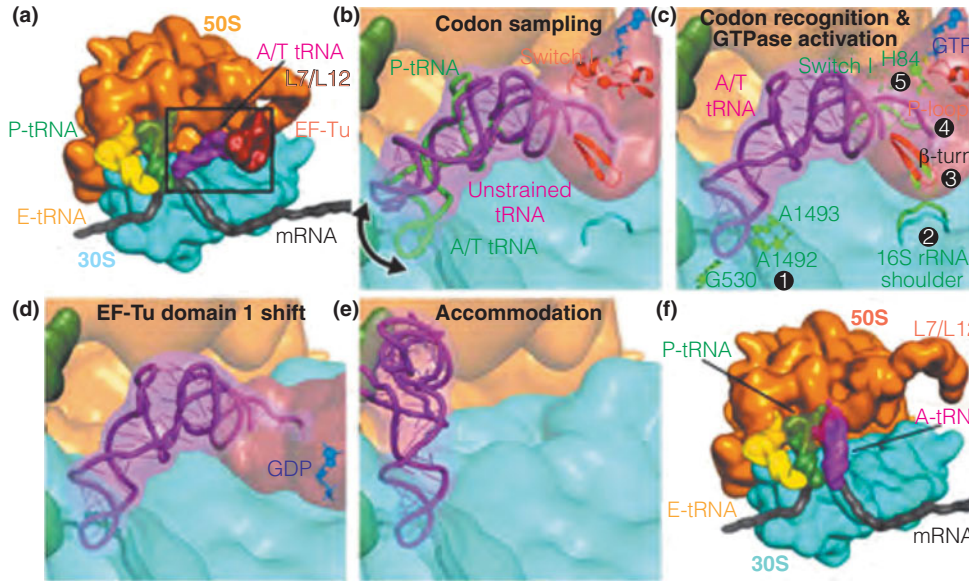


FIGURE 28.27

The ribosomal decoding pathway. (a) L7/L12 stalk on 50S subunit recruits ternary complex (aminoacyl-tRNA-EF-Tu-GTP) to a ribosome with deacylated tRNA in E site and peptidyl-tRNA in P site. The black frame is enlarged in subsequent panels. A/T tRNA is a tRNA molecule temporarily distorted, to interact simultaneously with the decoding center of the 30S subunit and EF-Tu, bound at a site in the intersubunit space. **(b)** tRNA samples codon–anticodon pairing. **(c)** The match is sensed by specific nucleotides in coding site (G530, A1492, A1493). Codon recognition triggers 30S subunit domain closure. A chain of conformational changes (shown as 2–5) opens a hydrophobic gate, allowing His84 on EF-Tu to initiate hydrolysis of GTP. **(d)** GTP hydrolysis and P_i release cause conformational change in EF-Tu, leading to its release from the ribosome. **(e)** and **(f)** EF-Tu release leads to relaxation of aminoacyl-tRNA structure and its accommodation at both the coding site and the peptidyl-transferase site. PDB ID 2WRN, 2WRO,2WRQ, and 2WRR.

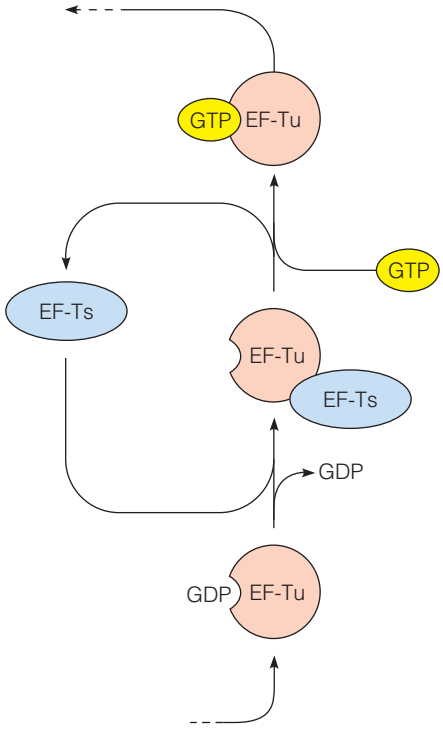


FIGURE 28.28

Regeneration of EF-Tu–GTP by Tu–Ts exchange. This figure gives details of the regeneration cycle shown at the top of Figure 28.26. Binding of the factor EF-Ts to EF-Tu allows the release of GDP and binding of a new GTP to prepare EF-Tu for another cycle.

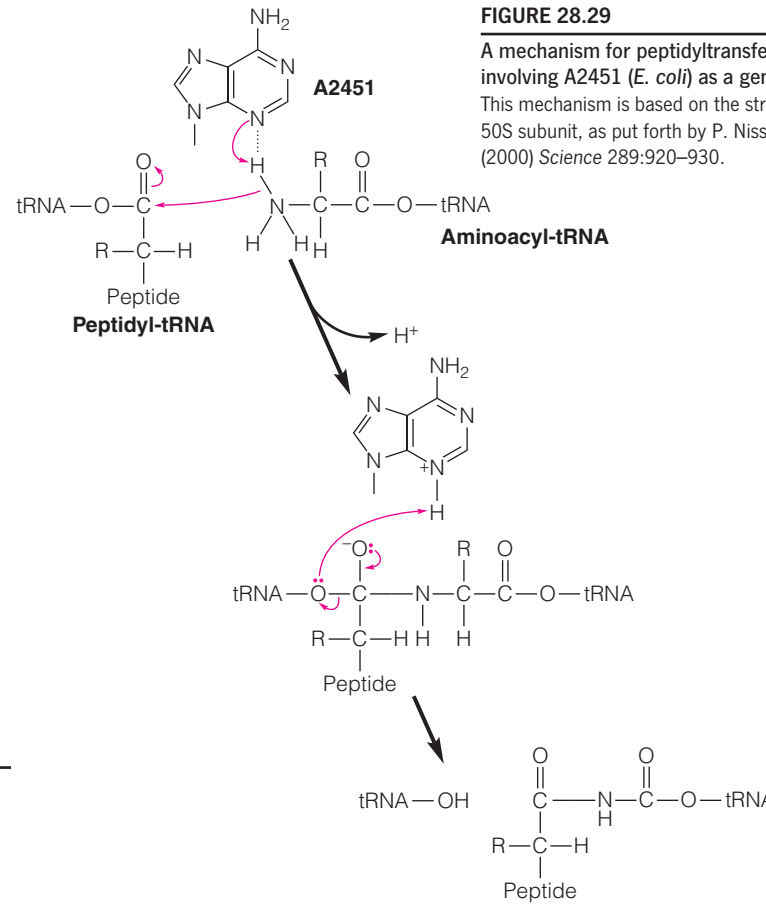
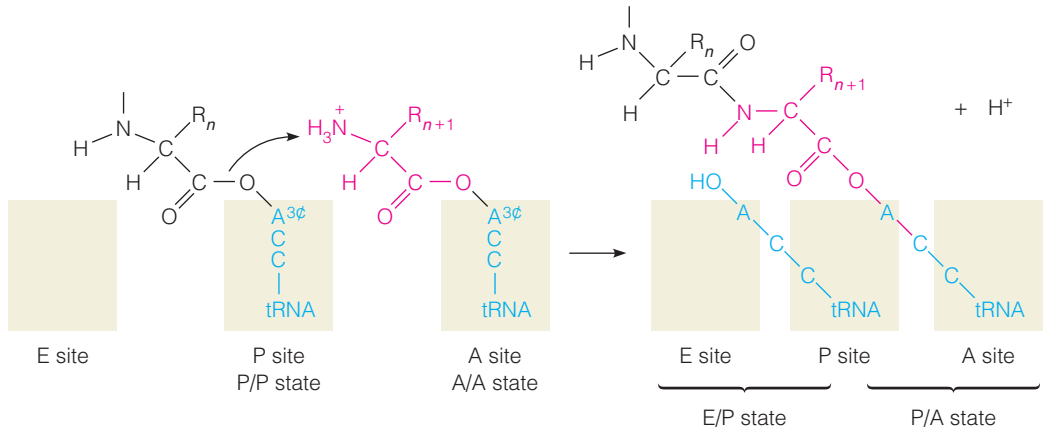


FIGURE 28.29

A mechanism for peptidyltransferase involving A2451 (*E. coli*) as a general base. This mechanism is based on the structure of the 50S subunit, as put forth by P. Nissen et al. (2000) *Science* 289:920–930.

N3 then stabilizes a tetrahedral carbon intermediate by binding to the oxyanion. The proton is then transferred to the peptidyl-tRNA 3′ hydroxyl as the newly formed peptide deacylates. Concomitant with this transfer is a switch from the simple P and A states to hybrid states, in which the acceptor ends of the two tRNA molecules move into the leftward positions while the codon ends remain fixed as before. These hybrid sites are indicated as E/P and P/A. This can be considered the first half of the *translocation* step (step 3 in Figure 28.26).



In elongation, the growing peptide chain at the P site is transferred to the newly arrived aminoacyl tRNA in the A site. Translocation then moves this tRNA to the P site and the previous tRNA to the E site.

To complete the translocation step, the anticodon end of the now uncharged tRNA in the P site is transferred to the E site, and the tRNA in the A site (the tRNA that now has the nascent polypeptide chain attached to it) is moved completely to the P site. In the process, the ribosome moves a three-nucleotide step in the 3′ direction along the mRNA, placing a new codon adjacent to the now empty A site. Like peptidyl transfer, this step requires a protein factor (EF-G) bound to GTP and requires GTP hydrolysis. Crystallographic studies reveal a remarkable "molecular mimicry" between EF-G–GTP and the ternary complex aa-tRNA–EF-Tu–GTP. As Figure 28.30 shows, the protein and the RNA–protein complex have almost exactly the same shape, even though they differ entirely in composition and sequence. It is speculated that the reason for this similarity is to allow EF-G–GTP to move temporarily into the A site, facilitating the displacement of the peptidyl–tRNA complex. Structural studies support this model.

During translation the ribosome is "ratcheted" along the mRNA molecule by a process involving rotation of the two ribosomal subunits with respect to each other. Based on structural analysis of the ribosome in intermediate states of rotation, the ratcheting process has been schematized as shown in Figure 28.31.

At this point the E and P sites are occupied, but A is empty. As the deacylated tRNA is released from E (step 4, Figure 28.26), the A site gains high affinity and accepts the aminoacyl tRNA dictated by the next codon. A cycle of elongation is now complete. All is as it was at the start, except that now:

1. The polypeptide chain has grown by one residue.
2. The ribosome has moved along the mRNA by three nucleotide residues—one codon.
3. At least two molecules of GTP have been hydrolyzed.

The whole process is repeated again and again until a termination signal is reached, with the newly synthesized polypeptide chain being forced to exit the ribosome through the tunnel mentioned previously.

