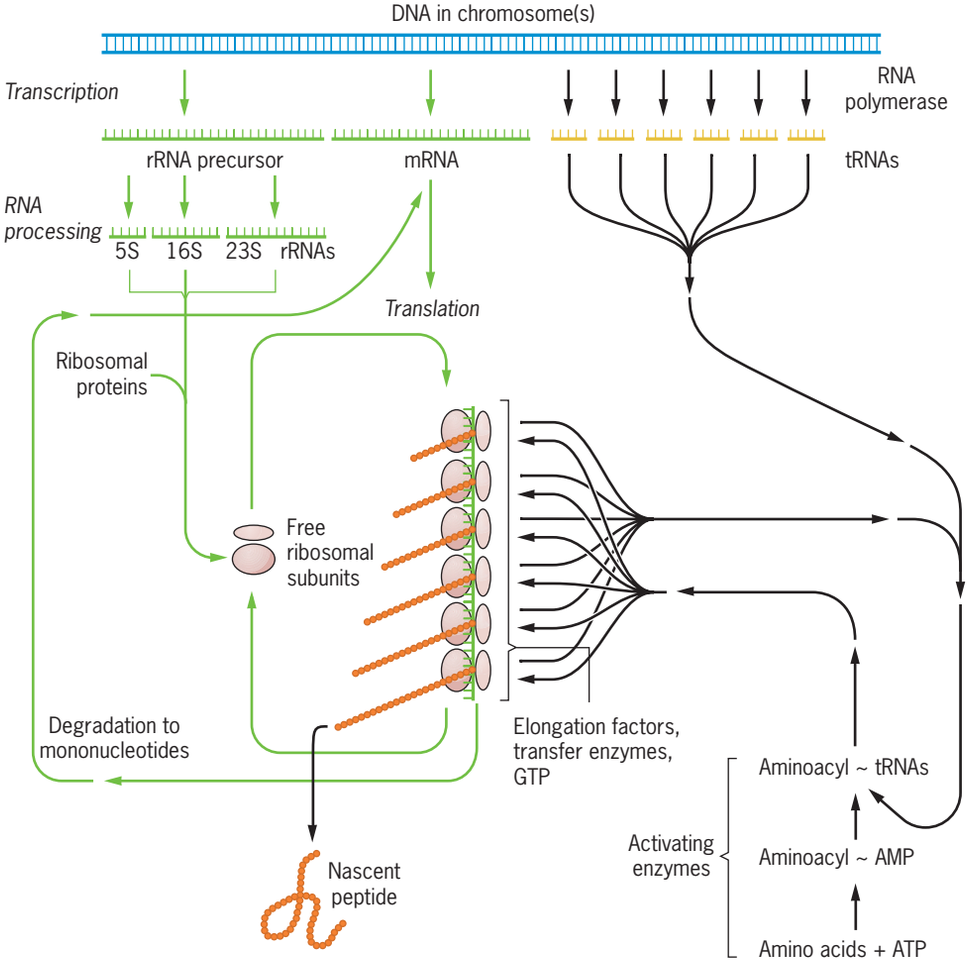
Protein Synthesis: Translation

The genetic information in mRNA molecules is translated into the amino acid sequences of polypeptides according to the specifications of the genetic code.

The process by which the genetic information stored in the sequence of nucleotides in an mRNA is translated, according to the specifications of the genetic code, into the sequence of amino acids in the polypeptide gene product is complex, requiring the functions of a large number of macromolecules. These include (1) over 50 polypeptides and three to five RNA molecules present in each ribosome (the exact composition varies from species to species), (2) at least 20 amino acid-activating enzymes, (3) 40 to 60 different tRNA molecules, and (4) numerous soluble proteins involved in polypeptide chain initiation, elongation, and termination. Because many of these macromolecules, particularly the components of the ribosome, are present in large quantities in each cell, the translation system makes up a major portion of the metabolic machinery of each cell.

OVERVIEW OF PROTEIN SYNTHESIS

Before focusing on the details of the translation process, we should preview the process of protein synthesis in its entirety. An overview of protein synthesis, illustrating its complexity and the major macromolecules involved, is presented in ■ **Figure 12.8**. The first step in gene expression, transcription, involves the transfer of information stored in genes to messenger RNA (mRNA) intermediaries, which carry that information to the sites of polypeptide synthesis in the cytoplasm. Transcription is discussed in detail in Chapter 11. The second step, translation, involves the transfer of the information in mRNA molecules into the sequences of amino acids in polypeptide gene products.

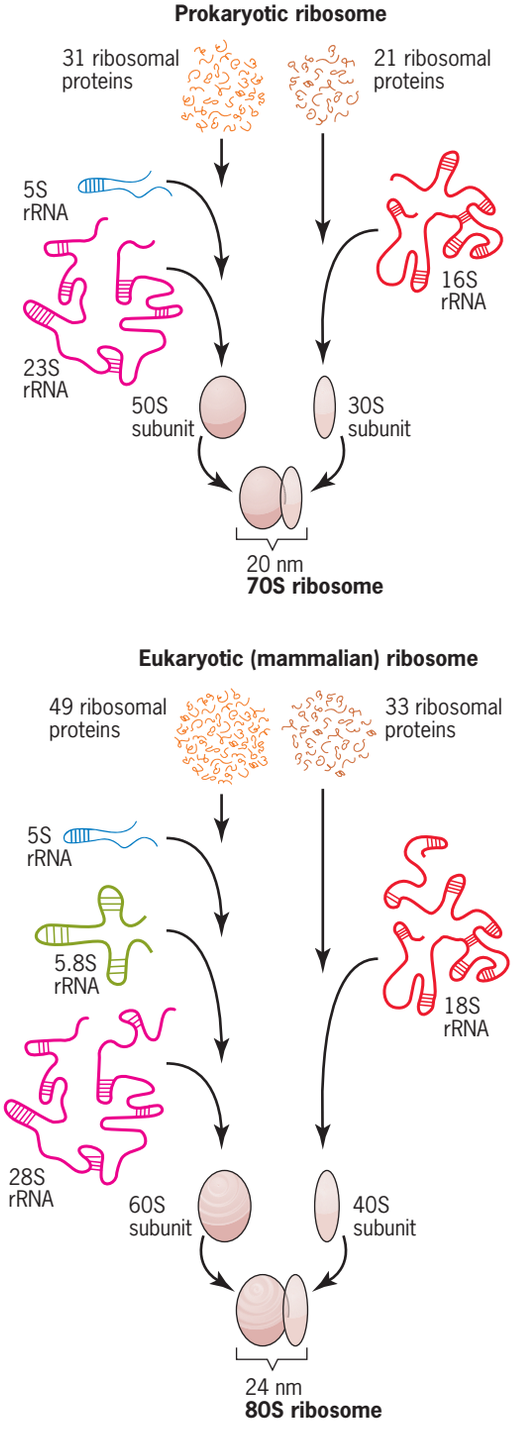


■ **FIGURE 12.8** Overview of protein synthesis. The sizes of the rRNA molecules shown are correct for bacteria; larger rRNAs are present in eukaryotes. For simplicity, all RNA species have been transcribed from contiguous segments of a single DNA molecule. In reality, the various RNAs are transcripts of genes located at different positions on from one to many chromosomes. Details of the various stages of protein synthesis are discussed in subsequent sections of this chapter.

Translation occurs on ribosomes, which are complex macromolecular structures located in the cytoplasm. Translation involves three types of RNA, all of which are transcribed from DNA templates (chromosomal genes). In addition to mRNAs, three to five RNA molecules (rRNA molecules) are present as part of the structure of each ribosome, and 40 to 60 small RNA molecules (tRNA molecules) function as adaptors by mediating the incorporation of the proper amino acids into polypeptides in response to specific nucleotide sequences in mRNAs. The amino acids are attached to the correct tRNA molecules by a set of activating enzymes called **aminoacyl-tRNA synthetases.**

The nucleotide sequence of an mRNA molecule is translated into the appropriate amino acid sequence according to the dictations of the genetic code. Some nascent polypeptides contain short amino acid sequences at the amino or carboxyl termini that function as signals for their transport into specific cellular compartments such as the endoplasmic reticulum, mitochondria, chloroplasts, or nuclei. Nascent secretory proteins, for example, contain a short *signal sequence* at the amino terminus that directs the emerging polypeptide to the membranes of the endoplasmic reticulum. Similar targeting sequences are present at the amino termini of proteins destined for import into mitochondria and chloroplasts. Some nuclear proteins contain targeting extensions at the carboxyl termini. In many cases, the targeting peptides are removed enzymatically by specific peptidases after transport of the protein into the appropriate cellular compartment.

The ribosomes may be thought of as workbenches, complete with machines and tools needed to make a polypeptide. They are nonspecific in the sense that they can synthesize any polypeptide (any amino acid sequence) encoded by a particular mRNA molecule, even an mRNA from a different species. Each mRNA molecule is simultaneously translated by several ribosomes, resulting in the formation of a polyribosome, or polysome. Given this brief overview of protein synthesis, we will now examine some of the more important components of the translation machinery more closely.



■ **FIGURE 12.9** Macromolecular composition of prokaryotic and eukaryotic ribosomes.

COMPONENTS REQUIRED FOR PROTEIN SYNTHESIS: RIBOSOMES

Living cells devote more energy to the synthesis of proteins than to any other aspect of metabolism. About one-third of the total dry mass of most cells consists of molecules that participate directly in the biosynthesis of proteins. In *E. coli*, the approximately 200,000 ribosomes account for 25 percent of the dry weight of each cell. This commitment of a major proportion of the metabolic machinery of cells to the process of protein synthesis documents its importance in the life forms that exist on our planet.

When the sites of protein synthesis were labeled in cells grown for short intervals in the presence of radioactive amino acids and were visualized by autoradiography, the results showed that proteins are synthesized on the ribosomes. In prokaryotes, ribosomes are distributed throughout cells; in eukaryotes, they are located in the cytoplasm, frequently on the extensive intracellular membrane network of the endoplasmic reticulum.

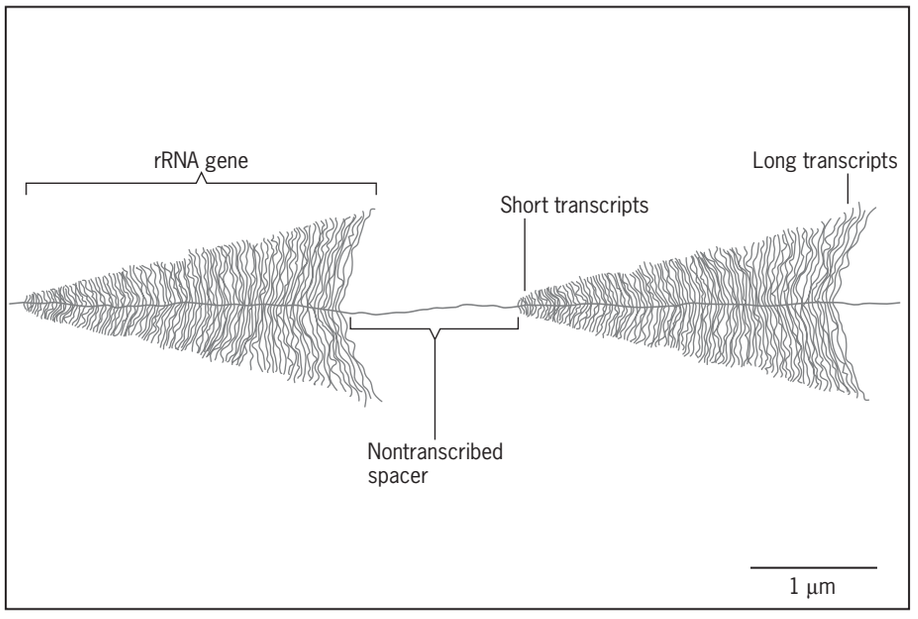
Ribosomes are approximately half protein and half RNA (■ **Figure 12.9**). They are composed of two subunits, one large and one small, which dissociate when the translation of an mRNA molecule is completed and reassociate during the initiation of translation. Each subunit contains a large, folded RNA molecule on which the ribosomal proteins assemble. Ribosome sizes are most frequently expressed in terms of their rates of sedimentation during centrifugation, in Svedberg (S) units. [One Svedberg unit is equal to a sedimentation coefficient (velocity/centrifugal force) of 10^{-13} seconds.] The *E. coli* ribosome, like the ribosomes of other prokaryotes, has a molecular weight of 2.5×10^6, a size of 70S, and dimensions of about 20 nm × 25 nm. The ribosomes of eukaryotes are larger (usually about 80S); however, size varies from species to species. The ribosomes present in the mitochondria and chloroplasts of eukaryotic cells are smaller (usually about 60S).

Although the size and macromolecular composition of ribosomes vary, the overall three-dimensional structure of the ribosome is basically the same in all organisms. In

E. coli, the small (30S) ribosomal subunit contains a 16S (molecular weight about 6×10^5) RNA molecule plus 21 different polypeptides, and the large (50S) subunit contains two RNA molecules (5S, molecular weight about 4×10^4, and 23S, molecular weight about 1.2×10^6) plus 31 polypeptides. In mammalian ribosomes, the small subunit contains an 18S RNA molecule plus 33 polypeptides, and the large subunit contains three RNA molecules of sizes 5S, 5.8S, and 28S plus 49 polypeptides. In organelles, the corresponding rRNA sizes are 5S, 13S, and 21S.

Masayasu Nomura and his colleagues were able to disassemble the 30S ribosomal subunit of *E. coli* into the individual macromolecules and then reconstitute functional 30S subunits from the components. In this way, they studied the functions of individual rRNA and ribosomal protein molecules.

The ribosomal RNA molecules, like mRNA molecules, are transcribed from a DNA template. In eukaryotes, rRNA synthesis occurs in the nucleolus (see Figure 2.1) and is catalyzed by RNA polymerase I. The nucleolus is a highly specialized component of the nucleus devoted exclusively to the synthesis of rRNAs and their assembly into ribosomes. The ribosomal RNA genes are present in tandemly duplicated arrays separated by intergenic spacer regions. The transcription of these tandem sets of rRNA genes can be visualized directly by electron microscopy. (■ **Figure 12.10**) shows a schematic diagram of the observed transcription.