Termination

The completion of polypeptide synthesis is signaled by the translocation of one of the *stop codons* (UAA, UAG, or UGA) into the A site. Because there are no tRNAs that recognize these codons under normal circumstances, termination of the

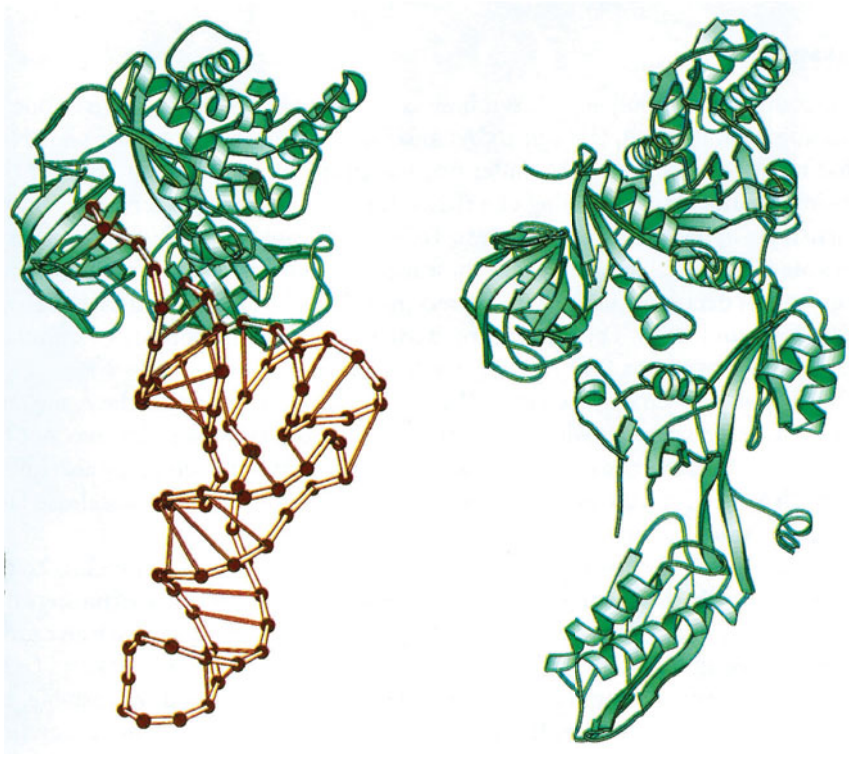


FIGURE 28.30

The striking structural similarity between the translocation factor EF-G (right) and the ternary complex aa~tRNA-EF-Tu–GTP (left). Protein is shown in green, RNA in brown.

From *Science* 270:1464–1472, P. Nissen, M. Kjeldgaard, S. Thirup, G. Polekhina, L. Reshetnikova, B. F. C. Clark, and J. Nyborg, Crystal structure of the ternary complex of Phe-tRNA^{Phe}, EF-Tu, and a GTP analog. © 1995. Reprinted with permission from AAAS.

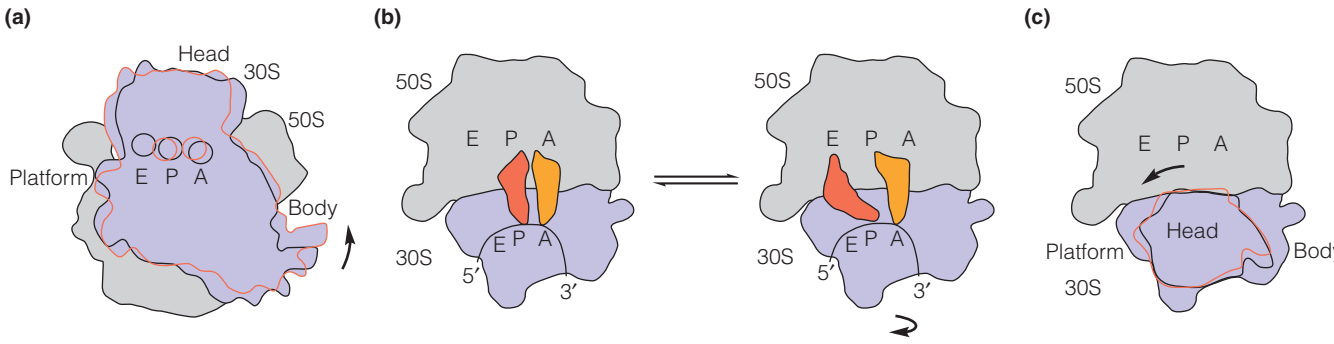


FIGURE 28.31

A schematic view of ribosome subunit rotational motions, based on crystal structures of ribosomes in intermediate states. (a) View from the bottom. 30S subunit (blue) is shown in starting conformation after termination (outlined in red) to a fully rotated conformation seen during elongation (black outline). **(b)** Side view. During transition to the fully rotated state, tRNAs shift from binding in A/A and P/P sites (30S/50S) to occupying hybrid sites A/P and P/E. **(c)** Rotation in another plane can move the head domain of the 30S subunit as much as 14° toward the E site.

From *Science* 325:1014–1017, W. Zhang, J. A. Dunkle, and J. H. D. Cate, Structures of the ribosome in intermediate states of ratcheting. © 2009. Reprinted with permission from AAAS.

chain does not involve binding of a tRNA. Instead, protein *release factors* participate in the termination process. The three release factors found in prokaryotes are listed in Table 28.4. Two of these factors can bind to the ribosome when a stop codon occupies the A site: RF1 recognizes UAA and UAG, and RF2 recognizes UAA and UGA. The third factor, RF3, is a GTPase that appears to stimulate the release process, via GTP binding and hydrolysis. Structural analysis of the ribosome complexed with RF2 show that the release factor interacts directly with a UGA termination codon (Figure 28.32).

The sequence of termination events is as shown in Figure 28.33. After RF1 or RF2 has bound to the ribosome, the peptidyltransferase transfers the C-terminal residue of the polypeptide chain from the P-site tRNA to a water molecule, releasing the peptide chain from the ribosome. The chemistry of this reaction is similar to peptide bond formation (Figure 28.29), except that water replaces the α-amino group as the attacking nucleophile. The RF factors and GDP are then released, followed by the tRNA. The 70S ribosome is now unstable. Its instability is accentuated by the presence of a protein called *ribosome recycling factor,* and also by the initiation factors IF3 and IF1, and the ribosome readily dissociates to 50S and 30S subunits prepared for another round of translation.

When the ribosomal subunits separate, the 30S subunit may or may not dissociate from its mRNA. In some cases in which polycistronic messages are being translated, the 30S subunit may simply slide along the mRNA until the next

Termination requires protein release factors that somehow recognize stop codons.

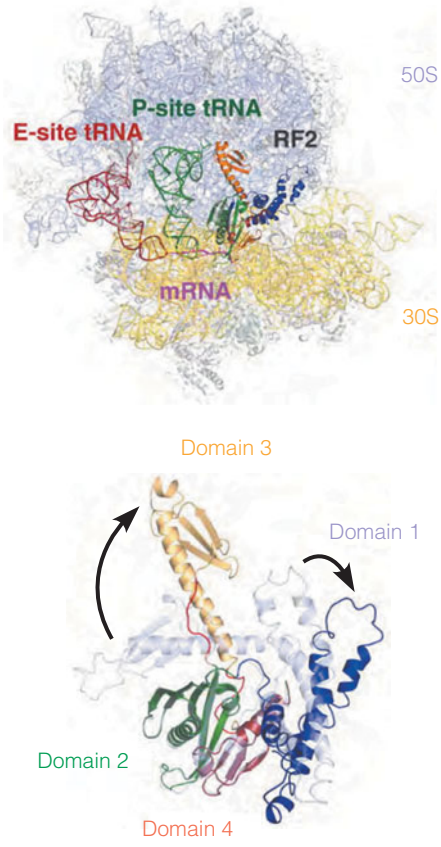


FIGURE 28.32

Interaction of RF2 with a UGA stop codon in the decoding center. Above, the ribosome in complex with RF2. UGA is in magenta, and RF2 is in green, except for those parts interacting directly with the codon, which are shown in red. RF2 helical domains that move significantly are shown in color, with the extent of the movements (domains 1 and 3), in the same color scheme, shown below. PDB ID 2jl5 to 2jl8.

From *Science* 322:953–956, A. Wexelbaumer, H. Jin, C. Neubauer, R. M. Voorhees, S. Petry, A. C. Kelley, and V. Ramakrishnan, Insights into translational termination from the structure of RF2 bound to the ribosome. © 2008. Reprinted with permission from AAAS.

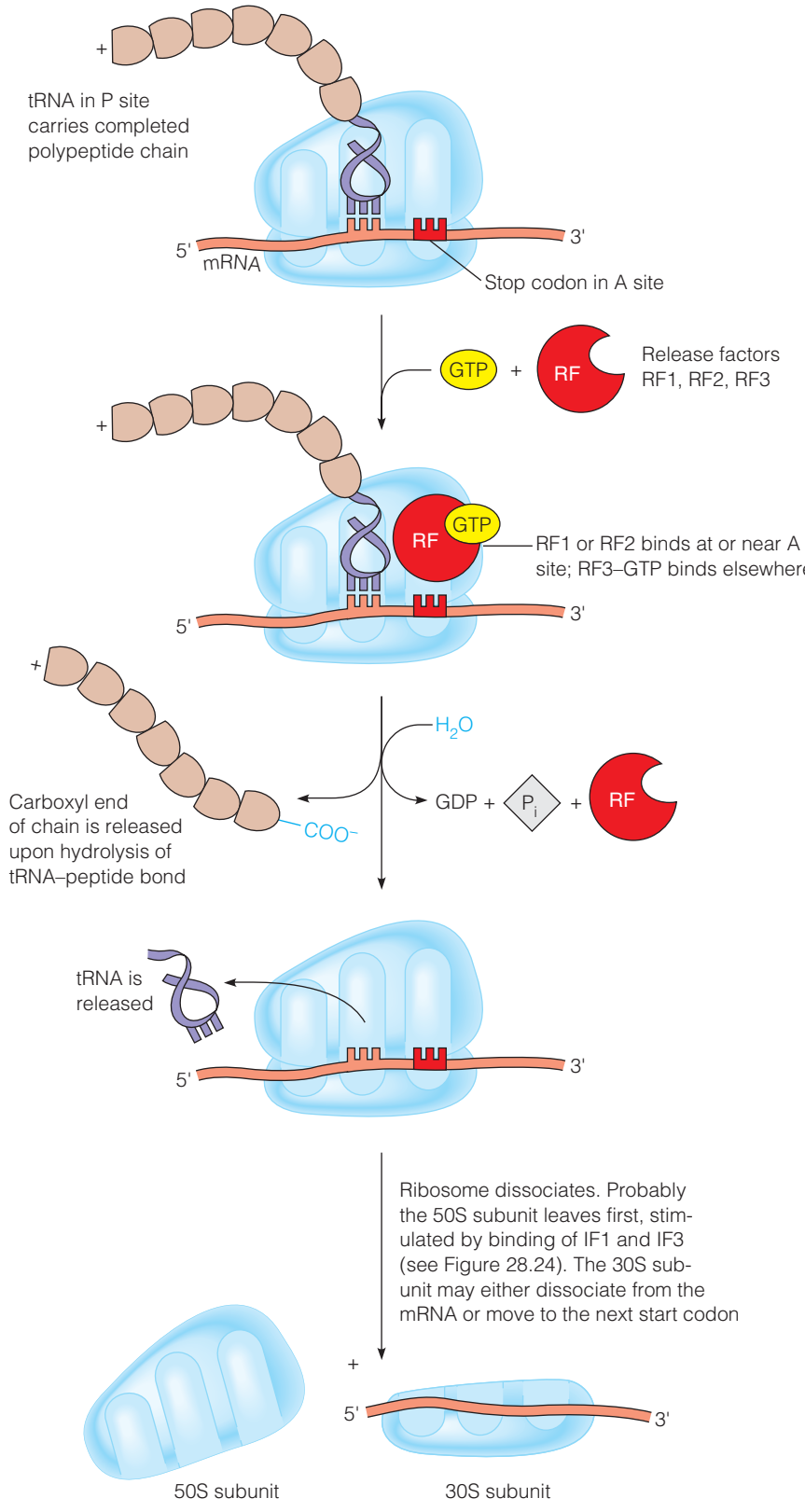


FIGURE 28.33

Termination of translation in prokaryotes.