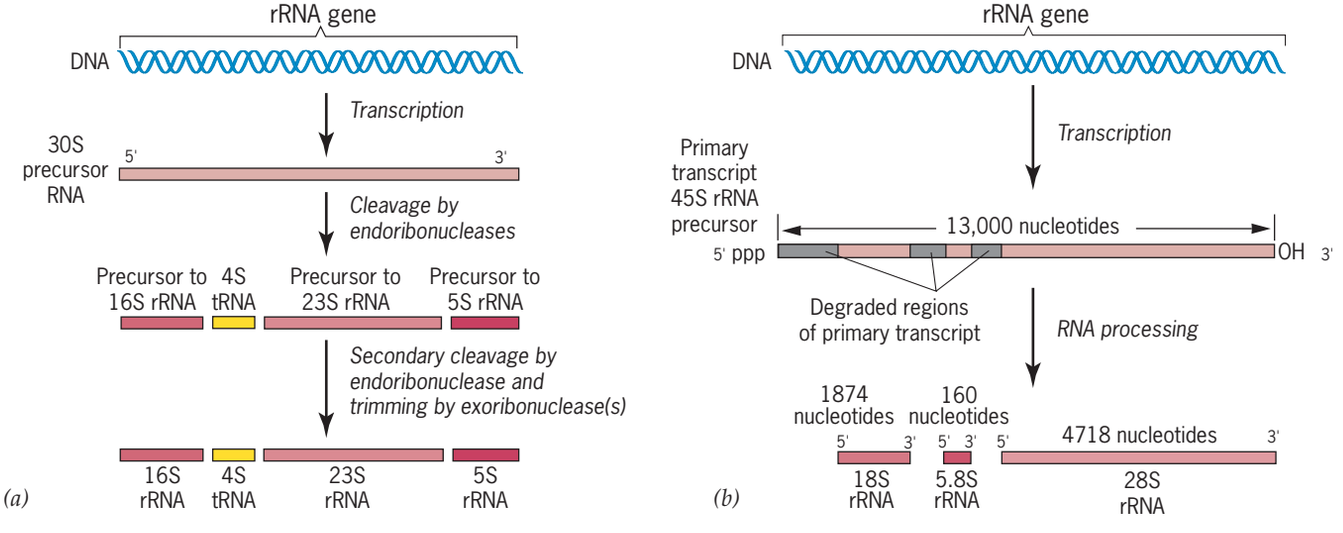


■ **FIGURE 12.10** Schematic diagram of an electron micrograph showing the transcription of tandemly repeated rRNA genes in the nucleolus of the new *Triturus viridescens*. A gradient of fibrils of increasing length is observed for each rRNA gene, and nontranscribed spacer regions separate the genes.

The transcription of the rRNA genes produces RNA precursors that are much larger than the RNA molecules found in ribosomes. These rRNA precursors undergo posttranscriptional processing to produce the mature rRNA molecules. In *E. coli*, the rRNA gene transcript is a 30S precursor, which undergoes endonucleolytic cleavages to produce the 5S, 16S, and 23S rRNAs plus one 4S transfer RNA molecule (■ **Figure 12.11*a***). In mammals, the 5.8S, 18S, and 28S rRNAs are cleaved from a 45S precursor (■ **Figure 12.11*b***), whereas the 5S rRNA is produced by posttranscriptional processing of a separate gene transcript. In addition to the posttranscriptional cleavages of rRNA precursors, many of the nucleotides in rRNAs are posttranscriptionally methylated. The methylation is thought to protect rRNA molecules from degradation by ribonucleases.

Multiple copies of the genes for rRNA are present in the genomes of all organisms that have been studied to date. This redundancy of rRNA genes is not surprising considering the large number of ribosomes present per cell. In *E. coli*, seven rRNA genes

■ **FIGURE 12.11** Synthesis and processing of (*a*) the 30S rRNA precursor in *E. coli* and (*b*) the 45S rRNA precursor in mammals.

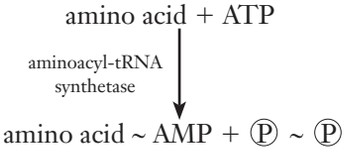


(*rrnA—rrnE*, *rrnG*, *rrnH*) are distributed among three distinct sites on the chromosome. In eukaryotes, the rRNA genes are present in hundreds to thousands of copies. The 5.8S-18S-28S rRNA genes of eukaryotes are present in tandem arrays in the **nucleolar organizer regions** of the chromosomes. In some eukaryotes, such as maize, there is a single pair of nucleolar organizers (on chromosome 6 in maize). In *Drosophila* and the South African clawed toad, *Xenopus laevis*, the sex chromosomes carry the nucleolar organizers. Humans have five pairs of nucleolar organizers located on the short arms of chromosomes 13, 14, 15, 21, and 22. The 5S rRNA genes in eukaryotes are not located in the nucleolar organizer regions. Instead, they are distributed over several chromosomes. However, the 5S rRNA genes are highly redundant, just as are the 5.8S-18S-28S rRNA genes.

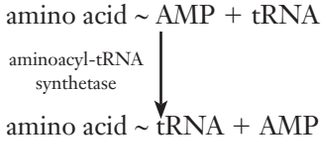
COMPONENTS REQUIRED FOR PROTEIN SYNTHESIS: TRANSFER RNAs

Although the ribosomes provide many of the components required for protein synthesis, and the specifications for each polypeptide are encoded in an mRNA molecule, the translation of a coded mRNA message into a sequence of amino acids in a polypeptide requires one additional class of RNA molecules, the transfer RNA (tRNA) molecules. Chemical considerations suggested that direct interactions between the amino acids and the nucleotide triplets or codons in mRNA were unlikely. Thus, in 1958, Francis Crick proposed that some kind of an adaptor molecule must mediate the specification of amino acids by codons in mRNAs during protein synthesis. The adaptor molecules were soon identified by other researchers and shown to be small (4S, 70–95 nucleotides long) RNA molecules. These molecules, first called soluble RNA (sRNA) molecules and subsequently transfer RNA (tRNA) molecules, contain a triplet nucleotide sequence, the anticodon, which is complementary to and base-pairs with the codon sequence in mRNA during translation. There are one to four tRNAs for each of the 20 amino acids.

The amino acids are attached to the tRNAs by high-energy (very reactive) bonds (symbolized ~) between the carboxyl groups of the amino acids and the 3′-hydroxyl termini of the tRNAs. The tRNAs are activated or charged with amino acids in a two-step process, with both reactions catalyzed by the same enzyme, aminoacyl-tRNA synthetase. There is at least one aminoacyl-tRNA synthetase for each of the 20 amino acids. The first step in aminoacyl-tRNA synthesis involves the activation of the amino acid using energy from adenosine triphosphate (ATP):



The amino acid~AMP intermediate is not normally released from the enzyme before undergoing the second step in aminoacyl-tRNA synthesis, namely, the reaction with the appropriate tRNA:



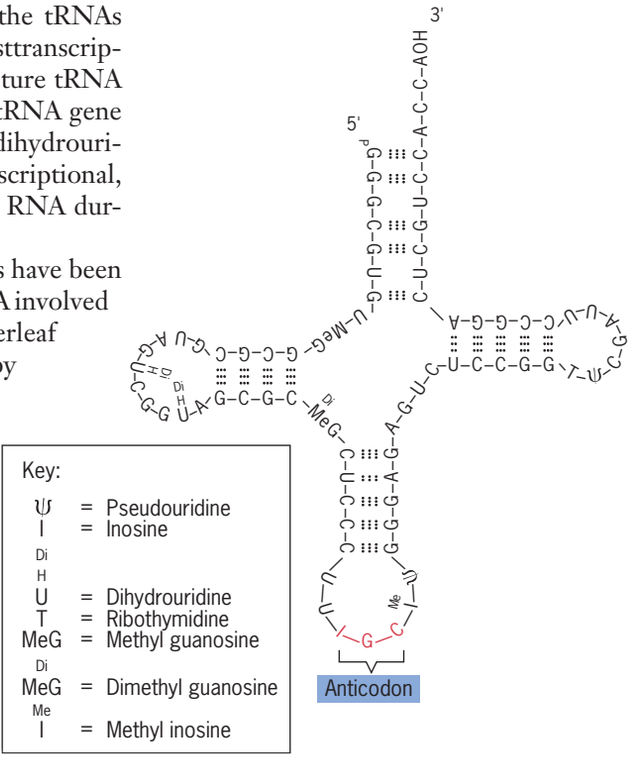
The aminoacyl~tRNAs are the substrates for polypeptide synthesis on ribosomes, with each activated tRNA recognizing the correct mRNA codon and presenting the amino acid in a steric configuration (three-dimensional structure) that facilitates peptide bond formation.

The tRNAs are transcribed from genes. As in the case of rRNAs, the tRNAs are transcribed in the form of larger precursor molecules that undergo posttranscriptional processing (cleavage, trimming, methylation, and so forth). The mature tRNA molecules contain several nucleosides that are not present in the primary tRNA gene transcripts. These unusual nucleosides, such as inosine, pseudouridine, dihydrouridine, 1-methyl guanosine, and several others, are produced by posttranscriptional, enzyme-catalyzed modifications of the four nucleosides incorporated into RNA during transcription.

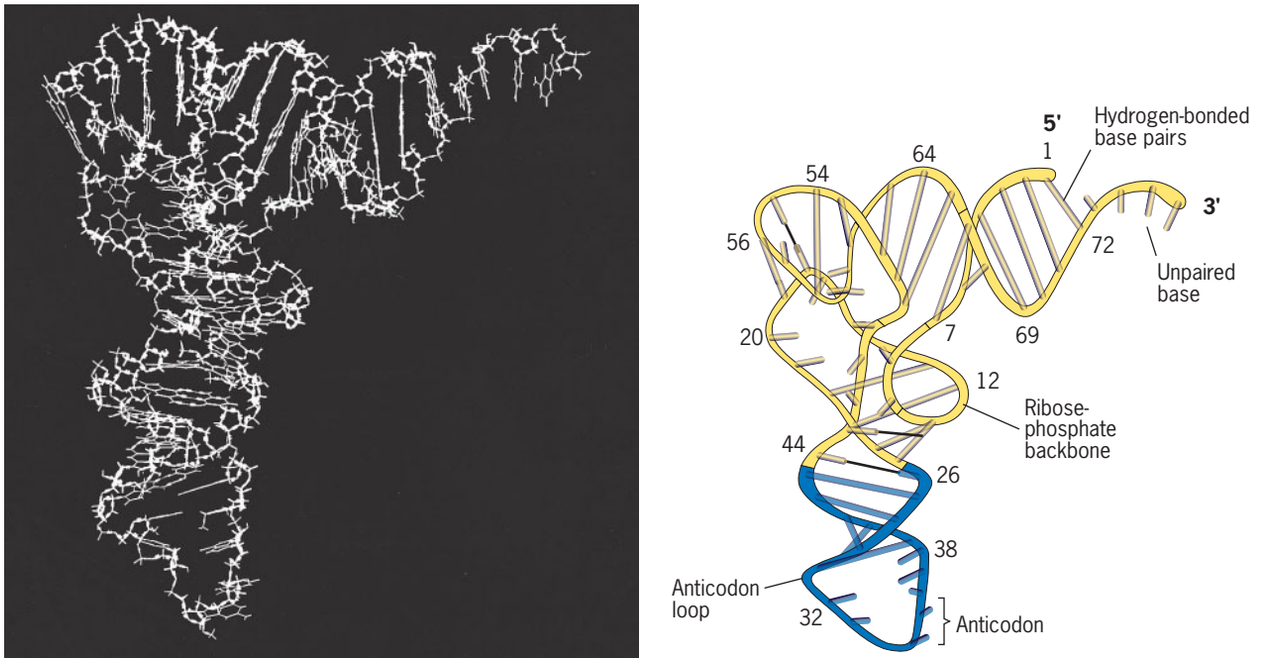
Because of their small size (most are 70 to 95 nucleotides long), tRNAs have been more amenable to structural analysis than the other, larger molecules of RNA involved in protein synthesis. The complete nucleotide sequence and proposed cloverleaf structure of the alanine tRNA of yeast (■ **Figure 12.12**) were published by Robert W. Holley and colleagues in 1965; Holley shared the 1968 Nobel Prize in Physiology or Medicine for this work. The three-dimensional structure of the phenylalanine tRNA of yeast was determined by X-ray diffraction studies in 1974 (■ **Figure 12.13**). The anticodon of each tRNA occurs within a loop (nonhydrogen-bonded region) near the middle of the molecule.

It should be apparent that tRNA molecules must contain a great deal of specificity despite their small size. Not only must they (1) have the correct anticodon sequences, so as to respond to the right codons, but they also must (2) be recognized by the correct aminoacyl-tRNA synthetases, so that they are activated with the correct amino acids, and (3) bind to the appropriate sites on the ribosomes to carry out their adaptor functions.

There are three tRNA binding sites on each ribosome (■ **Figure 12.14*a–b***). The ***A*** or **aminoacyl site** binds the incoming aminoacyl-tRNA, the tRNA carrying the next amino acid to be added to the growing polypeptide chain. The ***P*** or **peptidyl site** binds the tRNA to which the growing polypeptide is attached. The ***E*** or **exit site** binds the departing uncharged tRNA.

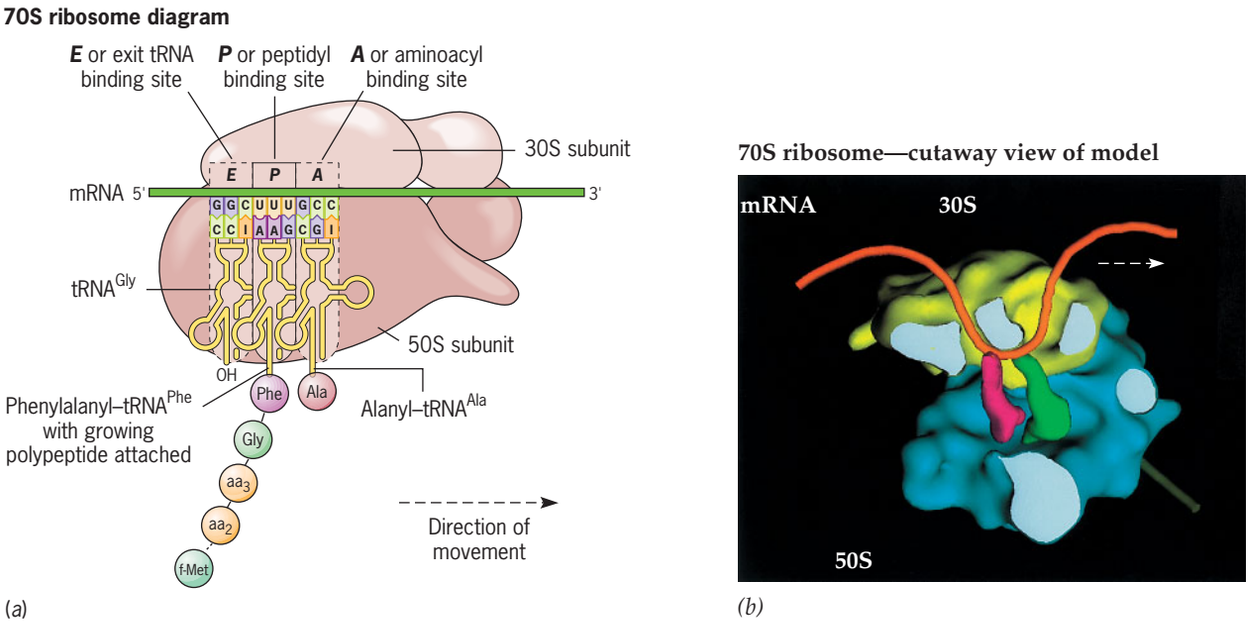


■ **FIGURE 12.12** Nucleotide sequence and cloverleaf configuration of the alanine tRNA of *S. cerevisiae*. The names of the modified nucleosides present in the tRNA are shown in the inset.



(a) *(b)*

■ **FIGURE 12.13** Photograph (*a*) and interpretative drawing (*b*) of a molecular model of the yeast phenylalanine tRNA based on X-ray diffraction data.