Shine–Dalgarno sequence and initiation codon are encountered and then begin a new round of translation. If the 30S subunit does dissociate from the message, it will soon reattach to another one.

Suppression of Nonsense Mutations

Understanding the process of termination helped clarify some peculiar observations concerning nonsense mutations. Recall from Chapter 7 that a *nonsense mutation* is one in which a codon for some amino acid has been mutated into a stop codon so that the polypeptide chain terminates prematurely. These mutations were originally discovered because their phenotypic expression could be *suppressed* by a class of mutations located in other genes. Recall from Chapter 25 that suppression is defined genetically as restoration of wild-type function by a second mutation at a different site. When this second mutation occurs in a different gene, the phenomenon is called **intergenic suppression**. Upon examination, the suppressors of nonsense mutations were found to lie in tRNA genes.

Consider the example shown in Figure 28.34. A nonsense mutation has changed a codon that normally specifies the amino acid tyrosine into a stop codon, causing premature termination of the polypeptide chain. If, however, one of the several tyrosine tRNAs mutates in its anticodon region so as to recognize

The effects of nonsense mutations can sometimes be suppressed by suppressor mutations, in which a tRNA mutates to recognize a stop codon and inserts an amino acid instead.

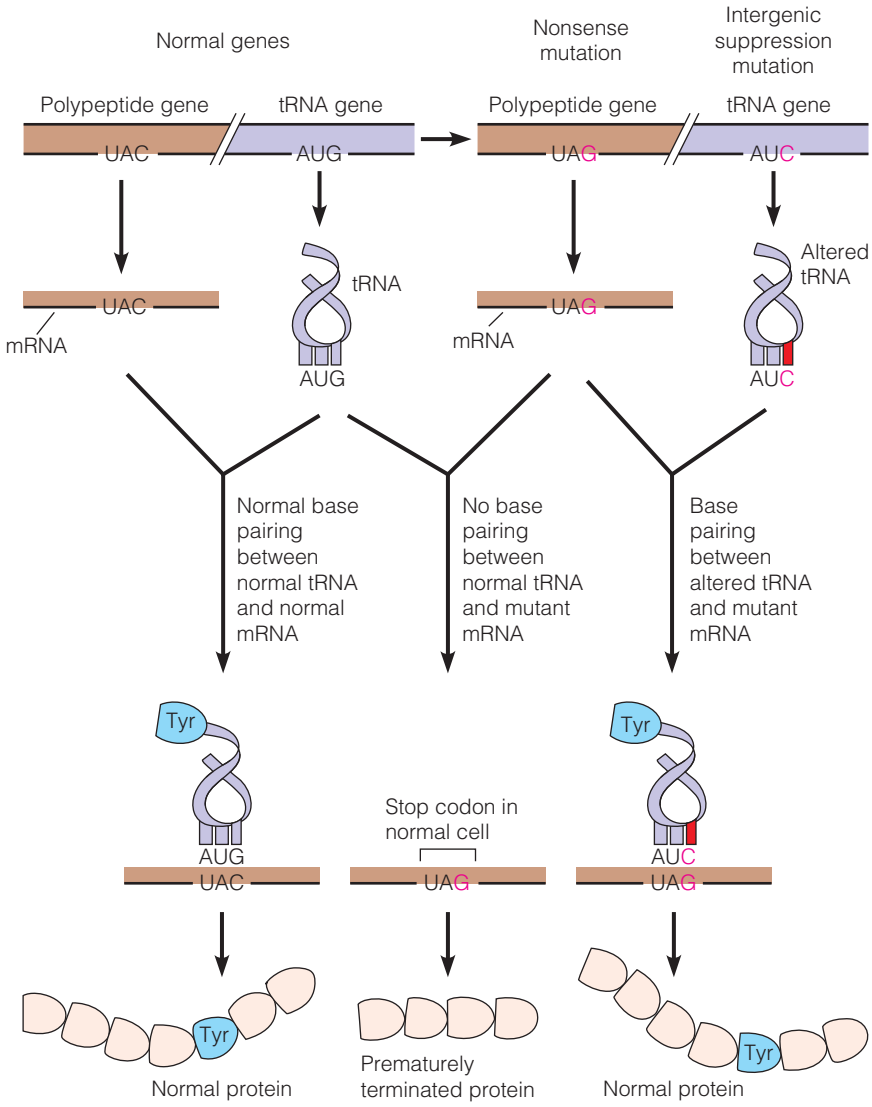


FIGURE 28.34

How an intergenic suppression mutation can overcome a nonsense mutation. A nonsense mutation in a protein-coding gene changes a codon for an amino acid into a stop codon, causing translation to terminate prematurely. Another mutation, in a tRNA gene, can circumvent the first mutation by altering the tRNA anticodon so that it will base-pair with the mutant mRNA. A functional protein is produced in this situation, even though suppression might not restore the original amino acid at that site.

the stop codon, translation can sometimes proceed in the normal fashion. Thus, a mutation that might otherwise be lethal can be suppressed by such a change, and the organism can survive. Clearly, it will still have problems, for the presence of such a mutated tRNA will interfere with the normal termination of other proteins. That they can survive at all depends on the fact that the suppressor mutation usually involves a minor tRNA species, little used in normal translation. Furthermore, such effects may be minimized by the frequent occurrence of two or more different stop signals in tandem in mRNAs. Even if the first stop codon is suppressed, the "emergency brake" still holds.

Suppressor mutations are by no means confined to correction of nonsense mutations. Some mutated tRNAs correct missense mutations, and some even contain two or four bases acting as the anticodon. These can therefore serve as **frameshift suppressors**.

Inhibition of Translation by Antibiotics

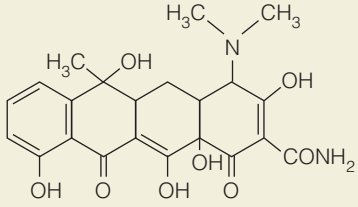
Many antibiotics have been shown to act by inhibiting specific steps in bacterial protein synthesis. Some of these antibiotics have proven to be useful reagents in analyzing mechanisms in translation, as well as in combating infections. We have already described the action of some kinds of antibiotics. In Chapter 9, we saw that the penicillins inhibit bacterial cell wall synthesis, and in Chapter 10 we discussed antibiotics such as gramicidin and valinomycin, which interfere with the ionic balance across membranes. Other antibiotics, such as rifampicin and streptolydigin (Chapter 27), block transcription in prokaryotes.

A number of important antibiotics act by inhibiting translation in bacterial cells.

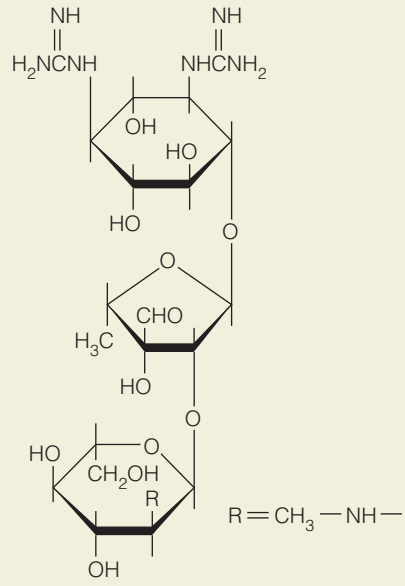
A host of naturally occurring substances interfere with various stages of protein synthesis. Some of these are shown in Figure 28.35. Each inhibits translation in a different way. Their importance to medicine stems largely from the fact that the translational machinery of eukaryotes is sufficiently different from that of prokaryotes that these antibiotics can be used safely in humans. In some cases (for example, the tetracyclines), antibiotics that would also inhibit eukaryotic translation are nevertheless harmless to eukaryotes because they cannot traverse the cell membranes of higher organisms.

A major problem with the therapeutic use of antibiotics is that microorganisms can develop resistance to many of them. An important example is *erythromycin* resistance. The erythromycin-binding site on the ribosome includes a specific region of the 23S RNA, and binding of the antibiotic can be inhibited by an enzyme that methylates a specific adenine residue in this region. Molecular biologists use erythromycin resistance in screening bacterial clones in recombinant DNA research. Resistance to erythromycin can be conferred to a bacterium by the insertion of a resistance gene coding for the methyltransferase on a bacterial plasmid. Bacteria containing the plasmid carrying the methylase gene will grow in an erythromycin-containing medium, whereas those lacking the plasmid will be killed. Thus, growth on such a medium automatically selects for only those clones that carry the plasmid. Because many such resistance genes are carried on plasmids, which are easily transferred from bacterium to bacterium, the frequency with which antibiotic-resistant strains arise is far higher than if the elements were carried on chromosomal genes. Another problem is the widespread use of antibiotics in animal husbandry, not to cure an infection, but rather to suppress any possible infections for improved animal weight maintenance and to prevent the spread of infection from animal to animal, under the crowded conditions in feedlots. That may be fine in the short run, but the resultant increased emergence of antibiotic-resistant strains makes many question the wisdom of this practice.

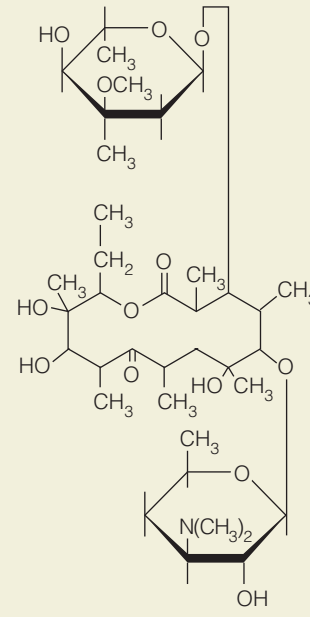
Structural studies on ribosomes have given enormous impetus to the development of new classes of antimicrobial agents to which resistance might not develop. In the same sense that knowledge of enzyme and receptor structure makes it possible to design entirely new inhibitors with therapeutic properties, the



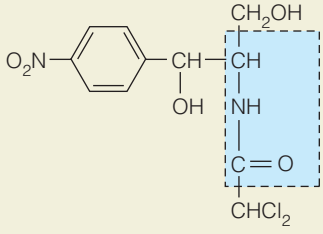
Tetracycline: Inhibits the binding of aminoacyl tRNAs to the ribosome and thereby blocks continued translation



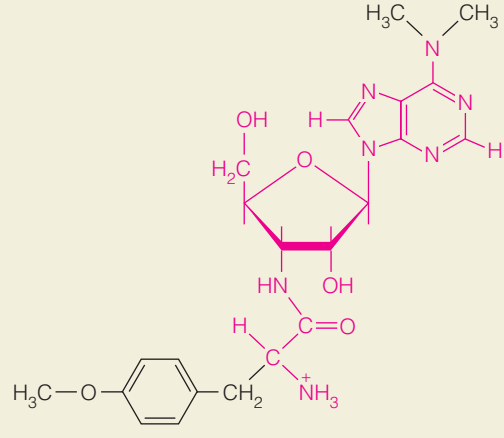
Streptomycin: Interferes with normal pairing between aminoacyl tRNAs and message codons, causing misreading, and thereby producing aberrant proteins



Erythromycin: Binds to a specific site on the 23S RNA and blocks elongation by interfering with the translocation step



Chloramphenicol: Blocks elongation, apparently by acting as competitive inhibitor for the peptidyltransferase complex. The amide link (in blue) resembles a peptide bond



Puromycin: Causes premature chain termination. The red portion of the molecule resembles the 3′ end of the aminoacylated tRNA. It will enter the A site and transfer to the growing chain, causing premature chain release

FIGURE 28.35

Some antibiotics that act by interfering with protein biosynthesis. Erythromycin is one of the polyketide antibiotics whose biosynthesis was discussed in Chapter 17.

ribosome, because of its many activities and its structural conservation among bacteria, is an extremely attractive target for drug development.