■ **FIGURE 12.14** Ribosome structure in *E. coli*. (*a*) Each ribosome/mRNA complex contains three aminoacyl-tRNA binding sites. The *A* or aminoacyl-tRNA site is occupied by alanyl-tRNA^{Ala}. The *P* or peptidyl site is occupied by phenylalanyl-tRNA^{Phe}, with the growing polypeptide chain covalently linked to the phenylalanine tRNA. The *E* or exit site is occupied by tRNA^{Gly} prior to its release from the ribosome. (*b*) An mRNA molecule (orange), which is attached to the 30S subunit (light green) of the ribosome, contributes specificity to the tRNA-binding sites, which are located largely on the 50S subunit (blue) of the ribosome. The aminoacyl-tRNAs located in the *P* and *A* sites are shown in red and dark green, respectively. The *E* site is unoccupied.

The three-dimensional structure of the 70S ribosome of the bacterium *Thermus thermophilus* has been solved with resolution to 0.55 nm by X-ray crystallography (■ Figure 12.15*a–c*). The crystal structure shows the positions of the three tRNA binding sites at the 50S–30S interface and the relative positions of the rRNAs and ribosomal proteins.

Although the aminoacyl-tRNA binding sites are located largely on the 50S subunit and the mRNA molecule is bound by the 30S subunit, the specificity for aminoacyl-tRNA binding in each site is provided by the mRNA codon that makes up part of the binding site (see Figure 12.14*b*). As the ribosome moves along an mRNA (or as the mRNA is shuttled across the ribosome), the specificity for the aminoacyl-tRNA binding in the *A*, *P*, and *E* sites changes as different mRNA codons move into register in the binding sites. The ribosomal binding sites by themselves (minus mRNA) are thus capable of binding any aminoacyl-tRNA.

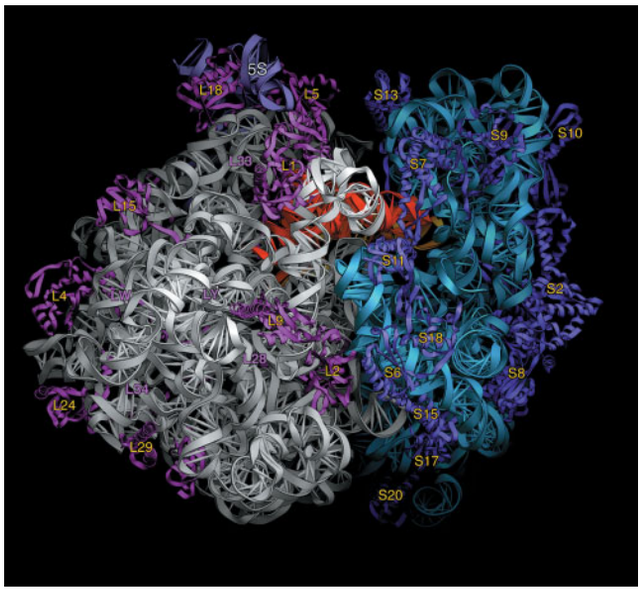
TRANSLATION: THE SYNTHESIS OF POLYPEPTIDES USING mRNA TEMPLATES

We now have reviewed all the major components of the protein-synthesizing system. The mRNA molecules provide the specifications for the amino acid sequences of the polypeptide gene products. The ribosomes provide many of the macromolecular components required for the translation process. The tRNAs provide the adaptor molecules needed to incorporate amino acids into polypeptides in response to codons in mRNAs. In addition, several soluble proteins participate in the process. The translation of the sequence of nucleotides in an mRNA molecule into the sequence of amino acids in its polypeptide product can be divided into three stages: (1) polypeptide chain initiation, (2) chain elongation, and (3) chain termination.

Translation: Polypeptide Chain Initiation

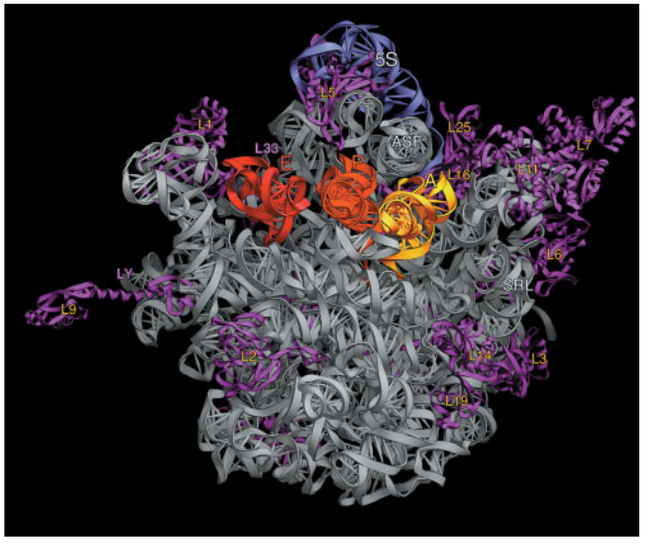
The **initiation** of translation includes all events that precede the formation of a peptide bond between the first two amino acids of the new polypeptide chain. Although several aspects of the initiation process are the same in prokaryotes and eukaryotes, some are different. Accordingly, we will first examine the initiation of polypeptide chains in *E. coli*, and we will then look at the unique aspects of translational initiation in eukaryotes.

70S ribosome—crystal structure



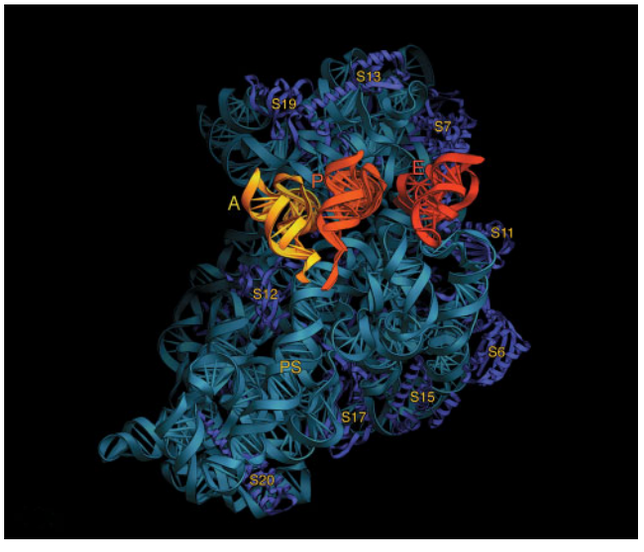
(a)

50S subunit—crystal structure



(b)

30S subunit—crystal structure

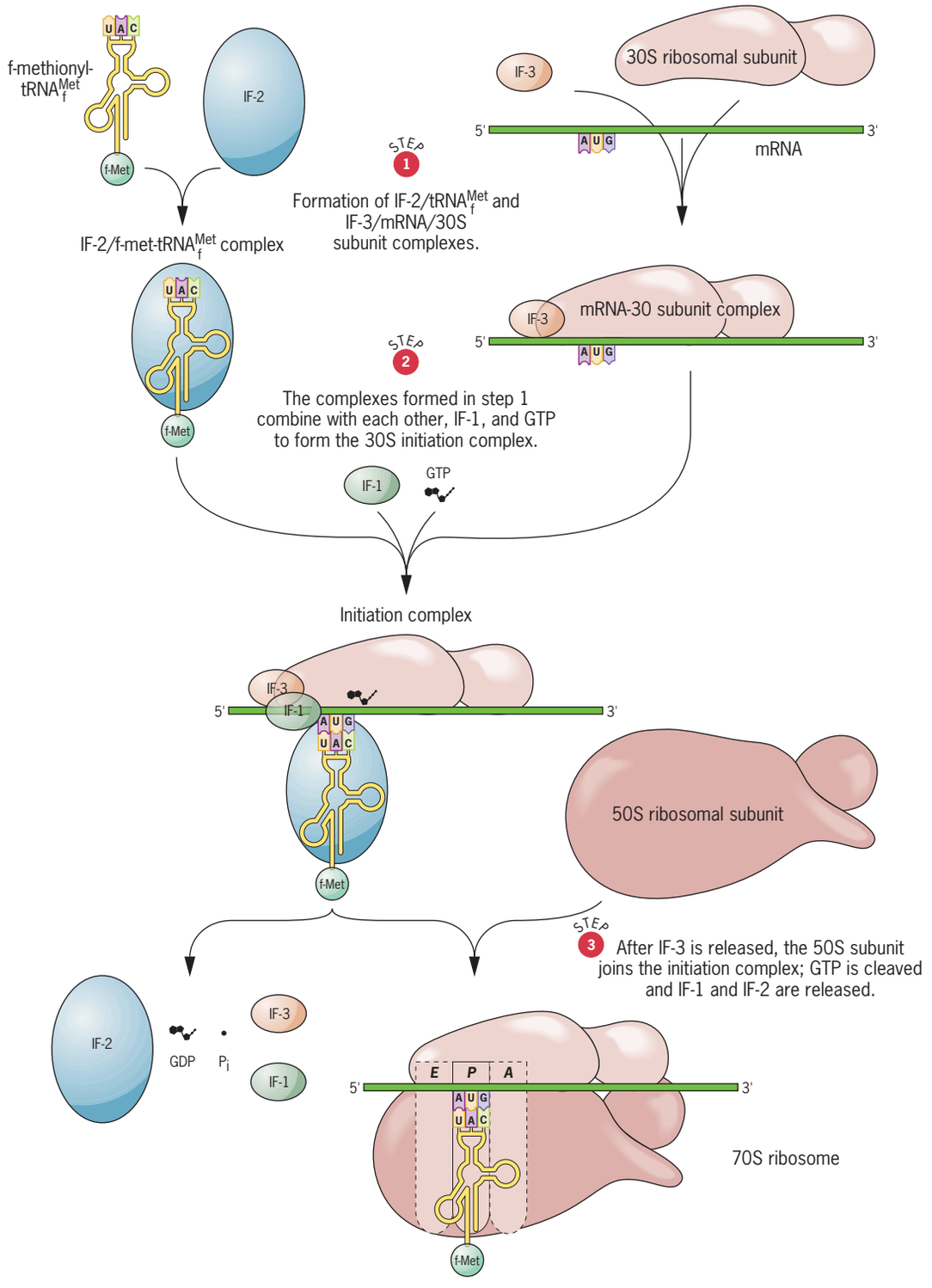


(c)

■ **FIGURE 12.15** Ribosome structure in *Thermus thermophilus*. Crystal structure of the 70S ribosome with 0.55 nm resolution, showing the complete ribosome (*a*) and the interfaces of the 50S (*b*) and 30S (*c*) subunits. (*a*) 50S subunit on the left; 30S subunit on the right. (*b*, *c*) Interfaces of the 50S subunit and the 30S subunit obtained by rotating the structures shown in (*a*) 90° to the left (*b*) or to the right (*c*), respectively. The tRNAs in the *A*, *P*, and *E* sites are shown in gold, orange, and red, respectively. Components: 16S rRNA (cyan); 23S rRNA (gray); 5S rRNA (light blue); 30S subunit proteins (dark blue); and 50S subunit proteins (magenta). L1, large subunit protein 1; S7, small subunit protein 7.

In *E. coli*, the initiation process involves the 30S subunit of the ribosome, a special initiator tRNA, an mRNA molecule, three soluble protein **initiation factors: IF-1, IF-2,** and **IF-3,** and one molecule of GTP (■ Figure 12.16). Translation occurs on 70S ribosomes, but the ribosomes dissociate into their 30S and 50S subunits each time they complete the synthesis of a polypeptide chain. In the first stage of the initiation of translation, a free 30S subunit interacts with an mRNA molecule and the initiation factors. The 50S subunit joins the complex to form the 70S ribosome in the final step of the initiation process.

The synthesis of polypeptides is initiated by a special tRNA, designated **$tRNA_f^{Met}$,** in response to a translation **initiation codon** (usually AUG, sometimes GUG). Therefore, all polypeptides begin with methionine during synthesis. The amino-terminal methionine is subsequently cleaved from many polypeptides. Thus, functional proteins need not have an amino-terminal methionine. The methionine on the initiator $tRNA_f^{Met}$ has the

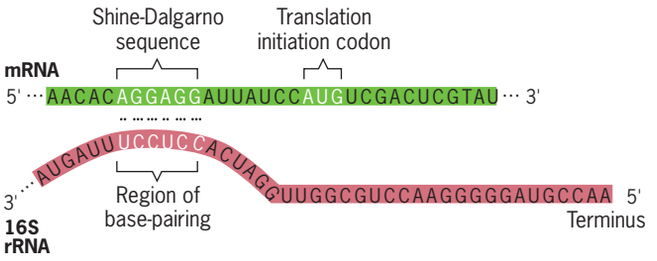


■ **FIGURE 12.16** The initiation of translation in *E. coli*.

amino group blocked with a formyl (—C(=O)—H) group (thus the "f" subscript in tRNA$_f^{Met}$). A distinct methionine tRNA, **tRNA^{Met}**, responds to internal methionine codons. Both methionine tRNAs have the same anticodon, and both respond to the same codon (AUG) for methionine. However, only methionyl-tRNA$_f^{Met}$ interacts with protein initiation factor IF-2 to begin the initiation process (Figure 12.16). Thus, only methionyl-tRNA$_f^{Met}$ binds to the ribosome in response to AUG initiation codons in mRNAs, leaving methionyl-tRNA^{Met} to bind in response to internal AUG codons. Methionyl-tRNA$_f^{Met}$ also binds to ribosomes in response to the alternate initiator codon, GUG

(a valine codon when present at internal positions), that occurs in some mRNA molecules.

Polypeptide chain initiation begins with the formation of two complexes: (1) one contains initiation factor IF-2 and methionyl-tRNA$_f^{Met}$, and (2) the other contains an mRNA molecule, a 30S ribosomal subunit and initiation factor IF-3 (Figure 12.16). The 30S subunit/mRNA complex will form only in the presence of IF-3; thus, IF-3 controls the ability of the 30S subunit to begin the initiation process. The formation of the 30S subunit/mRNA complex depends in part on base-pairing between a nucleotide sequence near the 3′ end of the 16S rRNA and a sequence near the 5′ end of the mRNA molecule (■ **Figure 12.17**). Prokaryotic mRNAs contain a conserved polypurine tract, consensus AGGAGG, located about seven nucleotides upstream from the AUG initiation codon. This conserved hexamer, called the **Shine-Dalgarno sequence** after the scientists who discovered it, is complementary to a sequence near the 3′ terminus of the 16S ribosomal RNA. When the Shine-Dalgarno sequences of mRNAs are experimentally modified so that they can no longer base-pair with the 16S rRNA, the modified mRNAs either are not translated or are translated very inefficiently, indicating that this base-pairing plays an important role in translation.



■ **FIGURE 12.17** Base-pairing between the Shine-Dalgarno sequence in a prokaryotic mRNA and a complementary sequence near the 3′ terminus of the 16S rRNA is involved in the formation of the mRNA/30S ribosomal subunit initiation complex.

The IF-2/methionyl-tRNA$_f^{Met}$ complex and the mRNA/30S subunit/IF-3 complex subsequently combine with each other and with initiation factor IF-1 and one molecule of GTP to form the complete 30S initiation complex. The final step in the initiation of translation is the addition of the 50S subunit to the 30S initiation complex to produce the complete 70S ribosome. Initiation factor IF-3 must be released from the complex before the 50S subunit can join the complex; IF-3 and the 50S subunit are never found to be associated with the 30S subunit at the same time. The addition of the 50S subunit requires energy from GTP and the release of initiation factors IF-1 and IF-2.

The addition of the 50S ribosomal subunit to the complex positions the initiator tRNA, methionyl-tRNA$_f^{Met}$, in the peptidyl (*P*) site with the anticodon of the tRNA aligned with the AUG initiation codon of the mRNA. Methionyl-tRNA$_f^{Met}$ is the only aminoacyl-tRNA that can enter the *P* site directly, without first passing through the aminoacyl (*A*) site. With the initiator AUG positioned in the *P* site, the second codon of the mRNA is in register with the *A* site, dictating the aminoacyl-tRNA binding specificity at that site and setting the stage for the second phase in polypeptide synthesis, chain elongation.

The initiation of translation is more complex in eukaryotes, involving several soluble initiation factors. Nevertheless, the overall process is similar except for two features. (1) The amino group of the methionine on the initiator tRNA is not formylated as in prokaryotes. (2) The initiation complex forms at the 5′ terminus of the mRNA, not at the Shine-Dalgarno/AUG translation start site as in *E. coli*. In eukaryotes, the initiation complex scans the mRNA, starting at the 5′ end, searching for an AUG translation-initiation codon. Thus, in eukaryotes, translation frequently begins at the AUG closest to the 5′ terminus of the mRNA molecule, although the efficiency with which a given AUG is used to initiate translation depends on the contiguous nucleotide sequence. The optimal initiation sequence is 5′-GCC(A or G)CC**AUG**G-3′. The purine (A or G) three bases upstream from the **AUG** initiator codon and the G immediately following it are the most important—influencing initiation efficiency by tenfold or more. Changes of other bases in the sequence cause smaller decreases in initiation efficiency. These sequence requirements for optimal translation initiation in eukaryotes are called **Kozak's rules,** after Marilyn Kozak, who first proposed them.

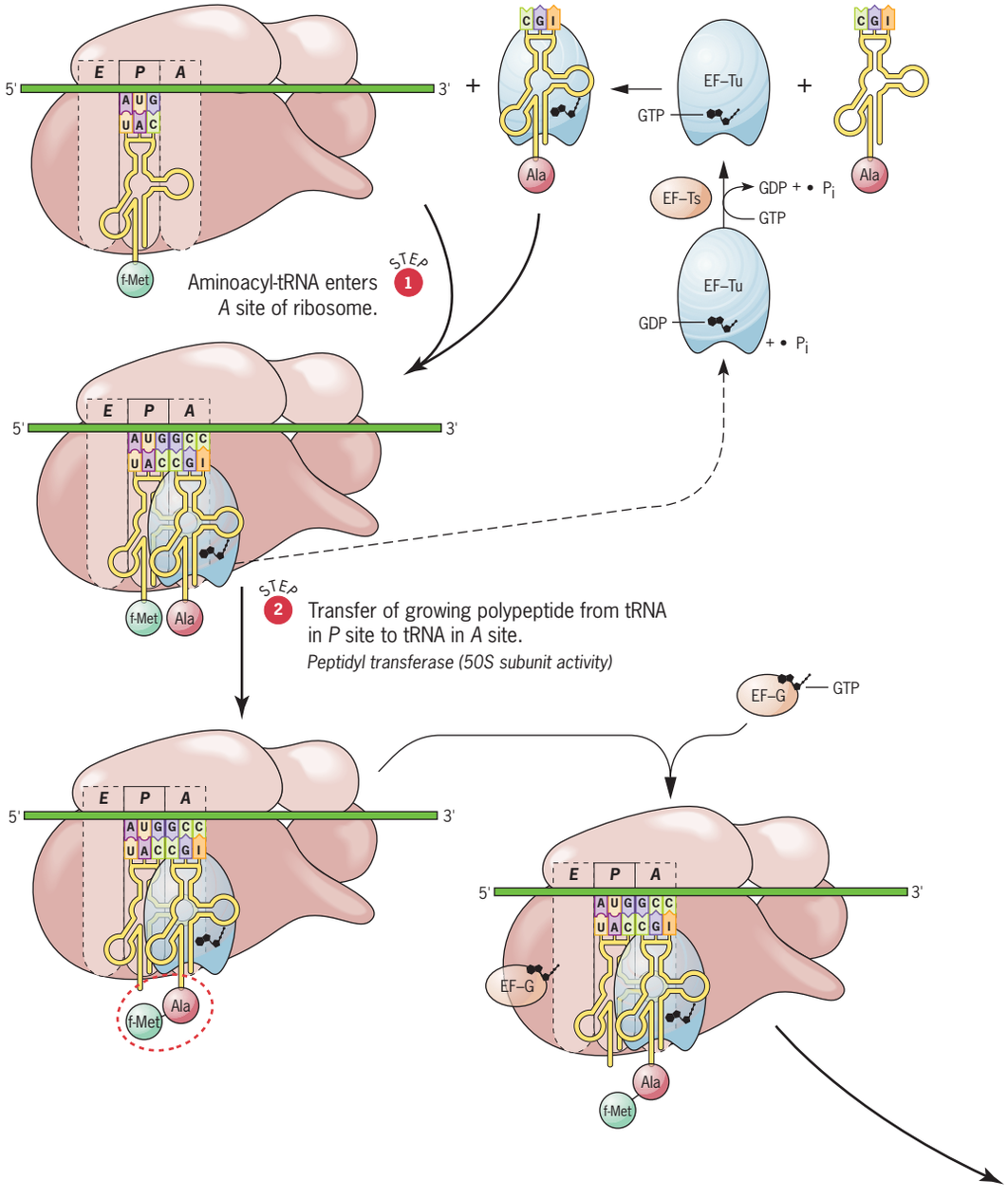
Like prokaryotes, eukaryotes contain a special initiator tRNA, **tRNA$_i^{Met}$** ("i" for initiator), but the amino group of the methionyl-tRNA$_i^{Met}$ is not formylated. The initiator methionyl-tRNA$_i^{Met}$ interacts with a soluble initiation factor and enters the *P* site directly during the initiation process, just as in *E. coli*.

In eukaryotes, a cap-binding protein (CBP) binds to the 7-methyl guanosine cap at the 5′ terminus of the mRNA. Then, other initiation factors bind to the CBP-mRNA complex, followed by the small (40S) subunit of the ribosome. The entire initiation complex moves 5′ → 3′ along the mRNA molecule, searching for an AUG codon.

When an AUG triplet is found, the initiation factors dissociate from the complex, and the large (60S) subunit binds to the methionyl-tRNA/mRNA/40S subunit complex, forming the complete (80S) ribosome. The 80S ribosome/mRNA/tRNA complex is ready to begin the second phase of translation, chain elongation. Try Solve It: Control of Translation in Eukaryotes to explore this process further.

Translation: Polypeptide Chain Elongation

The process of polypeptide chain **elongation** is basically the same in both prokaryotes and eukaryotes. The addition of each amino acid to the growing polypeptide occurs in three steps: (1) binding of an aminoacyl-tRNA to the *A* site of the ribosome, (2) transfer of the growing polypeptide chain from the tRNA in the *P* site to the tRNA in the *A* site by the formation of a new peptide bond, and (3) translocation of the ribosome along the mRNA to position the next codon in the *A* site (■ **Figure 12.18**).



■ **FIGURE 12.18** Polypeptide chain elongation in *E. coli*.

During step 3, the nascent polypeptide-tRNA and the uncharged tRNA are translocated from the *A* and *P* sites to the *P* and *E* sites, respectively. These three steps are repeated in a cyclic manner throughout the elongation process. The soluble factors involved in chain elongation in *E. coli* are described here. Similar factors participate in chain elongation in eukaryotes.

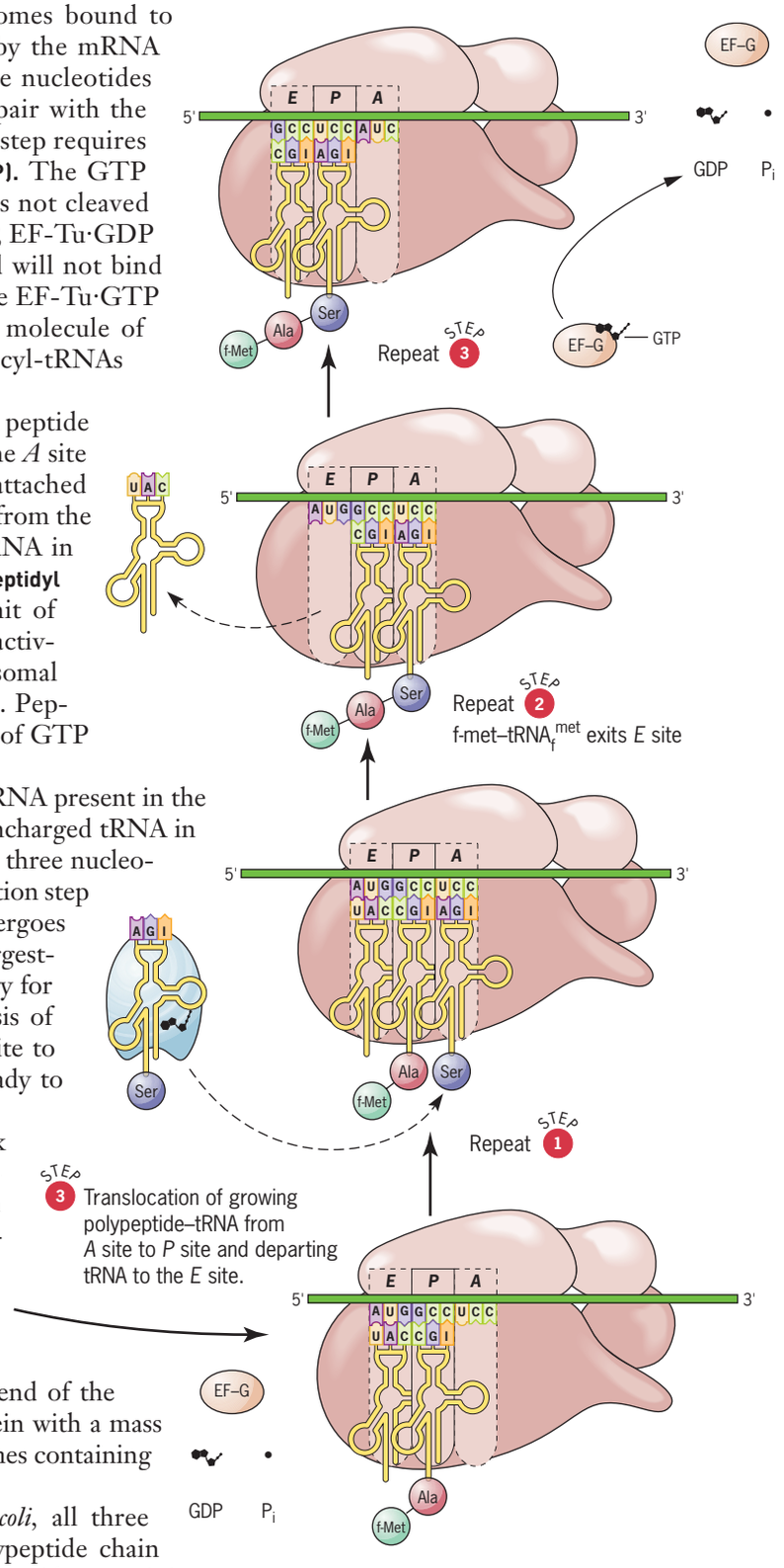
In the first step, an aminoacyl-tRNA enters and becomes bound to the *A* site of the ribosome, with the specificity provided by the mRNA codon in register with the *A* site (Figure 12.18). The three nucleotides in the anticodon of the incoming aminoacyl-tRNA must pair with the nucleotides of the mRNA codon present at the *A* site. This step requires **elongation factor Tu** carrying a molecule of GTP **(EF-Tu·GTP).** The GTP is required for aminoacyl-tRNA binding at the *A* site but is not cleaved until the peptide bond is formed. After the cleavage of GTP, EF-Tu·GDP is released from the ribosome. EF-Tu·GDP is inactive and will not bind to aminoacyl-tRNAs. EF-Tu·GDP is converted to the active EF-Tu·GTP form by **elongation factor Ts (EF-Ts),** which hydrolyzes one molecule of GTP in the process. EF-Tu interacts with all of the aminoacyl-tRNAs except methionyl-tRNA.

The second step in chain elongation is the formation of a peptide bond between the amino group of the aminoacyl-tRNA in the *A* site and the carboxyl terminus of the growing polypeptide chain attached to the tRNA in the *P* site. This uncouples the growing chain from the tRNA in the *P* site and covalently joins the chain to the tRNA in the *A* site (Figure 12.18). This key reaction is catalyzed by **peptidyl transferase,** an enzymatic activity built into the 50S subunit of the ribosome. We should note that the peptidyl transferase activity resides in the 23S rRNA molecule rather than in a ribosomal protein, perhaps another relic of an early RNA-based world. Peptide bond formation requires the hydrolysis of the molecule of GTP brought to the ribosome by EF-Tu in step 1.

During the third step in chain elongation, the peptidyl-tRNA present in the *A* site of the ribosome is translocated to the *P* site, and the uncharged tRNA in the *P* site is translocated to the *E* site, as the ribosome moves three nucleotides toward the 3′ end of the mRNA molecule. The translocation step requires GTP and **elongation factor G (EF-G).** The ribosome undergoes changes in conformation during the translocation process, suggesting that it may shuttle along the mRNA molecule. The energy for the movement of the ribosome is provided by the hydrolysis of GTP. The translocation of the peptidyl-tRNA from the *A* site to the *P* site leaves the *A* site unoccupied and the ribosome ready to begin the next cycle of chain elongation.

The elongation of one eukaryotic polypeptide, the silk protein fibroin, can be visualized with the electron microscope by using techniques developed by Oscar Miller, Barbara Hamkalo, and colleagues. Most proteins fold up on the surface of the ribosome during their synthesis. However, fibroin remains extended from the surface of the ribosome under the conditions used by Miller and coworkers. As a result, nascent polypeptide chains of increasing length can be seen attached to the ribosomes as they are scanned from the 5′ end of the mRNA to the 3′ end (■ **Figure 12.19**). Fibroin is a large protein with a mass of over 200,000 daltons; it is synthesized on large polyribosomes containing 50 to 80 ribosomes.

Polypeptide chain elongation proceeds rapidly. In *E. coli*, all three steps required to add one amino acid to the growing polypeptide chain occur in about 0.05 second. Thus, the synthesis of a polypeptide containing



■ **FIGURE 12.18** (*continued*)

Solve It!

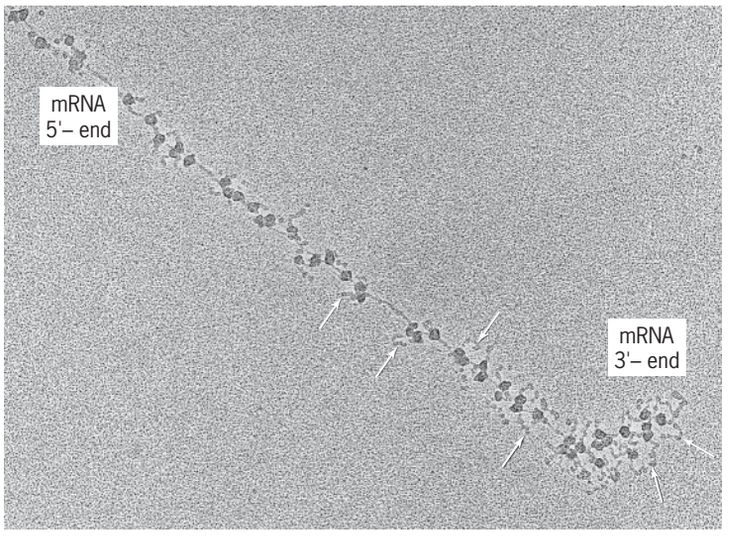
Control of Translation in Eukaryotes

The nucleotide sequence of the nontemplate strand of a portion of the human *HBB* (β-globin) gene specifying the 5′-terminus of the *HBB* mRNA is given below. Remember that the nontemplate strand will have the same sequence as the transcript of the gene, but with Ts in place of Us. Position 1 is the nucleotide corresponding to the 5′-end of the mRNA.

```
1 ACATTTGCTT   CTGACACAAC
TGTGTTCACT   AGCAACCTCA
AACAGACACC   ATGGTGCATC
TGACTCCTGA   GGAGAAGTCT
GCCGTTACTG   CCCTGTGGGG
```

Based on this sequence, the genetic code (see Table 12.1), and your knowledge of the initiation of translation in eukaryotes, predict the amino-terminal amino acid sequence of human β-globin.