Translation in Eukaryotes

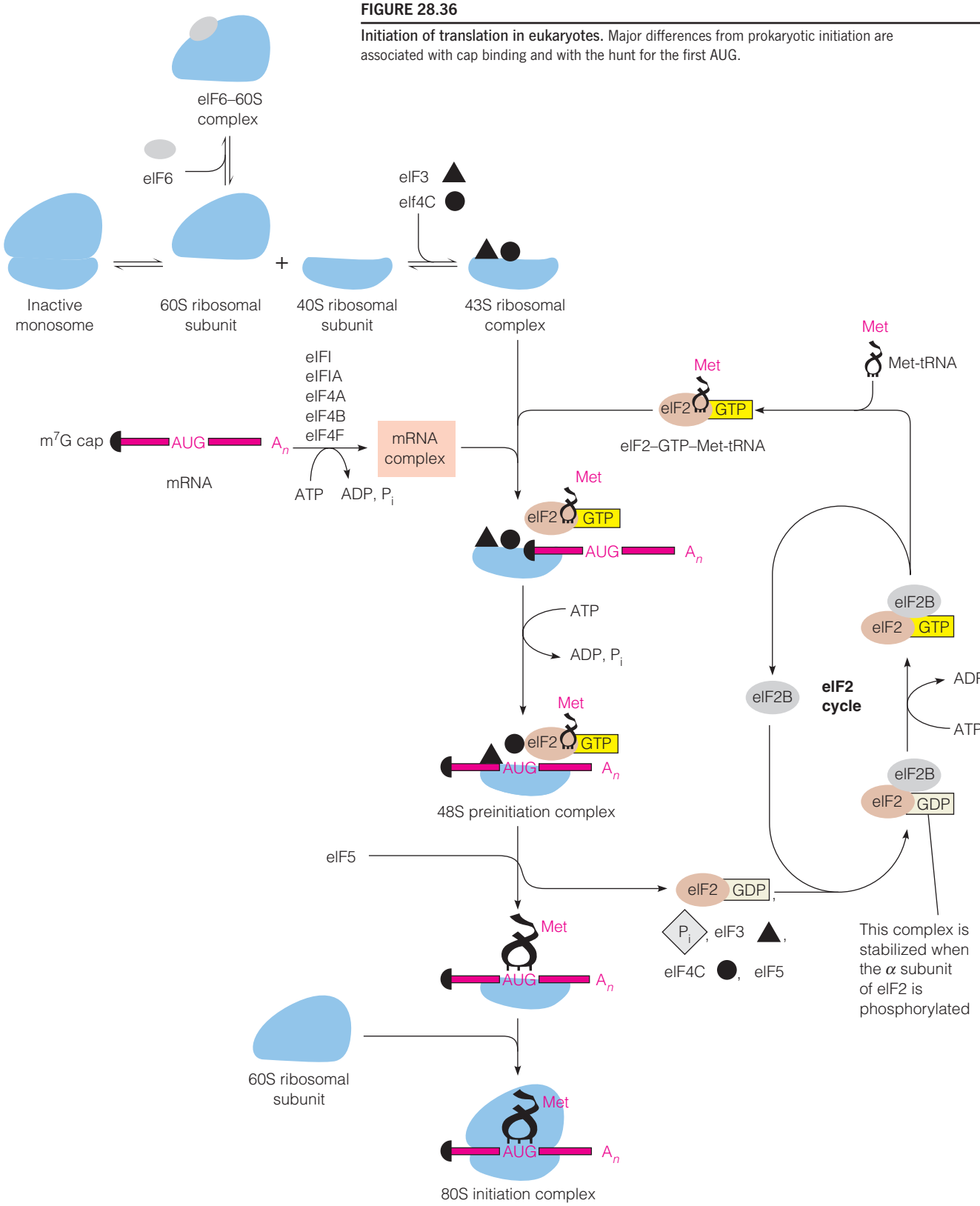
In eukaryotes, translational initiation is more complex and requires more protein factors than in prokaryotes.

The mechanism for translating messenger RNA into protein in eukaryotic cells is basically the same as in prokaryotes. In eukaryotes the ribosomes are larger and more complex, and virtually all mRNAs are monocistronic. There are more

soluble protein factors, as we saw in Table 28.4, but the functions performed are comparable to those we have discussed for bacteria. The most significant differences are in initiation mechanisms; these are schematized for eukaryotic cells in Figure 28.36. Aside from the greater complexity of the ribosome and soluble

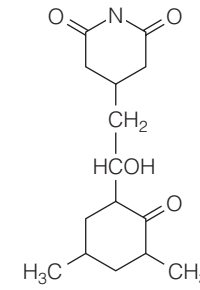
FIGURE 28.36

Initiation of translation in eukaryotes. Major differences from prokaryotic initiation are associated with cap binding and with the hunt for the first AUG.



protein factors, the major differences are (1) that the 5′ end of a message is sensed not by a Shine-Dalgarno sequence, but by the 7-methylguanine cap, and (2) the N-terminal amino acid, inserted at the initiator AUG, is methionine, not *N*-formylmethionine. After detecting the 5′ cap, the ribosomal 40S subunit then scans along the mRNA (an ATP-dependent process) until the first AUG is found. At this point the initiation factors are released, and the 60S subunit is attached to begin translation.

A number of the common inhibitors of prokaryotic translation are also effective in eukaryotic cells. They include pactamycin, tetracycline, and puromycin. There are also inhibitors that are effective *only* in eukaryotes. Two important ones are *cycloheximide* and *diphtheria toxin.* Cycloheximide inhibits the translocation activity in the eukaryotic ribosome and is often used in biochemical studies when processes must be studied in the absence of protein synthesis. Diphtheria toxin is an enzyme, coded for by a bacteriophage that is lysogenic in the bacterium *Corynebacterium diphtheriae.* It catalyzes a reaction in which NAD^+ adds an *ADP-ribose* group to a specially modified histidine in the translocation factor eEF2, the eukaryotic equivalent of EF-G (Figure 28.37). Because the toxin is a catalyst, minute amounts can irreversibly block a cell's protein synthetic machinery; pure diphtheria toxin is one of the deadliest substances known.



Cycloheximide

Protein Synthesis in Organelles

As described in Chapter 15, the mitochondrial genome (mtDNA) contains 37 genes that encode 13 proteins (in humans), all of which are subunits of the respiratory chain complexes. The remaining mtDNA genes encode 22 tRNAs and 2 rRNAs. These tRNAs and rRNAs are part of the mitochondrial protein synthesis machinery, required to translate the 13 proteins encoded in mitochondrial DNA. Reflecting its evolution from an ancient alpha-proteobacterium, the mitochondrial protein synthesis machinery is more closely related to the bacterial system than to the eukaryotic cytosolic system. Like the prokaryotic process, translation initiates with formylated Met-tRNA in mitochondria and requires only a handful of initiation and elongation factors. Also, mitochondrial protein synthesis is inhibited by some antibiotics that interfere with steps in bacterial protein synthesis. However, mitochondrial ribosomes have undergone a major remodeling during mitochondrial evolution. The rRNAs of mammalian mitochondrial ribosomes are smaller than their bacterial counterparts, and the large subunit of mitochondrial ribosomes completely lacks a 5S rRNA component. On the other hand, mitochondrial ribosomes have more protein subunits so that mitochondrial ribosomes have a protein:RNA ratio of 2:1, compared to a 1:2 ratio for the bacterial ribosome. Chloroplasts also possess their own protein synthesis machinery, but much less is known about the chloroplast system.

Rates and Energetics of Translation

Translation is a rapid process in prokaryotes. At 37 °C an *E. coli* ribosome can synthesize a 300-residue polypeptide chain in about 20 seconds. This means that a single ribosome passes through about 15 codons, or 45 nucleotides, in each second. This rate is almost exactly the same as our best estimates of the rate of prokaryotic *transcription,* which means that mRNA can be translated as fast as it is transcribed. That equality is not a coincidence. Recent studies with *E. coli* show that a ribosomal protein, NusE, interacts in the cell with an RNA polymerase component, NusG, and that through this interaction transcription and translation are physically coupled, as shown in Figure 28.38, with the rate of transcription being controlled by the rate of translation. Direct coupling of this type cannot occur in eukaryotic cells because the two processes take place in separate compartments.

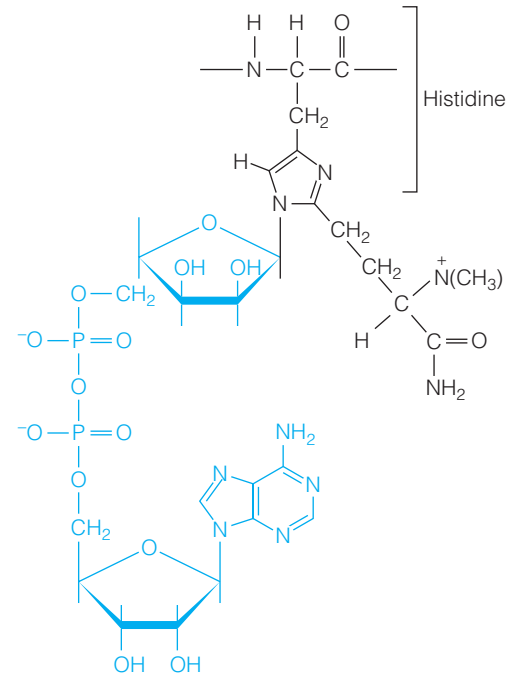


FIGURE 28.37

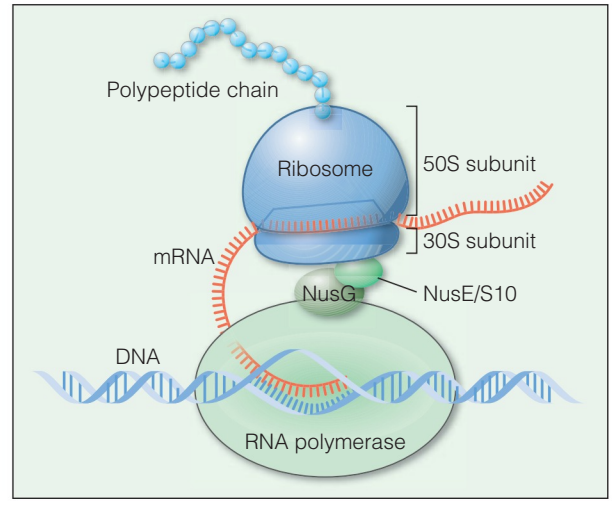
ADP-ribosylated diphthamide derivative of histidine in eEF2. Synthesis of this derivative of a modified histidine in eEF2 using NAD^+ is catalyzed by diphtheria toxin. eEF2 is inactivated, and protein synthesis is therefore blocked. ADP ribose from NAD^+ is shown in blue. Diphthamide is in black.