▶ *To see the solution to this problem, visit the Student Companion site.*



■ **FIGURE 12.19** Visualization of the elongation of fibroin polypeptides in the posterior silk gland of the silkworm *Bombyx mori*. The arrows point to growing fibroin polypeptides. Note their increasing length as one approaches the 3′ end of the mRNA molecule.

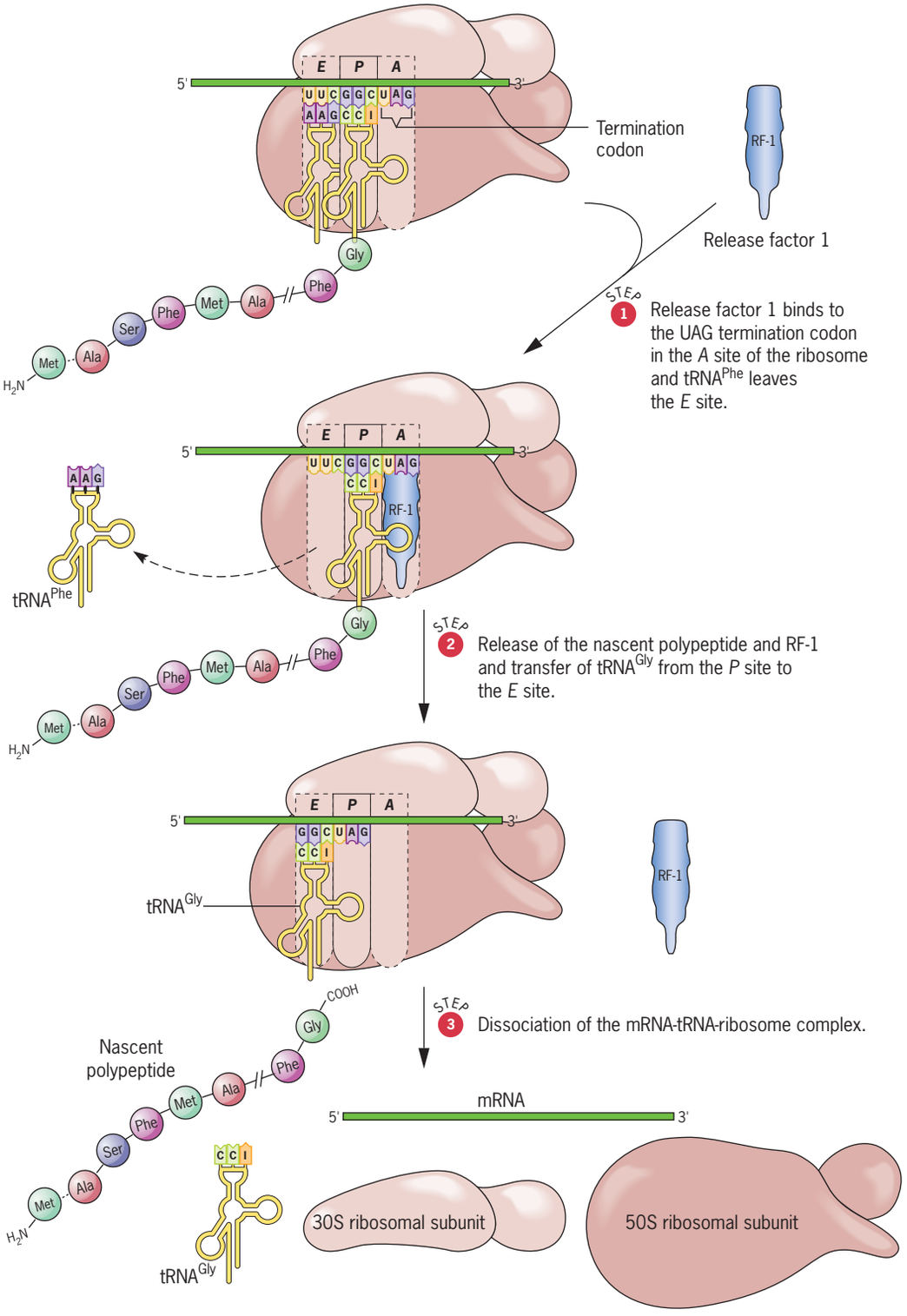
300 amino acids takes only about 15 seconds. Given its complexity, the accuracy and efficiency of the translational apparatus are indeed amazing.

Translation: Polypeptide Chain Termination

Polypeptide chain elongation undergoes **termination** when any of three **chain-termination codons** (UAA, UAG, or UGA) enters the *A* site on the ribosome (■ Figure 12.20). These three stop codons are recognized by soluble proteins called **release factors (RFs).** In *E. coli*, there are two release factors, RF-1 and RF-2. RF-1 recognizes termination codons UAA and UAG; RF-2 recognizes UAA and UGA. In eukaryotes, a single release factor (**eRF**) recognizes all three termination codons. The presence of a release factor in the *A* site alters the activity of peptidyl transferase such that it adds a water molecule to the carboxyl terminus of the nascent polypeptide. This reaction releases the polypeptide from the tRNA molecule in the *P* site and triggers the translocation of the free tRNA to the *E* site. Termination is completed by the release of the mRNA molecule from the ribosome and the dissociation of the ribosome into its subunits. The ribosomal subunits are then ready to initiate another round of protein synthesis, as previously described.

KEY POINTS

- *Genetic information carried in the sequences of nucleotides in mRNA molecules is translated into sequences of amino acids in polypeptide gene products by intricate macromolecular machines called ribosomes.*
- *The translation process is complex, requiring the participation of many different RNA and protein molecules.*
- *Transfer RNA molecules serve as adaptors, mediating the interaction between amino acids and codons in mRNA.*
- *The process of translation involves the initiation, elongation, and termination of polypeptide chains and is governed by the specifications of the genetic code.*



■ **FIGURE 12.20** Polypeptide chain termination in *E. coli*. The formyl group of formylmethionine is removed during translation.

The Genetic Code

The genetic code is a nonoverlapping code, with each amino acid plus polypeptide initiation and termination specified by RNA codons composed of three nucleotides.

As it became evident that genes controlled the structure of polypeptides, attention focused on how the sequence of the four different nucleotides in DNA could control the sequence of the 20 amino acids present in proteins. With the discovery of the mRNA intermediary (Chapter 11), the question became one of how the sequence of the four bases present in mRNA molecules could specify the amino acid sequence of a polypeptide. What is the nature of the genetic code relating mRNA base sequences to amino acid sequences? Clearly, the symbols or letters used in the code must be the bases; but what comprises a codon, the unit or word specifying one amino acid or, actually, one aminoacyl-tRNA?

PROPERTIES OF THE GENETIC CODE: AN OVERVIEW

The main features of the genetic code were worked out during the 1960s. Cracking the code was one of the most exciting events in the history of science, with new information reported almost daily. By the mid-1960s, the genetic code was largely solved. Before focusing on specific features of the code, let us consider its most important properties.

1. *The genetic code is composed of nucleotide triplets*. Three nucleotides in mRNA specify one amino acid in the polypeptide product; thus, each codon contains three nucleotides.
2. *The genetic code is nonoverlapping*. Each nucleotide in mRNA belongs to just one codon except in rare cases where genes overlap and a nucleotide sequence is read in two different reading frames.
3. *The genetic code is comma-free*. There are no commas or other forms of punctuation within the coding regions of mRNA molecules. During translation, the codons are read consecutively.
4. *The genetic code is degenerate*. All but two of the amino acids are specified by more than one codon.
5. *The genetic code is ordered*. Multiple codons for a given amino acid and codons for amino acids with similar chemical properties are closely related, usually differing by a single nucleotide.
6. *The genetic code contains start and stop codons*. Specific codons are used to initiate and to terminate polypeptide chains.
7. *The genetic code is nearly universal*. With minor exceptions, the codons have the same meaning in all living organisms, from viruses to humans.

THREE NUCLEOTIDES PER CODON

Twenty different amino acids are incorporated into polypeptides during translation. Thus, at least 20 different codons must be formed with the four bases available in mRNA. Two bases per codon would result in only 4^2 or 16 possible codons—clearly not enough. Three bases per codon yields 4^3 or 64 possible codons—an apparent excess.

In 1961, Francis Crick and colleagues published the first strong evidence in support of a *triplet code* (three nucleotides per codon). Crick and coworkers carried out a genetic analysis of mutations induced at the *r*II locus of bacteriophage T4 by the chemical proflavin. Proflavin is a mutagenic agent that causes single base-pair additions and deletions (Chapter 13). Phage T4 *r*II mutants are unable to grow in cells of *E. coli* strain K12, but grow like wild-type phage in cells of *E. coli* strain B. Wild-type T4 grows equally well on either strain. Crick and coworkers isolated proflavin-induced

revertants of a proflavin-induced *r*II mutation. These revertants were shown to result from the occurrence of additional mutations at nearby sites rather than reversion of the original mutation. Second-site mutations that restore the wild-type phenotype in a mutant organism are called **suppressor mutations** because they cancel, or suppress, the effect(s) of the original mutation.

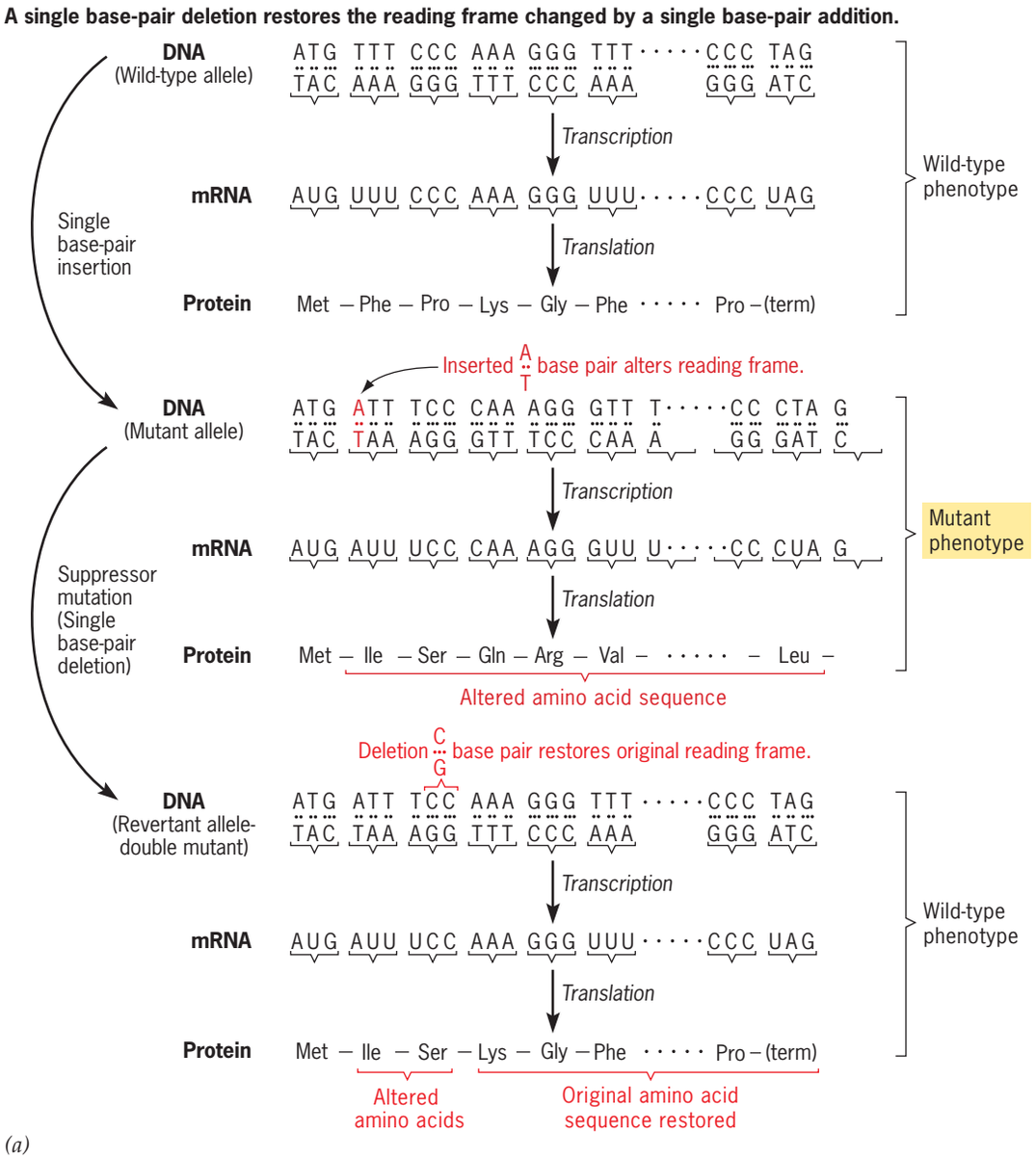
Crick and colleagues reasoned that if the original mutation was a single base-pair addition or deletion, then the suppressor mutations must be single base-pair deletions or additions, respectively, occurring at a site or sites near the original mutation. If sequential nucleotide triplets in an mRNA specify amino acids, then every nucleotide sequence can be recognized or read during translation in three different ways. For example, the sequence AAAGGGCCCTTT can be read (1) AAA, GGG, CCC, TTT, (2) A, AAG, GGC, CCT, TT, or (3) AA, AGG, GCC, CTT, T. The **reading frame** of an mRNA is the series of nucleotide triplets that are read (positioned in the *A* site of the ribosome) during translation. A single base-pair addition or deletion will alter the reading frame of the gene and mRNA for that portion of the gene distal to the mutation. This effect is illustrated in ■ **Figure 12.21*a***. The suppressor mutations were then isolated as single mutants by screening progeny of backcrosses to wild-type. Like the original mutation, the suppressor mutations were found to produce *r*II mutant phenotypes. Crick and colleagues next isolated proflavin-induced suppressor mutations of the original suppressor mutations, and so on.

Crick and colleagues then classified all the isolated mutations into two groups, plus (+) and minus (−) (for additions and deletions, although they had no idea which group was which), based on the reasoning that a (+) mutation would suppress a (−) mutation but not another (+) mutation, and vice versa (Figure 12.21). Then, Crick and coworkers constructed recombinants that carried various combinations of the (+) and the (−) mutations. Like the single mutants, recombinants with two (+) mutations or two (−) mutations always had the mutant phenotype. The critical result was that recombinants with three (+) mutations (■ **Figure 12.21*b***) or three (−) mutations often exhibited the wild-type phenotype. This indicated that the addition of three base pairs or the deletion of three base pairs left the distal portion of the gene with the wild-type reading frame. This result would be expected only if each codon contained three nucleotides.

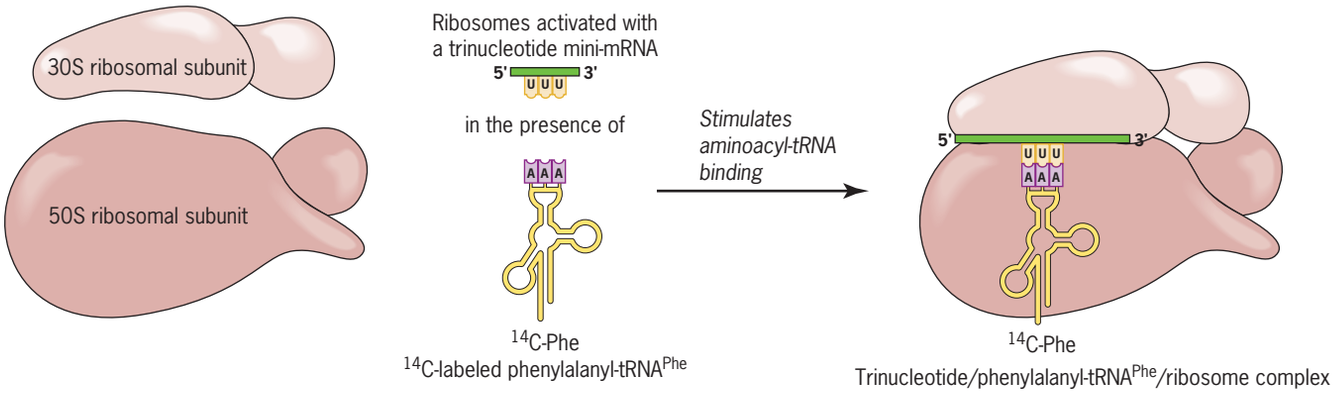
Evidence from *in vitro* translation studies soon supported the results of Crick and colleagues and firmly established the triplet nature of the code. Some of the more important results follow: (1) Trinucleotides were sufficient to stimulate specific binding of aminoacyl-tRNAs to ribosomes. For example, 5′-UUU-3′ stimulated the binding of phenylalanyl-$tRNA^{Phe}$ to ribosomes. (2) Chemically synthesized mRNA molecules that contained repeating dinucleotide sequences directed the synthesis of copolymers (large chainlike molecules composed of two different subunits) with alternating amino acid sequences. For example, when $poly(UG)_n$ was used as an artificial mRNA in an *in vitro* translation system, the repeating copolymer $(cys\text{-}val)_m$ was synthesized. (The subscripts *n* and *m* refer to the number of nucleotides and amino acids in the respective polymers.) (3) In contrast, mRNAs with repeating trinucleotide sequences directed the synthesis of a mixture of three homopolymers (initiation being at random on such mRNAs in the *in vitro* systems). For example, $poly(UUG)_n$ directed the synthesis of a mixture of polyleucine, polycysteine, and polyvaline. These results are consistent only with a triplet code, with its three different reading frames. When $poly(UUG)_n$ is translated in reading frame 1, UUG, UUG, polyleucine is produced, whereas translation in reading frame 2, UGU, UGU, yields polycysteine, and translation in reading frame 3, GUU, GUU, produces polyvaline. Ultimately, the triplet nature of the code was definitively established by comparing the nucleotide sequences of genes and mRNAs with the amino acid sequences of their polypeptide products.

DECIPHERING THE CODE

The cracking of the genetic code in the 1960s took several years and involved intense competition between many different research laboratories. New information



■ **FIGURE 12.21** Early evidence that the genetic code is a triplet code. See the text for details.



■ **FIGURE 12.22** Stimulation of aminoacyl-tRNA binding to ribosomes by synthetic trinucleotide mini-mRNAs. The results of these trinucleotide-activated ribosome binding assays helped scientists crack the genetic code.

accumulated rapidly but sometimes was inconsistent with earlier data. Indeed, cracking the code proved to be a major challenge.

Deciphering the genetic code required scientists to obtain answers to several questions. (1) Which codons specify each of the 20 amino acids? (2) How many of the 64 possible triplet codons are utilized? (3) How is the code punctuated? (4) Do the codons have the same meaning in viruses, bacteria, plants, and animals? The answers to these questions were obtained primarily from the results of two types of experiments, both of which were performed with cell-free systems. The first type of experiment involved translating artificial mRNA molecules *in vitro* and determining which of the 20 amino acids were incorporated into proteins. In the second type of experiment, ribosomes were activated with mini-mRNAs just three nucleotides long. Then, researchers determined which aminoacyl-tRNAs were stimulated to bind to ribosomes activated with each of the trinucleotide messages (■ **Figure 12.22**).

The decade of the 1960s—the era of the cracking of the genetic code—was one of the most exciting times in the history of biology. Deciphering the genetic code was a difficult and laborious task, and progress came in a series of breakthroughs. We discuss these important developments in A Milestone in Genetics: Cracking the Genetic Code on the Student Companion web site. By combining the results of *in vitro* translation experiments performed with synthetic mRNAs and trinucleotide binding assays, Marshall Nirenberg, Severo Ochoa, H. Ghobind Khorana, Philip Leder, and their colleagues worked out the meaning of all 64 triplet codons (**Table 12.1**). Nirenberg and Khorana shared the 1968 Nobel Prize in Physiology or Medicine for their work on the code with Robert Holley, who determined the complete nucleotide sequence of the yeast alanine tRNA. Ochoa had already received the 1959 Nobel Prize for his discovery of RNA polymerase.

INITIATION AND TERMINATION CODONS

The genetic code also provides for punctuation of genetic information at the level of translation. In both prokaryotes and eukaryotes, the codon AUG is used to initiate polypeptide chains (Table 12.1). In rare instances, GUG is used as an initiation codon. In both cases, the initiation codon is recognized by an initiator tRNA, $tRNA_f^{Met}$ in prokaryotes and $tRNA_i^{Met}$ in eukaryotes. In prokaryotes, an AUG codon must follow an appropriate nucleotide sequence, the Shine-Delgarno sequence, in the 5′ nontranslated segment of the mRNA molecule in order to serve as translation initiation codon. In eukaryotes, the codon must be the first AUG encountered by the ribosome as it scans from the 5′ end of the mRNA molecule. At internal positions, AUG is recognized by $tRNA^{Met}$, and GUG is recognized by a valine tRNA.

TABLE 12.1

The Genetic Code[a]

First (5′) letter	Second letter: U	Second letter: C	Second letter: A	Second letter: G	Third (3′) letter
U	UUU Phe (F)	UCU Ser (S)	UAU Tyr (Y)	UGU Cys (C)	U
	UUC Phe (F)	UCC Ser (S)	UAC Tyr (Y)	UGC Cys (C)	C
	UUA Leu (L)	UCA Ser (S)	UAA Stop (terminator)	UGA Stop (terminator)	A
	UUG Leu (L)	UCG Ser (S)	UAG Stop (terminator)	UGG Trp (W)	G
C	CUU Leu (L)	CCU Pro (P)	CAU His (H)	CGU Arg (R)	U
	CUC Leu (L)	CCC Pro (P)	CAC His (H)	CGC Arg (R)	C
	CUA Leu (L)	CCA Pro (P)	CAA Gln (Q)	CGA Arg (R)	A
	CUG Leu (L)	CCG Pro (P)	CAG Gln (Q)	CGG Arg (R)	G
A	AUU Ile (I)	ACU Thr (T)	AAU Asn (N)	AGU Ser (S)	U
	AUC Ile (I)	ACC Thr (T)	AAC Asn (N)	AGC Ser (S)	C
	AUA Ile (I)	ACA Thr (T)	AAA Lys (K)	AGA Arg (R)	A
	AUG Met (M) (initiator)	ACG Thr (T)	AAG Lys (K)	AGG Arg (R)	G
G	GUU Val (V)	GCU Ala (A)	GAU Asp (D)	GGU Gly (G)	U
	GUC Val (V)	GCC Ala (A)	GAC Asp (D)	GGC Gly (G)	C
	GUA Val (V)	GCA Ala (A)	GAA Glu (E)	GGA Gly (G)	A
	GUG Val (V)	GCG Ala (A)	GAG Glu (E)	GGG Gly (G)	G

= Polypeptide chain initiation codon

= Polypeptide chain termination codon

[a]Each triplet nucleotide sequence or codon refers to the nucleotide sequence in **mRNA** (not DNA) that specifies the incorporation of the indicated amino acid or polypeptide chain termination. The one-letter symbols for the amino acids are given in parentheses after the standard three-letter abbreviations.

Three codons—UAG, UAA, and UGA—specify polypeptide chain termination (Table 12.1). These codons are recognized by protein release factors, rather than by tRNAs. Prokaryotes contain two release factors, RF-1 and RF-2. RF-1 terminates polypeptides in response to codons UAA and UAG, whereas RF-2 causes termination at UAA and UGA codons. Eukaryotes contain a single release factor that recognizes all three termination codons.

A DEGENERATE AND ORDERED CODE

All the amino acids except methionine and tryptophan are specified by more than one codon (Table 12.1). Three amino acids—leucine, serine, and arginine—are each specified by six different codons. Isoleucine has three codons. The other amino acids each have either two or four codons. The occurrence of more than one codon per amino acid is called **degeneracy** (although the usual connotations of the term are hardly appropriate). The degeneracy in the genetic code is not at random; instead, it is highly ordered. In most cases, the multiple codons specifying a given amino acid differ by only one base, the third or 3′ base of the codon. The degeneracy is primarily of two types. (1) Partial degeneracy occurs when the third base may be either of the two pyrimidines (U or C) or, alternatively, either of the two purines (A or G). With partial degeneracy, changing the third base from a purine to a pyrimidine, or vice versa, will change the amino acid specified by the codon. (2) In the case of complete degeneracy,

any of the four bases may be present at the third position in the codon, and the codon will still specify the same amino acid. For example, valine is encoded by GUU, GUC, GUA, and GUG (Table 12.1).

Scientists have speculated that the **order** in the genetic code has evolved as a way of minimizing mutational lethality. Many base substitutions at the third position of codons do not change the amino acid specified by the codon. Moreover, amino acids

PROBLEM-SOLVING SKILLS

Predicting Amino Acid Substitutions Induced by Mutagens

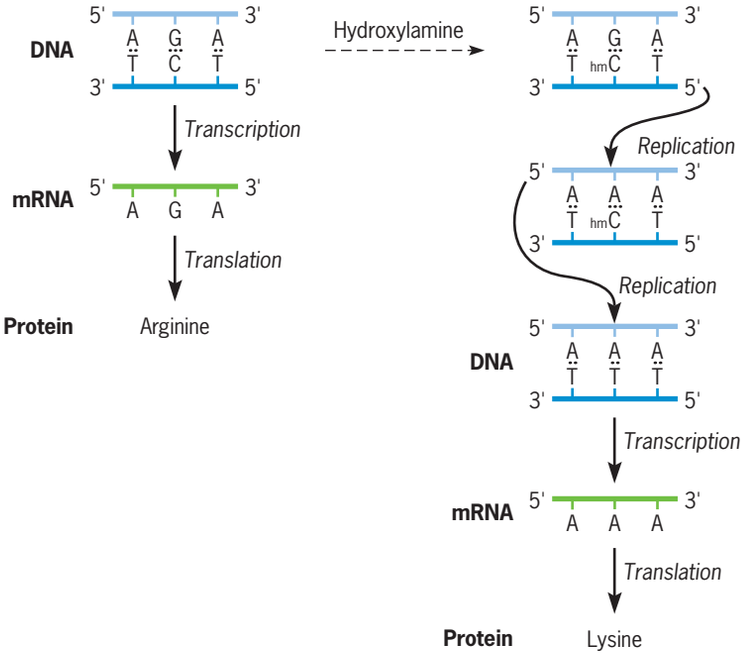
THE PROBLEM

The chemical hydroxylamine (NH_2OH) transfers a hydroxyl (-OH) group to cytosine producing hydroxymethylcytosine (hmC), which, unlike cytosine, pairs with adenine. Therefore, hydroxylamine induces G:C to A:T base-pair substitutions in DNA. If you treat the double-stranded DNA of a virus such as phage T4 with hydroxylamine, what amino acid substitutions will be induced in the proteins encoded by the virus?

FACTS AND CONCEPTS

1. The nature of the genetic code—the meaning of the 64 triplet nucleotide sequences in mRNA—is shown in Table 12.1.
2. Complete degeneracy occurs when the first two nucleotides in an mRNA codon are sufficient to determine the amino acid in the polypeptide specified by the mRNA.
3. Partial degeneracy occurs when the same amino acid is specified if the base in the 3′ nucleotide of a codon is either of the two pyrimidines or either of the two purines.
4. Hydroxylamine will only alter codons specified by DNA base-pair triplets that contain G:C base pairs.
5. If the G:C base pair occupies the third (3′) position of the triplet, hydroxylamine will induce amino acid substitutions only in cases where the genetic code is NOT degenerate, that is, where the base present as the 3′ nucleotide of the codon determines its meaning. Only two codons are not degenerate at the 3′ position; they are 5′-AUG-3′ (methionine) and 5′-UGG-3′ (tryptophan).
6. For codons with complete or partial degeneracy at the 3′ position, hydroxylamine will not induce amino acid substitutions by modifying the base pair specifying the 3′ base in the codon. It will induce G:C → A:T and C:G → T:A substitutions (where the first base given is in the template strand). However, given the partial or complete degeneracy, the resulting codons will still specify the same amino acids. An AAG lysine codon, for example, could be changed to an AAA lysine codon, or a UUC phenylalanine codon could be changed to a UUU phenylalanine codon. However, no amino acid substitution will occur in either case.

(3′) position. Indeed, there are more potential targets in genomes than nontargets; 51 of the 64 DNA triplets contain G:C or C:G base pairs. Consider as an example the arginine codon 5′-AGA-3′; it will be transcribed from a DNA template strand with the sequence 3′-TCT-5′ (reversing the polarity to keep the bases in the same order). The C in this sequence can be hydroxymethylated, producing hmC, which will pair with adenine. After two semiconservative replications, the DNA template strand will contain the sequence 3′-TTT-5′ at this site, and transcription of this sequence will yield a 5′-AAA-3′ mRNA codon. Translation of the mRNA will result in the insertion of lysine in the resulting polypeptide because AAA is a lysine codon. Thus, one example of the effects of hydroxylamine will be the replacement of arginine residues with lysines. This process is diagrammed below.



ANALYSIS AND SOLUTION

The answer to the question of which amino acid substitutions will be induced by hydroxylamine requires a careful analysis of the nature of the genetic code (Table 12.1). Potential targets of hydroxylamine mutagenesis are DNA triplets specifying mRNA codons containing C's and G's at the first (5′) and second positions in the codons and triplets specifying nondegenerate codons with G's or C's at the third

The only amino acids specified by codons with no targets of hydroxylamine-induced amino acid substitutions are phenylalanine (UUU & UUC), isoleucine (AUU, AUC, & AUA), tyrosine (UAU & UAC), asparagine (AAU & AAC), and lysine (AAA & AAG). The other amino acids are all specified by DNA base-pair triplets that contain one or more G:C's, with the C's being potential targets of hydroxylamine mutagenesis. For further discussion visit the Student Companion site.

with similar chemical properties (such as leucine, isoleucine, and valine) have codons that differ from each other by only one base. Thus, many single base-pair substitutions will result in the substitution of one amino acid for another amino acid with very similar chemical properties (for example, valine for isoleucine). In most cases, conservative substitutions of this type will yield active gene products, which minimizes the effects of mutations. Try Problem-Solving Skills: Predicting Amino Acid Substitutions Induced by Mutagens to test your understanding of the genetic code.