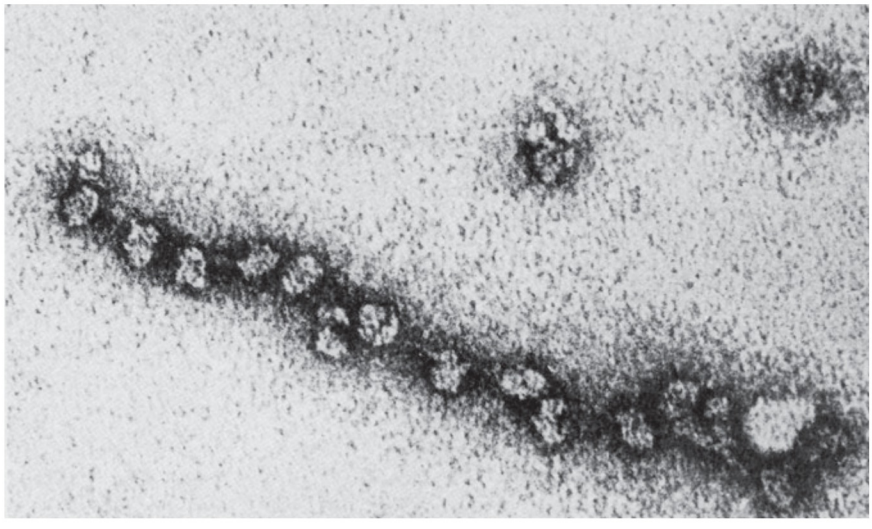
But the rate mentioned above represents the growth of individual polypeptide chains and does not account for the total rate of protein synthesis in the cell because many ribosomes may be simultaneously translating a given message. In fact, if we were to carefully lyse *E. coli* cells, we would observe **polyribosomes** (also called polysomes) like those shown in Figure 28.39. Apparently, as soon as one

FIGURE 28.38

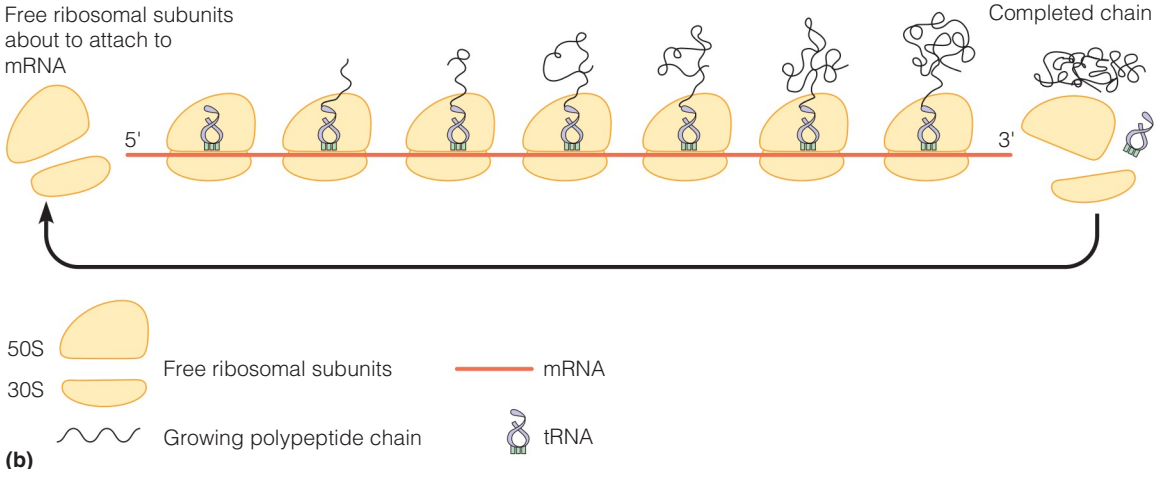
Coupling of transcription to translation in *E. coli* via the interaction between NusE and NusG.

From *Science* 328:436–437, J. W. Roberts, Syntheses that stay together. © 2010. Reprinted with permission from AAAS.





(a)



(b)

FIGURE 28.39

Polyribosomes. (a) Electron micrograph showing *E. coli* polyribosomes. The ribosomes are closely clustered on an mRNA molecule. **(b)** Schematic picture of a polyribosome like that shown in **(a)**. Each ribosome is to be imagined as moving from left to right.

(a) Courtesy of Barbara Hamkalo; (b) *Molecular Biology of the Gene*, 4th ed., James D. Watson, Nancy H. Hopkins, Jeffrey W. Roberts, Joan Argetsinger Steitz, and Alan M. Weiner. © 1987. Reprinted by permission of Pearson Education Inc., Upper Saddle River, NJ.

ribosome has moved clear of the 5′ region of the mRNA, another attaches. Under some conditions, as many as 50 ribosomes may be packed onto an mRNA, with one finishing translation every few seconds. Because each *E. coli* cell contains 15,000 ribosomes or more, all of them operating at full capacity can synthesize about 750 protein molecules of 300 residues each second.

The energy cost for this process is high. If we examine the individual steps in protein synthesis described earlier, we can make the following estimate of the total energy budget for synthesizing a protein of N residues:

	The *equivalent* of
$2N$	ATPs are required to charge the tRNAs because the ATP is cleaved to AMP and PP_i, and PP_i is subsequently hydrolyzed.
1	GTP is needed for initiation.
N–1	GTPs are required to form the N—1 peptide bonds, in the EF-Tu–GTP hydrolysis step.
N–1	GTPs are necessary for the N—1 translocation steps.
1	GTP is required in termination.
Sum	$= 4N$

Translation is fast but energy-expensive. About four ATP equivalents are needed for each amino acid added.

Altogether, then, about $4N$ high-energy phosphate molecules must be hydrolyzed to complete a chain of N units. This is a minimal estimate, for it does not include the energy required to formylate methionine, nor any extra GTPs that may be expended in proofreading and replacing incorrectly bound tRNAs. Furthermore, there have been persistent, although debated, reports that *two* GTPs must be hydrolyzed for every aa-tRNA bound to the A site. But even at the conservative estimate, a typical protein of 300 residues costs the cell about 60,000 kJ of free energy per mole, if we assume ATP or GTP hydrolysis yields about 50 kJ/mol under cellular conditions. Proteins are expensive!

If we express the same data in terms of the energy requirement for synthesis per mole of *peptide bond,* we obtain a cost of about 200 kJ/mol. Given that the free energy change required to form a peptide bond in dilute aqueous solution is only about +20 kJ/mol, the price seems exorbitant. Why does the cell have no mechanism for making peptide bonds for a few dozen kilojoules each? Certainly, an input of even 40 kJ/mol would be enough to make the synthesis process very favorable—with an equilibrium constant of about 3000.

The key to this great energy expenditure is found in the fundamental nature of life. The cell is making polypeptides of *defined* sequence. If it were simply throwing together amino acids at random, the free energy price could be much cheaper. But a chain of 300 residues, made from 20 different amino acids, can be put together in 20^{300} different ways, whereas the cell needs *one* specific sequence. There is, in other words, a large entropy price to be paid in making specific sequences—and making them correctly. What this means at the mechanistic level is that every step in the assembly not only must be done with a free energy excess but also must involve a specific *choice.* Furthermore, the product must, at critical points, be checked by a proofreading mechanism, which in turn costs more energy. It is expensive to get a good translation of a book, for not only must the translators be expert and careful but their work must also be rechecked with great care.

The Final Stages in Protein Synthesis: Folding and Covalent Modification

The polypeptide chain that emerges from the ribosome is not a completed, functional protein. It must fold into its tertiary structure, and it may have to associate with other subunits. In some cases, disulfide bonds must be formed, and other covalent modifications, such as hydroxylation of specific prolines and lysines

must take place. Complexation with carbohydrate or lipid occurs after translation. In addition, many proteins are subjected to specific proteolytic cleavage to remove portions of the nascent chain.

Chain Folding

The cell need not wait until the entire chain is released from the ribosome to commence its finishing touches. The first portion of the nascent chain (about 30 residues) is protected as it passes through the tunnel in the ribosome. However, changes begin almost as soon as the N-terminal end emerges. There is good evidence that folding into the tertiary structure starts during translation and is nearly complete by the time the chain is released. For example, antibodies to *E. coli* β-galactosidase, which recognize the tertiary folding of the molecule, will attach to polyribosomes synthesizing this protein. This enzyme displays catalytic activity only as a tetramer. It has been demonstrated that nascent β-galactosidase chains, still attached to ribosomes, can associate with free subunits to form a functional tetramer. Thus, even quaternary structure can be partially established before synthesis is complete.

This behavior should not be surprising, if we recall (from Chapters 6 and 7) that formation of the secondary, tertiary, and quaternary levels of protein structure is thermodynamically favored. However, as we have seen in Chapter 6, in some cases this spontaneous folding must be aided by chaperone proteins.

Covalent Modification

Some of the covalent modifications of polypeptide chains also occur during translation. We mentioned earlier that the *N*-formyl group is removed from the initial *N*-fMet of most prokaryotic proteins. A specific *deformylase* catalyzes this reaction. In many cases, deformylation seems to happen almost as soon as the N-terminus emerges from the ribosome. Removal of the N-terminal methionine itself can also be an early event, but whether it happens or not apparently depends on the cotranslational folding of the chain. Presumably, in some cases this residue is "tucked away" and protected from proteolysis in the folded structure.

Translation is immediately followed by various kinds of protein processing, including chain folding, covalent modification, and directed transport.

Some prokaryotic (and many eukaryotic) proteins experience much more severe proteolytic modifications. These proteins are usually the ones that are going to be exported from the cell or are destined for membrane or organelle locations. We discuss the more complicated eukaryotic protein processing in the next section and concentrate here on what happens in prokaryotes.

Bacterial proteins that are destined for secretion (**translocation** across the cell membrane) are characterized by highly hydrophobic **signal sequences** or **leader sequences** in the N-terminal regions. Representatives are listed in Table 28.5. After the protein has passed through the membrane, the leader sequence is cleaved off at the point indicated by the arrow in the table.

A current model for translocation in bacteria is shown in Figure 28.40. In many, but not all, cases the protein to be translocated (the pro-protein) is first complexed in the cytoplasm with a chaperone—the SecB protein in the example shown. This complexing keeps the protein from folding prematurely, which

TABLE 28.5 N-terminal signal sequences of representative prokaryotic proteins

Protein	−20					−15					−10					−5					−1 ↓	+1	
Leucine-binding protein	M	K	A	N	A	K	T	I	I	A	G	M	I	A	L	A	I	S	H	T	A	M	A	E	E
Prealkaline phosphatase			M	K	Q	S	T	I	A	L	A	L	L	P	L	L	F	T	P	V	T	K	A	R	T
Prelipoprotein				M	K	A	T	K	L	V	L	G	A	V	I	L	G	S	T	L	L	A	G	C	S

Note: Hydrophobic residues are in magenta. The cleavage site is designated by the arrow.