A NEARLY UNIVERSAL CODE

Vast quantities of information are now available from *in vitro* studies, from amino acid replacements due to mutations, and from correlated nucleic acid and polypeptide sequencing, which allow a comparison of the meaning of the 64 codons in different species. These data all indicate that the genetic code is nearly **universal;** that is, the codons have the same meaning, with minor exceptions, in all species.

The most important exceptions to the universality of the code occur in mitochondria of mammals, yeast, and several other species. Mitochondria have their own chromosomes and protein-synthesizing machinery (Chapter 15). Although the mitochondrial and cytoplasmic systems are similar, there are some differences. In the mitochondria of humans and other mammals, (1) UGA specifies tryptophan rather than chain termination, (2) AUA is a methionine codon, not an isoleucine codon, and (3) AGA and AGG are chain-termination codons rather than arginine codons. The other 60 codons have the same meaning in mammalian mitochondria as in nuclear mRNAs (Table 12.1). There are also rare differences in codon meaning in the mitochondria of other species and in nuclear transcripts of some protozoa. However, since these exceptions are rare, the genetic code should be considered nearly universal.

KEY POINTS

- *Each of the 20 amino acids in proteins is specified by one or more nucleotide triplets in mRNA.*
- *Of the 64 possible triplets, given the four bases in mRNA, 61 specify amino acids and 3 signal chain termination.*
- *The code is nonoverlapping, with each nucleotide part of a single codon, degenerate, with most amino acids specified by two or four codons, and ordered, with similar amino acids specified by related codons.*
- *The genetic code is nearly universal; with minor exceptions, the 64 triplets have the same meaning in all organisms.*

Codon-tRNA Interactions

Codons in mRNA molecules are recognized by aminoacyl-tRNAs during translation.

The translation of a sequence of nucleotides in mRNA into the correct sequence of amino acids in the polypeptide product requires the accurate recognition of codons by aminoacyl-tRNAs. Because of the degeneracy of the genetic code, either several different tRNAs must recognize the different codons specifying a given amino acid or the anticodon of a given tRNA must be able to base-pair with several different codons. Actually, both of these phenomena occur. Several tRNAs exist for certain amino acids, and some tRNAs recognize more than one codon.

RECOGNITION OF CODONS BY tRNAs: THE WOBBLE HYPOTHESIS

The hydrogen bonding between the bases in the anticodons of tRNAs and the codons of mRNAs follows strict base-pairing rules only for the first two bases of the codon. The base-pairing involving the third base of the codon is less stringent, allowing what Crick has called *wobble* at this site. On the basis of molecular distances and steric (three-dimensional structure) considerations, Crick proposed that wobble would allow

several types, but not all types, of base-pairing at the third codon base during the codon–anticodon interaction. His proposal has since been strongly supported by experimental data. Table 12.2 shows the base-pairing predicted by Crick's wobble hypothesis.

The **wobble hypothesis** predicted the existence of at least two tRNAs for each amino acid with codons that exhibit complete degeneracy, and this has proven to be true. The wobble hypothesis also predicted the occurrence of three tRNAs for the six serine codons. Three serine tRNAs have been characterized: (1) $tRNA^{Ser1}$ (anticodon AGG) binds to codons UCU and UCC, (2) $tRNA^{Ser2}$ (anticodon AGU) binds to codons UCA and UCG, and (3) $tRNA^{Ser3}$ (anticodon UCG) binds to codons AGU and AGC. These specificities were verified by the trinucleotide-stimulated binding of purified aminoacyl-tRNAs to ribosomes *in vitro*.

Finally, several tRNAs contain the base inosine, which is made from the purine hypoxanthine. Inosine is produced by a posttranscriptional modification of adenosine. Crick's wobble hypothesis predicted that when inosine is present at the 5′ end of an anticodon (the wobble position), it would base-pair with uracil, cytosine, or adenine in the codon. In fact, purified alanyl-tRNA containing inosine (I) at the 5′ position of the anticodon (see Figure 12.12) binds to ribosomes activated with GCU, GCC, or GCA trinucleotides (■ Figure 12.23). The same result has been obtained with other purified tRNAs with inosine at the 5′ position of the anticodon. Thus, Crick's wobble hypothesis nicely explains the relationships between tRNAs and codons given the degenerate, but ordered, genetic code.

TABLE 12.2

Base-Pairing between the 5′ Base of the Anticodons of tRNAs and the 3′ Base of Codons of mRNAs According to the Wobble Hypothesis

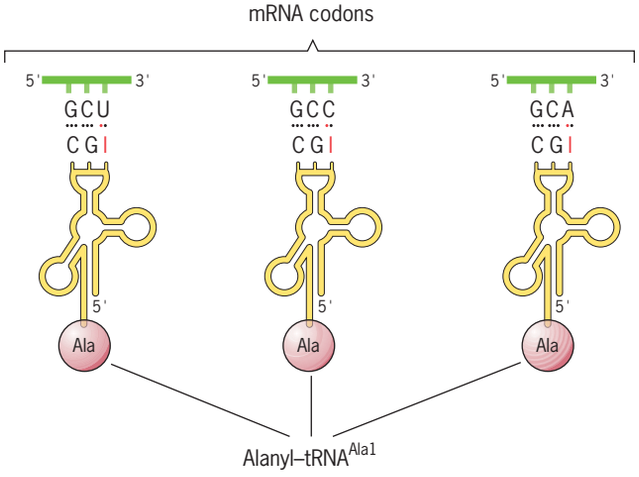
Base in Anticodon	Base in Codon
G	U or C
C	G
A	U
U	A or G
I	A, U, or C

SUPPRESSOR MUTATIONS THAT PRODUCE tRNAs WITH ALTERED CODON RECOGNITION

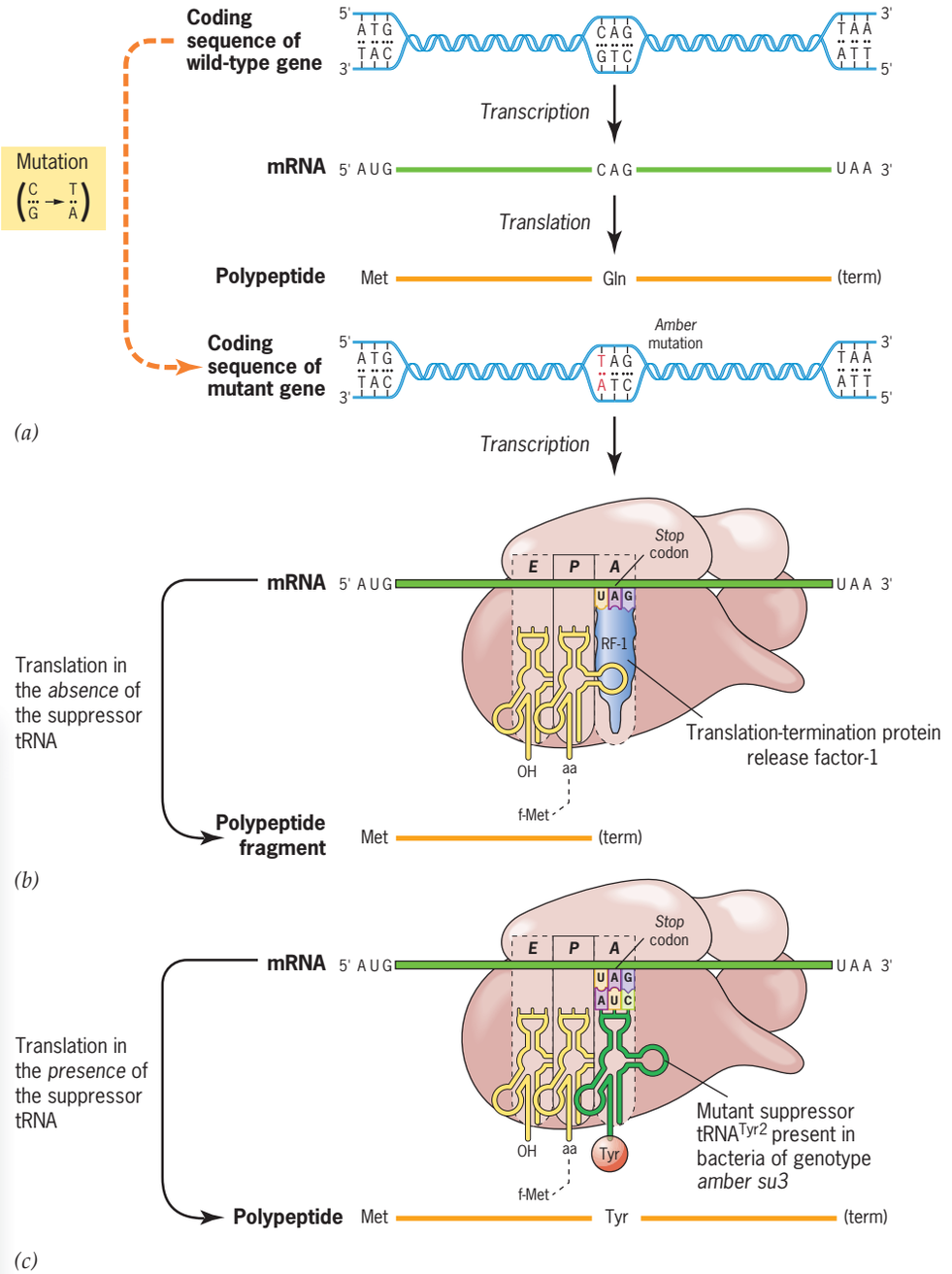
Even if we exclude the mitochondria, the genetic code is not absolutely universal. Minor variations in codon recognition and translation are well documented. In *E. coli* and yeast, for example, some mutations in tRNA genes alter the anticodons and thus the codons recognized by the mutant tRNAs. These mutations were initially detected as *suppressor mutations*, nucleotide substitutions that suppressed the effects of other mutations. The suppressor mutations were subsequently shown to occur in tRNA genes. Many of these suppressor mutations changed the anticodons of the altered tRNAs.

The best-known examples of suppressor mutations that alter tRNA specificity are those that suppress UAG chain-termination mutations within the coding sequences of genes. Such mutations, called *amber* mutations (after one of the researchers who discovered them), result in the synthesis of truncated polypeptides. Mutations that produce chain-termination triplets within genes have come to be known as **nonsense mutations,** in contrast to **missense mutations,** which change a triplet so that it specifies a different amino acid. A gene that contains a missense mutation encodes a complete polypeptide, but with an amino acid substitution in the polypeptide gene product. A nonsense mutation results in a truncated polypeptide, with the length of the chain depending on the position of the mutation within the gene. Nonsense mutations frequently result from single base-pair substitutions, as illustrated in ■ Figure 12.24*a*. The polypeptide fragments produced from genes containing nonsense mutations (■ Figure 12.24*b*) often are completely nonfunctional. See Solve It: Effects of Base-Pair Substitutions in the Coding Region of the *HBB* Gene.

Suppression of nonsense mutations has been shown to result from mutations in tRNA genes that cause the mutant tRNAs to recognize the termination (UAG, UAA, or UGA) codons, albeit with varying efficiencies. These mutant tRNAs are referred to as **suppressor tRNAs.** When the *amber* (UAG) suppressor tRNA produced by the *amber su3* mutation in *E. coli* was sequenced, it was found to have an altered anticodon. This particular *amber* suppressor mutation occurs in the $tRNA^{Tyr2}$ gene (one of two



■ **FIGURE 12.23** Base-pairing between the anticodon of alanyl-$tRNA^{Ala1}$ and mRNA codons GCU, GCC, and GCA according to Crick's wobble hypothesis. Trinucleotide-activated ribosome binding assays have shown that alanyl-$tRNA^{Ala1}$ does indeed base-pair with all three codons.



■ **FIGURE 12.24** (*a*) The formation of an *amber* (UAG) chain-termination mutation. (*b*) Its effect on the polypeptide gene product in the absence of a suppressor tRNA, and (*c*) in the presence of a suppressor tRNA. The *amber* mutation shown here changes a CAG glutamine (Gln) codon to a UAG chain-termination codon. The polypeptide containing the tyrosine inserted by the suppressor tRNA may or may not be functional; however, suppression of the mutant phenotype will occur only when the polypeptide is functional.

Solve It!

Effects of Base-Pair Substitutions in the Coding Region of the *HBB* Gene

The first 42 nucleotides, shown as triplets corresponding to mRNA codons, in the nontemplate strand of the coding region of the human *HBB* (β-globin) gene are given below. Recall that the nontemplate strand has the same sequence as the mRNA, but with T's in place of U's. The first (amino-terminal) 14 amino acids of the nascent human β-globin are also given using the single letter code (see Table 12.1). The methionine is subsequently removed to yield mature β-globin. Consider the potential phenotypic effects of the four single nucleotide substitutions numbered 1 through 4 below when present in homozygotes.

```
      1   2   3       4
      T   T   T       C
      ↑   ↑   ↑       ↑
1 ATG GTG CAT CTG ACT CCT GAG GAG AAG TCT GCC GTT ACT GCC
H2N M — V — H — L — T — P — E — E — K — S — A — V — T — A —
```

Which substitution would you expect to have the largest effect on phenotype? The second largest effect? No effect? No, or a very small, effect?

▶ *To see the solution to this problem, visit the Student Companion site.*

tyrosine tRNA genes in *E. coli*). The anticodon of the wild-type (nonsuppressor) $tRNA^{Tyr2}$ was shown to be 5′-G′UA-3′ (where G′ is a derivative of guanine). The anticodon of the mutant (suppressor) $tRNA^{Tyr2}$ is 5′-CUA-3′. Because of the single-base substitution, the anticodon of the suppressor $tRNA^{Tyr2}$ base-pairs with the 5′-UAG-3′ *amber* codon (recall that base-pairing always involves strands of opposite polarity); that is,

mRNA: 5′-UAG-3′ (codon)
tRNA: 3′-AUC-5′ (anticodon)

Thus, suppressor tRNAs allow complete polypeptides to be synthesized from mRNAs containing termination codons within genes (■ **Figure 12.24c**). Such polypeptides will be functional if the amino acid inserted by the suppressor tRNA does not significantly alter the protein's chemical properties. In addition, see On the Cutting Edge: Selenocysteine, the 21st Amino Acid.