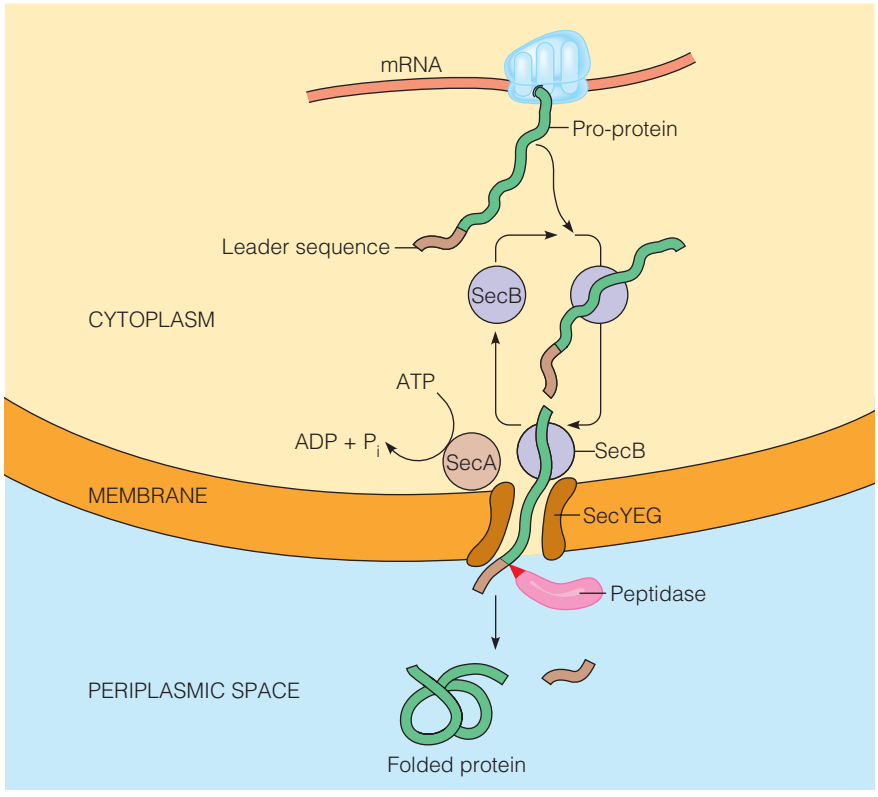


FIGURE 28.40

A current model for protein secretion by prokaryotes. The new polypeptide chain (the pro-protein) complexes with SecB, which prevents complete folding during transport to the membrane. At the membrane an ATPase, SecA, drives translocation through the membrane with the aid of SecYEG, which forms a membrane pore. The leader sequence is then cleaved off the secreted protein by a membrane peptidase.

would prevent it from being passed through a secretory pore in the membrane. This pore is composed of a heterotrimeric protein made up of SecE, SecY, and SecG—the "SecYEG translocon." The secretory pore is also a target for a fourth protein component, SecA. SecA is an ATPase, and both ATP hydrolysis and the electrochemical potential gradient across the membrane help drive translocation. Structural analysis of the SecA protein with and without bound adenine nucleotide suggests a mechanism comparable to that of DNA-dependent helicases in driving the protein through the membrane. After the pro-protein has been translocated, the leader peptide is cleaved off by a membrane-bound protease, and the protein can fold. The cleavage site, as shown in Table 28.5, usually lies between a small amino acid (often Gly or Ala) and an acidic or a basic one.

Protein Splicing

A small, but significant, number of proteins, mostly from single-celled organisms—bacteria, archaea, and eukaryotic microbes—undergo post-translational splicing, in a process comparable to the RNA splicing discussed in Chapter 27. In protein splicing an **intein** (internal protein segment) is cleaved from within the polypeptide sequence, yielding a mature protein, the **extein** (external protein). Embedding an intein into a normally nonspliced protein retains its splicing activity for the new protein host, indicating that amino acid residues for catalysis of splicing lie within the intein.

Although we know little about the biological functions of protein splicing, we do know that the mechanism is comparable to that of RNA splicing, with, of course, different functional groups involved. As shown in Figure 28.41, an N-terminal serine or threonine hydroxyl (or cysteine thiol; serine as shown) attacks the C-terminal peptide carbon of the upstream intein segment (N-extein), and this is followed by a transesterification involving a thiol or hydroxyl in the N-terminal residue of the downstream extein segment (C-extein) upon that same

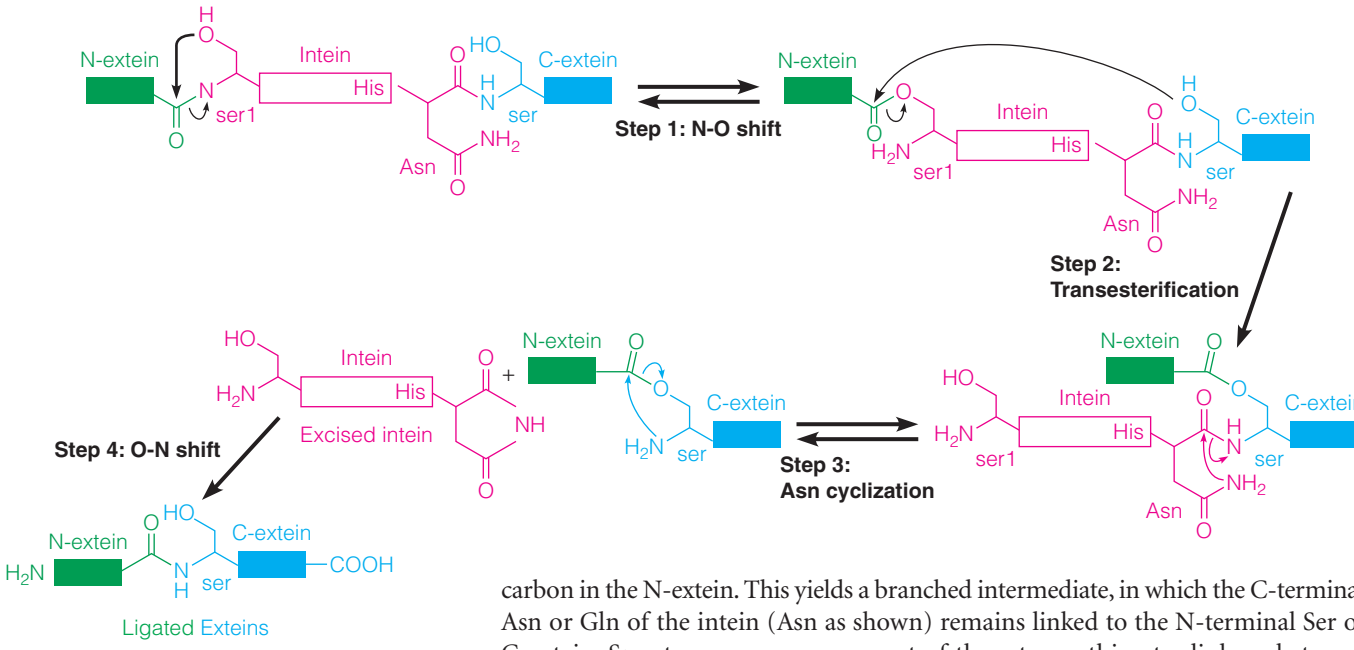


FIGURE 28.41

An outline of the mechanism of protein splicing.

carbon in the N-extein. This yields a branched intermediate, in which the C-terminal Asn or Gln of the intein (Asn as shown) remains linked to the N-terminal Ser of C-extein. Spontaneous rearrangement of the ester or thioester linkage between the ligated exteins yields the more stable peptide bond linking N-extein to C-extein. Most inteins encode a homing nuclease comparable to that seen in introns, suggesting that the functions of RNA and protein splicing are similar, in promoting their ability to move from gene to gene.

Protein Targeting in Eukaryotes

The eukaryotic cell is a multicompartmental structure. Its several organelles each require different proteins, only a few of which are synthesized within the organelles themselves. Most mitochondrial and chloroplast proteins, for example, are encoded by the nuclear genome and synthesized in the cytoplasm. They must be carefully distinguished from other newly synthesized proteins and selectively transported to their appropriate addresses. Other new proteins are destined for export out of the cell or into vesicles like lysosomes. The diversity of destinations for different proteins implies the existence of a complex system for labeling and sorting newly synthesized proteins and ensuring that they end up in their proper places. And, as seen with bacteria, there must be a process by which protein molecules, which may be hydrophilic, engage the hydrophobic membrane and find a way either to pass through or, as in the case of integral membrane proteins, to become embedded within the membrane.

Proteins Synthesized in the Cytoplasm

Proteins destined for the cytoplasm and those to be incorporated into mitochondria, chloroplasts, or nuclei are synthesized on polyribosomes free in the cytoplasm. The proteins targeted to organelles, as initially synthesized, contain specific signal sequences. These sequences probably aid in membrane insertion, but they also signal that these polypeptides will interact with a particular class of chaperones. These chaperones are members of the "heat-shock" Hsp70 family, and they act to ensure that the newly synthesized protein remains unfolded and is delivered to a receptor site on the organelle membrane. The unfolded protein then passes through membranes, through gates containing transport proteins that discriminate among proteins destined for the lumen, the membranes, or an organelle matrix. If it passes into an organelle matrix, the protein may be taken up by intraorganelle chaperones for final folding. The N-terminal targeting sequence is also cleaved off during this transport.

Transport of proteins into mitochondria is schematized in Figure 28.42. First, the Hsp70-bound protein attaches *via* a basic N-terminal signal sequence to a receptor protein as part of a structure called the TOM complex (translocation of outer membrane). An ATP-dependent reaction releases the protein from the receptor and inserts it into a pore, another part of the TOM complex. The signal sequence then interacts with another complex, the TIM complex (translocation of inner membrane), in the inner membrane. The electrochemical gradient across the inner membrane pulls the signal sequence through. A mitochondrial Hsp70 binds to the protein as it becomes exposed within the mitochondrial matrix and, in another energy-dependent reaction, pulls the rest of the protein through. The signal sequence is removed by a specific protease (MPP, matrix processing peptidase) within the matrix. Note that this process pulls the protein through both outer and inner mitochondrial membranes.

A quite different process occurs in nuclear transport. Originally it was thought that these proteins simply diffused into the nucleus through the nuclear pores and were then bound to chromatin. However, it is becoming clear that the nuclear pores are complex gates, rather than open channels. Proteins destined for the nucleus contain *nuclear localization sequences* (NLS) that help these proteins select the nucleus

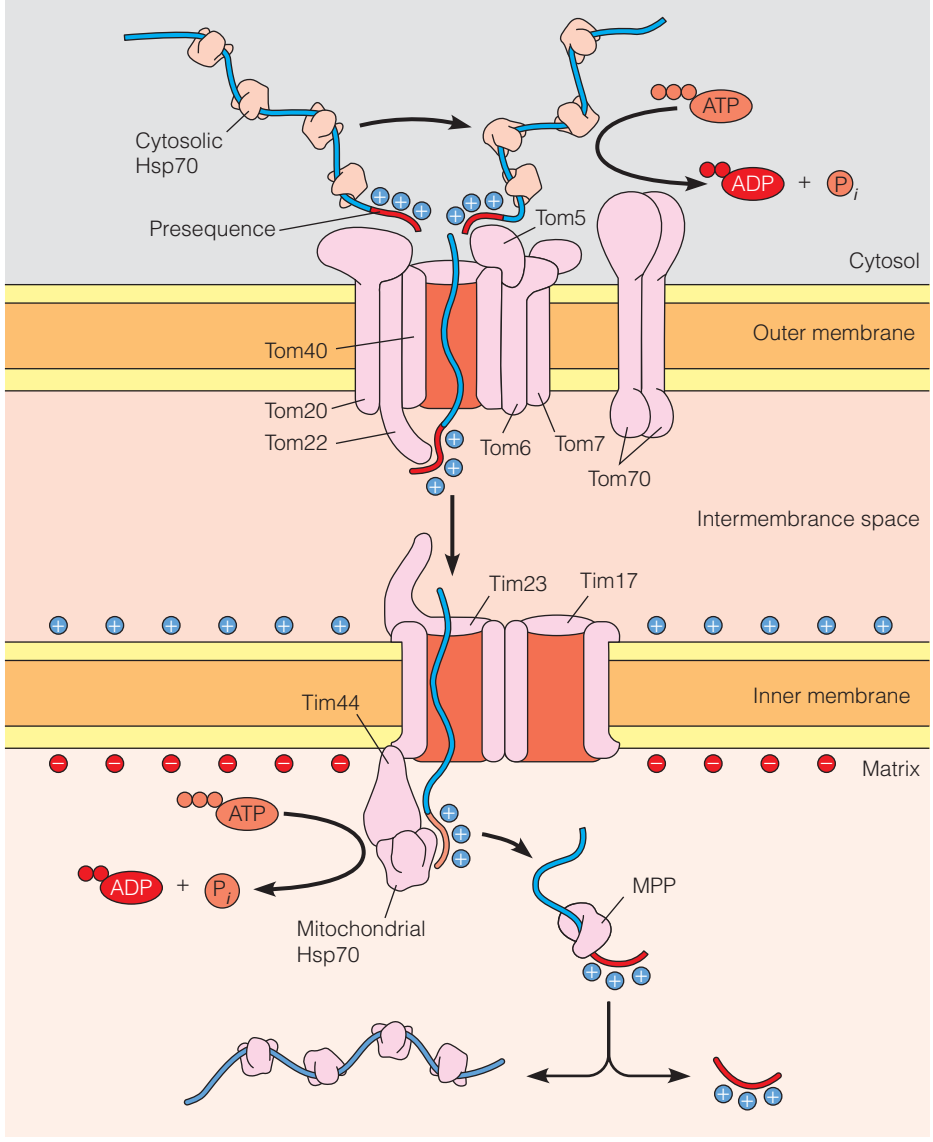


FIGURE 28.42

Transport of newly synthesized mitochondrial proteins into the matrix. Upper left, signal sequence of Hsp70-bound protein inserts into import receptor of the TOM complex (TOM20) in the outer membrane. Hsp70 dissociation is coupled to ATP hydrolysis. Insertion of the protein into the outer membrane (*via* TOM22) puts signal sequence in position to interact with TIM complex (TIM23) in the inner membrane. Potential across the inner membrane drives protein into intermembrane space. Signal sequence is cleaved off by MPP. Mitochondrial Hsp70 binds to protein in the matrix and uses the energy of ATP hydrolysis to pull the rest of the protein through.

Modified from *The Cell: A Molecular Approach*, 4th ed., G. M. Cooper and R. E. Hausmann (2007). American Society for Microbiology.

as their destination. Nuclear localization sequences can be found anywhere within the polypeptide sequence, not only the N-terminus. Moreover, the NLS is not removed as a consequence of transport. This is important because the nuclear membrane breaks down in each cell division cycle, and each nuclear protein must be re-transported into the nucleus after re-establishment of the nuclear envelope.

The nuclear localization signal on a prospective cargo protein interacts with a protein called **importin**, which carries the protein through the nuclear pore complex, as schematized in Figure 28.43. Energy for transport is provided by a monomeric G protein called **Ran** (Ras-related nuclear protein). Ran is similar to other G proteins we have encountered, in that the protein becomes activated by exchange of Ran-bound GDP for GTP, by a guanine nucleotide exchange factor (GEF), and inactivated by GTPase-activating protein (GAP), which hydrolyzes bound GTP to GDP. Ran passes the nuclear pore freely. Because GEF is localized to the nucleus and GAP to the cytoplasm, Ran-GTP predominates in the nucleus and Ran-GDP in the cytoplasm.

Proteins destined for the cytoplasm, nuclei, mitochondria, and chloroplasts are synthesized in the cytoplasm; those destined for organelles have specific targeting sequences.

Once the importin-cargo protein complex has passed into the nucleus, Ran-GTP binds to that complex, which displaces the cargo. The Ran-GTP complex is returned to the cytoplasm, where bound GTP is converted to GDP. Ran-GDP returns to the nucleus, where it exchanges GDP for GTP, and importin seeks a new NLS-containing cargo protein.

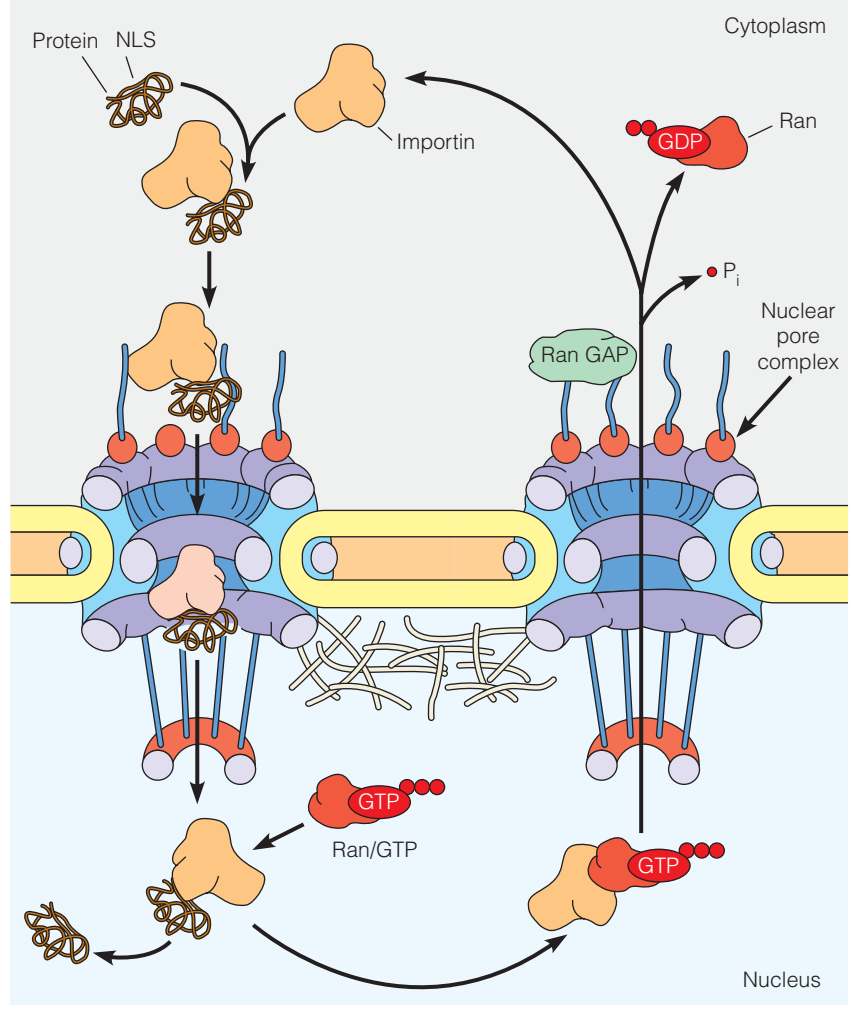


FIGURE 28.43

A schematic view of delivery of a protein, synthesized in cytoplasm, into the nucleus. A protein with a nuclear localization signal (NLS) binds to importin. The complex binds to a nuclear pore and passes through. Within the nucleus Ran-GTP binds to the importin-cargo complex and displaces the cargo. The resultant Ran-importin complex is returned to the cytoplasm, where Ran-bound GTP is hydrolyzed to GDP. Ran-GDP returns to the nucleus, and its bound GDP is exchanged for GTP (not shown).

Modified from *The Cell: A Molecular Approach*, 4th ed., G. M. Cooper and R. E. Hausmann (2007). American Society for Microbiology.

Proteins Synthesized on the Rough Endoplasmic Reticulum

Proteins destined for cellular membranes, lysosomes, or extracellular transport use a quite different distribution system. The key structures in this system are the **rough endoplasmic reticulum** (RER) and the Golgi complex (see also Chapter 9). The rough endoplasmic reticulum is a network of membrane-enclosed spaces within the cytoplasm. The RER membrane is heavily coated on the outer, cytosolic surface with polyribosomes; this coating is what gives the membrane its rough appearance. The Golgi complex resembles the RER in that it is a stack of thin, membrane-bound sacs. However, the Golgi sacs are not interconnected, nor do they carry polyribosomes on their surfaces. The role of the Golgi complex is to act as a "switching center" for proteins with various destinations.

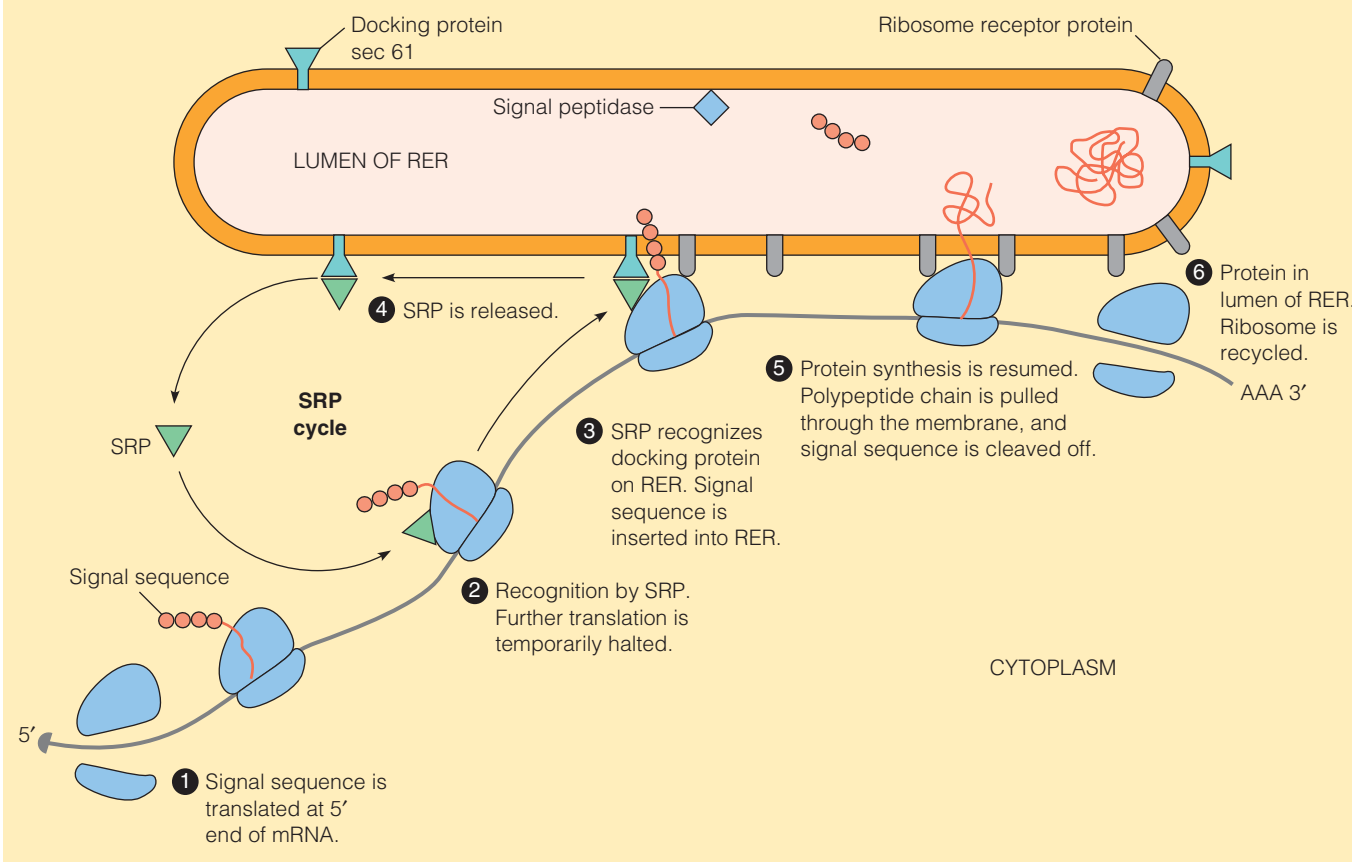
Proteins destined for cell membranes, lysosomes, or export are synthesized on the rough endoplasmic reticulum, then modified and transported via the Golgi apparatus.

Proteins that are to be directed to their destinations via the Golgi complex are synthesized by polyribosomes associated with the RER. Synthesis actually begins in the cytoplasm (Figure 28.44, step 1). The first sequence to be synthesized is an N-terminal *signal sequence,* part of a mechanism for attaching the ribosome and nascent protein to the RER. *Signal recognition particles (SRPs),* containing several proteins and a small (7S) RNA, recognize the signal sequences of the appropriate nascent proteins and bind to them as they are being extruded from the ribosomes (step 2).

The SRP has two functions. First, its binding temporarily halts translation so that no more than the N-terminal signal sequence extends from the ribosome. This pause prevents completion of the protein in the wrong place—that is, in the cytosol—and also inhibits premature folding of the polypeptide chain. Thus, the SRP is acting as a kind of chaperone. The second function of the SRP is to recognize a docking protein in the RER membrane. This is the trimeric Sec61 complex, homologous to the bacterial SecYEG. The docking protein binds the ribosome to the RER, and the signal sequence is inserted into the RER membrane (step 3). The SRP is then released (step 4), allowing translation to resume (step 5). The protein

FIGURE 28.44

The sequence of events in synthesis of proteins on the rough endoplasmic reticulum. The time sequence of events is from left to right. Recent cryo-electron microscopic analysis of ribosome-Sec61 complexes shows that the Sec61 functions as a monomer. Model building allowed investigators to trace the path of newly synthesized protein from the tRNA through the ribosomal tunnel through the monomeric Sec61 embedded in the membrane, as shown in Figure 28.45.



being synthesized is actually *pulled* through the membrane by an ATP-dependent process. Before translation is complete, signal sequences are cleaved from some proteins by an RER-associated protease. These proteins are released into the lumen of the RER and further transported (step 6). Proteins that will remain in the endoplasmic reticulum have resistant signal peptides and thereby remain anchored to the RER membrane. A model of the protein translocation process, based upon the structure of Sec61, is shown in Figure 28.45.

Role of the Golgi Complex

The proteins that enter the lumen of the RER undergo the first stages of glycosylation at this point. Vesicles carrying these proteins then bud off the RER and move to the Golgi complex (Figure 28.46). Here the carbohydrate moieties of glycoproteins are completed (see pages 339–343 in Chapter 9 for details), and a final

FIGURE 28.45

Schematic representation of an actively translating and translocating eukaryotic ribosome-Sec61 complex. NC, nascent chain; PCC, protein-conducting channel (Sec61). P-tRNA, peptidyl-tRNA with its nascent chain. PDB ID 2ww9, 2wwa, and 2wwb.

From *Science* 326:1369–1372, T. Becker, S. Bhushan, A. Jarasch, J.-P. Armache, S. Funes, F. Jossinet, J. Gumbart, T. Mielke, O. Berninghausen, K. Schulten, E. Westhof, R. Gilmore, E. C. Mandon, and R. Beckmann, Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome.

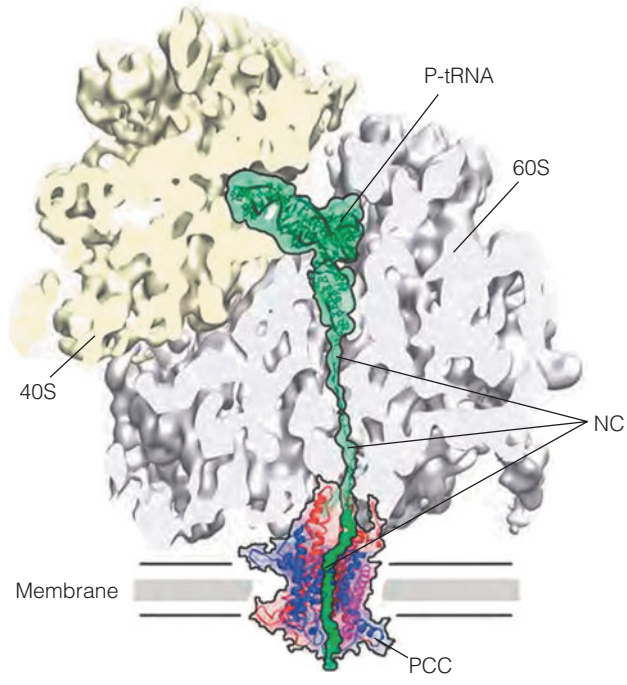
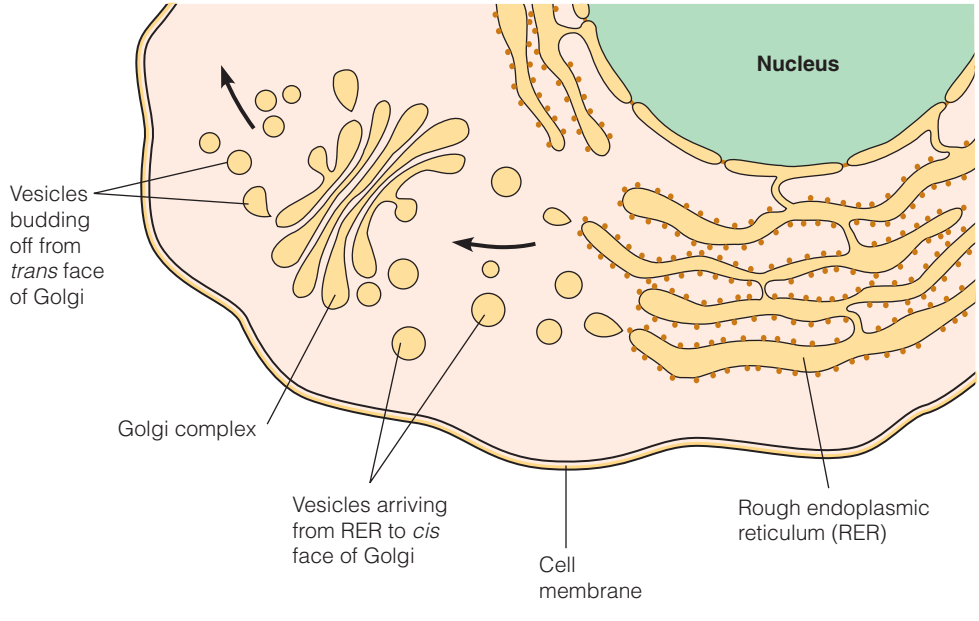


FIGURE 28.46

Transfer from the rough endoplasmic reticulum (RER) to the Golgi complex. Note that vesicles bud off the RER and move to the *cis* face of the Golgi. Primary lysosomal vesicles bud from the *trans* portion of the Golgi.



sorting occurs. The multiple membrane sacs that constitute the Golgi complex represent a multilayer arena for these processes. Vesicles from the RER enter at the *cis* face of the Golgi (that closest to the RER) and fuse with the Golgi membrane. Proteins are then passed, again via vesicles, to the intermediate layers. Finally, vesicles bud off from the *trans* face of the Golgi complex to form lysosomes, peroxisomes, or glyoxysomes or to travel to the plasma membrane. All of this transport of vesicles, from the RER to the *cis* face of the Golgi, to successive levels of the Golgi and on to their final destinations, requires high specificity in targeting. Transport of vesicles to the wrong destinations would cause cellular chaos. This sorting is accomplished by having each kind of protein cargo packed in a vesicle marked by specific vesicle membrane proteins. In some cases, the target membranes contain complementary proteins that interact with these and cause membrane fusion. These complementary pairs are called *SNARES* (soluble N-ethylmaleimide-sensitive factor attachment protein receptors)—v-SNARES on vesicles, t-SNARES on target membranes. The interaction of specific v- and t-SNARES, aided by cytosolic fusion proteins, leads to fusion of the vesicle and target membranes and delivery of the cargo (see Figure 28.47).

The Fate of Proteins: Programmed Destruction

In Chapter 11 we pointed out that one mechanism for control of enzymatic function is the selective degradation of certain enzymes. However, not only enzymes need to be destroyed in a programmed way. Regulatory proteins that are essential in certain parts of the cell cycle and deleterious in others must be eliminated at some point. Consider the cyclins (Chapter 24), for example, which must be broken down and resynthesized during each cell cycle. Proteins that have become damaged must also be removed. In some developmental processes, it is necessary to remove whole organelles or even entire cells and tissues.

Eukaryotic cells have two distinct methods for protein degradation. The lysosomes contain among their hydrolases proteolytic enzymes that will degrade any protein trapped within the organelle. Parallel to this process is a cytosolic degradation system, which is of necessity highly selective. The danger inherent in having nonspecific proteases loose in the cytosol should be evident. Both of these processes were described briefly in Chapter 20, and we supplement that information here.

The Lysosomal System

The lysosomal particles budded from the Golgi complex, known as **primary lysosomes**, are essentially bags of degradative enzymes. Over 50 different hydrolytic enzymes are contained in lysosomes, including proteases, nucleases, lipases, and carbohydrate-cleaving enzymes. The lysosomes play a number of important roles in cellular metabolism, as schematically depicted in Figure 28.48.

In some cell types, such as those in the pancreas that secrete degradative enzymes, primary lysosomes migrate to the cell surface and release their contents into the exterior medium (path A). Primary lysosomes may also fuse with *autophagic vesicles,* formed when smooth ER engulfs organelles destined for destruction (path B). The combined vesicle is called an *autophagic lysosome.* In some kinds of cells—mainly certain white blood cells—primary lysosomes may fuse with *phagocytic vacuoles* that have engulfed nutrient materials at the cell surface (path C). In these heterophagic lysosomes, the nutrients are digested and their amino acids, nucleotides, lipids, and other low-molecular-weight constituents released into the cytosol. Residual, undigested material is excreted when the heterophagic lysosomes and autophagic lysosomes find their way to the plasma membrane.

Cytosolic Protein Degradation

In contrast to the lysosomal enzymes, which are usually safely sequestered in their vesicles, any protease activity that is free in normal cytosol must be under rigid control. It must attack only the proteins whose destruction is needed. These may include