KEY POINTS

- *The wobble hypothesis explains how a single tRNA can respond to two or more codons.*
- *Some suppressor mutations alter the anticodons of tRNAs so that the mutant tRNAs recognize chain-termination codons and insert amino acids in response to their presence in mRNA molecules.*

ON THE CUTTING EDGE

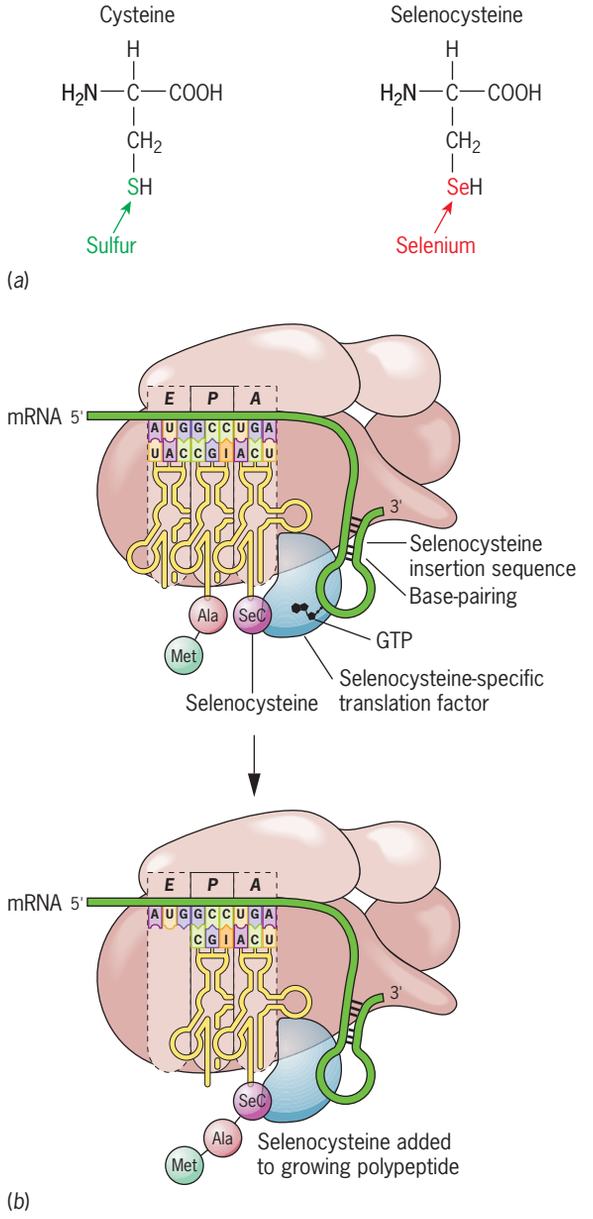
SELENOCYSTEINE, THE 21ST AMINO ACID

The structures of the 20 amino acids that are found in most proteins are shown in Figure 12.1, and the codons that specify each of these amino acids and the initiation and termination of polypeptide chains are shown in Table 12.1. However, in a few proteins, there is another amino acid—**selenocysteine**—that is specified by the genetic code at the time of translation. Selenocysteine contains the essential trace element selenium (atomic number 34) in place of the sulfur group in cysteine (■ **Figure 1*a***). When present in proteins—called **selenoproteins**—the reactive selenium is usually present at the active site and participates in oxidation/reduction (hydrogen removal/addition) reactions. Selenoproteins play important metabolic roles in all living organisms—prokaryotes, eukaryotes, and archaea.

Selenocysteine is incorporated into polypeptides during translation in response to the codon UGA, which normally functions as a chain-termination signal. The mRNAs encoding selenoproteins contain special selenocysteine insertion sequences (SECIS elements) that interact with specific translation factors leading to the incorporation of selenocysteine into polypeptides in response to UGA codons (■ **Figure 1*b***). The structures and locations of SECIS elements vary among procaryotes, archaea, and eukaryotes. However, in all cases, the SECIS elements form hairpin-like structures similar to those in tRNAs by intrastrand hydrogen bonding. In eukaryotes, these elements are located in the 3′-untranslated regions of mRNAs.

Selenocysteine has its own tRNA with a 5′-UCA-3′ anticodon and a unique hairpin domain. This tRNA is activated by the addition of serine, which is then converted to selenocysteine. During the translation of mRNAs encoding selenoproteins, the selenocysteyl-tRNAs respond to UGA codons with the aid of a selenocysteine-specific translation factor (Figure 1*b*). This selenocysteine-specific translation factor replaces elongation factor Tu during the entry of selenocysteine-tRNA into the A site on the ribosome. In the absence of selenium, translation of mRNAs encoding selenoproteins results in the synthesis of truncated polypeptides, with translation being terminated at the UGA selenocysteine codons. Thus, UGA codons in mRNAs lacking SECIS elements specify polypeptide chain termination, whereas UGA codons in mRNAs with downstream SECIS hairpin structures specify selenocysteine.

Are there other modified amino acids that are specified by codons during translation? So far, there is one other documented example, pyrrolysine—lysine with a pyrroline ring on the end of the side chain. Pyrrolysine is incorporated into polypeptides in some archaea and one bacterium, but not eukaryotes, in response to the codon UAG, which normally signals chain termination. The mechanism(s) by which UAG codons specify pyrrolysine incorporation rather than chain termination are still under investigation.



■ **FIGURE 1** (*a*) Comparison of the structures of cysteine and selenocysteine. (*b*) The incorporation of selenocysteine into a growing polypeptide in response to the codon UGA when a selenocysteine insertion sequence is present in the mRNA being translated.

Basic Exercises

Illustrate Basic Genetic Analysis

1. The human β-globin polypeptide is 146 amino acids long. How long is the coding portion of the human β-globin mRNA?

Answer: Each amino acid is specified by a codon containing three nucleotides. Therefore, the 146 amino acids in β-globin will be specified by 438 (146 × 3) nucleotides. However, a termination codon must be present at the end of the coding sequence, bringing the length to 438 + 3 = 441 nucleotides. In the case of β-globin and many other proteins, the amino-terminal methionine (specified by the initiation codon AUG) is removed from the β-globin during synthesis. Adding the initiation codon increases the coding sequence of the β-globin mRNA to 444 nucleotides (441 + 3).

2. If the coding segment of an mRNA with the sequence 5′-AUGUUUCCCAAAGGG-3′ is translated, what amino acid sequence will be produced?

Answer: (Amino-terminus)-methionine-phenylalanine-proline-lysine-glycine-(carboxyl-terminus). The amino acid sequence is deduced using the genetic code shown in **Table 12.1**. AUG is the methionine initiation codon followed by the phenylalanine codon UUU, the proline codon CCC, the lysine codon AAA, and the glycine codon GGG.

3. If a coding segment of the template strand of a gene (DNA) has the sequence 3′-TACAAAGGGTTTCCC-5′, what amino acid sequence will be produced if it is transcribed and translated?

Answer: The mRNA sequence produced by transcription of this segment of the gene will be 5′-AUGUUUCCCAAAGGG-3′. Note that this mRNA has the same nucleotide sequence as the one discussed in Exercise 2. Thus, it will produce the same peptide when translated: NH_2-Met-Phe-Pro-Lys-Gly-COOH.

4. What sequence of nucleotide pairs in a gene in *Drosophila* will encode the amino acid sequence methionine-tryptophan (reading from the amino terminus to the carboxyl terminus)?

Answer: The codons for methionine and tryptophan are AUG and UGG, respectively. Thus, the nucleotide sequence in the mRNA specifying the dipeptide sequence methionine-tryptophan must be 5′-AUGUGG-3′. The template DNA strand must be complementary and antiparallel to the mRNA sequence (3′-TACACC-5′), and the other strand of DNA must be complementary to the template strand. Therefore, the sequence of base pairs in the gene must be:

5′-ATGTGG-3′
3′-TACACC-5′

5. A wild-type gene contains the trinucleotide-pair sequence:

5′-GAG-3′
3′-CTC-5′

This triplet specifies the amino acid glutamic acid. If the second base pair in this gene segment were to change from A:T to T:A, yielding the following DNA sequence:

5′-GTG-3′
3′-CAC-5′

would it still encode glutamic acid?

Answer: No, it would now specify the amino acid valine. The codon for glutamic acid is 5′-GAG-3′, which tells us that the bottom strand of DNA is the template strand. Transcription of the wild-type gene yields the mRNA sequence 5′-GAG-3′, which is a glutamic acid codon. Transcription of the altered gene produces the mRNA sequence 5′-GUG-3′, which is a valine codon. Indeed, this is exactly the same nucleotide-pair change that gave rise to the altered hemoglobin in Herrick's sickle-cell anemia patient, discussed at the beginning of this chapter. See Figure 1.9 for further details.

Testing Your Knowledge

Integrate Different Concepts and Techniques

1. The average mass of the 20 common amino acids is about 137 daltons. Estimate the approximate length of an mRNA molecule that encodes a polypeptide with a mass of 65,760 daltons. Assume that the polypeptide contains equal amounts of all 20 amino acids.

Answer: Based on this assumption, the polypeptide would contain about 480 amino acids (65,760 daltons/137 daltons per amino acid). Since each codon contains three nucleotides, the coding region of the mRNA would have to be 1440 nucleotides long (480 amino acids × 3 nucleotides per amino acid).

2. The antibiotic streptomycin kills sensitive *E. coli* by inhibiting the binding of $tRNA_f^{Met}$ to the *P* site of the ribosome and by causing misreading of codons in mRNA. In sensitive bacteria, streptomycin is bound by protein S12 in the 30S subunit of the ribosome. Resistance to streptomycin can result from a mutation in the gene-encoding protein S12 so that the altered protein will no longer bind the antibiotic. In 1964, Luigi Gorini and Eva Kataja isolated mutants of *E. coli* that grew on minimal medium supplemented with either the amino acid arginine or streptomycin. That is, in

the absence of streptomycin, the mutants behaved like typical arginine-requiring bacteria. However, in the absence of arginine, they were streptomycin-dependent conditional-lethal mutants. That is, they grew in the presence of streptomycin but not in the absence of streptomycin. Explain the results obtained by Gorini and Kataja.

Answer: The streptomycin-dependent conditional-lethal mutants isolated by Gorini and Kataja contained missense mutations in genes encoding arginine biosynthetic enzymes. If arginine was present in the medium, these enzymes were unessential. However, these enzymes were required for growth in the absence of arginine (one of the 20 amino acids required for protein synthesis).

Streptomycin causes misreading of mRNA codons in bacteria. This misreading allowed the codons that contained the missense mutations to be translated ambiguously—with the wrong amino acids incorporated—when the antibiotic was present. When streptomycin was present in the mutant bacteria, an amino acid occasionally would be inserted (at the site of the mutation) that resulted in an active enzyme, which, in turn, allowed the cells to grow, albeit slowly. In the absence of streptomycin, no misreading occurred, and all of the mutant polypeptides were inactive.

Questions and Problems

Enhance Understanding and Develop Analytical Skills

12.1 In a general way, describe the molecular organization of proteins and distinguish proteins from DNA, chemically and functionally. Why is the synthesis of proteins of particular interest to geneticists?

12.2 At what locations in the cell does protein synthesis occur?

12.3 Is the number of potential alleles of a gene directly related to the number of nucleotide pairs in the gene? Is such a relationship more likely to occur in prokaryotes or in eukaryotes? Why?

12.4 Why was it necessary to modify Beadle and Tatum's one gene–one enzyme concept of the gene to one gene–one polypeptide?

12.5 (a) Why is the genetic code a triplet code instead of a singlet or doublet code? (b) How many different amino acids are specified by the genetic code? (c) How many different amino acid sequences are possible in a polypeptide 146 amino acids long?

12.6 What types of experimental evidence were used to decipher the genetic code?

12.7 In what sense and to what extent is the genetic code (a) degenerate, (b) ordered, and (c) universal?

12.8 GO The thymine analog 5-bromouracil is a chemical mutagen that induces single base-pair substitutions in DNA called transitions (substitutions of one purine for another purine and one pyrimidine for another pyrimidine). Using the known nature of the genetic code (Table 12.1), which of the following amino acid substitutions should you expect to be induced by 5-bromouracil with the highest frequency:

(a) Met → Val;
(b) Met → Leu;
(c) Lys → Thr;
(d) Lys → Gln;
(e) Pro → Arg; or
(f) Pro → Gln? Why?

12.9 Using the information given in Problem 12.8, would you expect 5-bromouracil to induce a higher frequency of His → Arg or His → Pro substitutions? Why?

12.10 What is the minimum number of tRNAs required to recognize the six codons specifying the amino acid leucine?

12.11 Characterize ribosomes in general as to size, location, function, and macromolecular composition.

12.12 (a) Where in the cells of higher organisms do ribosomes originate? (b) Where in the cells are ribosomes most active in protein synthesis?

12.13 Identify three different types of RNA that are involved in translation and list the characteristics and functions of each.

12.14 (a) How is messenger RNA related to polysome formation? (b) How does rRNA differ from mRNA and tRNA in specificity? (c) How does the tRNA molecule differ from that of DNA and mRNA in size and helical arrangement?

12.15 Outline the process of aminoacyl-tRNA formation.

12.16 How is translation (a) initiated and (b) terminated?

12.17 Of what significance is the wobble hypothesis?

12.18 GO If the average molecular mass of an amino acid is assumed to be 100 daltons, about how many nucleotides will be present in an mRNA coding sequence specifying a single polypeptide with a molecular mass of 27,000 daltons?

12.19 The bases A, G, U, C, I (inosine) all occur at the 5′ positions of anticodons in tRNAs.

(a) Which base can pair with three different bases at the 3′ positions of codons in mRNA?
(b) What is the minimum number of tRNAs required to recognize all codons of amino acids specified by codons with complete degeneracy?

12.20 GO Assume that in the year 2025, the first expedition of humans to Mars discovers several Martian life forms thriving in hydrothermal vents that exist below the planet's surface. Several teams of molecular biologists extract proteins and nucleic acids from these organisms and make some momentous discoveries. Their first discovery is that the proteins in Martian life forms contain only 14 different amino acids instead of the 20 present in life forms on Earth. Their second discovery is that the DNA and RNA in these organisms have only two different nucleotides instead of the four nucleotides present in living organisms on Earth. (a) Assuming that transcription and translation work similarly in Martians and Earthlings, what is the minimum number of nucleotides that must be present in the Martian codon to specify all the amino acids in Martians? (b) Assuming that the Martian code proposed above has translational start-and-stop signals, would you expect the Martian genetic code to be degenerate like the genetic code used on Earth?

12.21 What are the basic differences between translation in prokaryotes and translation in eukaryotes?

12.22 What is the function of each of the following components of the protein-synthesizing apparatus:

(a) aminoacyl-tRNA synthetase,
(b) release factor 1,
(c) peptidyl transferase,
(d) initiation factors,
(e) elongation factor G?

12.23 An *E. coli* gene has been isolated and shown to be 68 nm long. What is the maximum number of amino acids that this gene could encode?

12.24 (a) What is the difference between a nonsense mutation and a missense mutation? (b) Are nonsense or missense mutations more frequent in living organisms? (c) Why?

12.25 The human α-globin chain is 141 amino acids long. How many nucleotides in mRNA are required to encode human α-globin?

12.26 What are the functions of the *A*, *P*, and *E* aminoacyl-tRNA binding sites on the ribosome?

12.27 (a) In what ways does the order in the genetic code minimize mutational lethality? (b) Why do base-pair changes that cause the substitution of a leucine for a valine in the polypeptide gene product seldom produce a mutant phenotype?

12.28 (a) What is the function of the Shine-Dalgarno sequence in prokaryotic mRNAs? (b) What effect does the deletion of the Shine-Dalgarno sequence from an mRNA have on its translation?

12.29 (a) In what ways are ribosomes and spliceosomes similar? (b) In what ways are they different?

12.30 The 5′ terminus of a human mRNA has the following sequence:

5′cap-GAAGAGACAAGGTCAUGGCCAU-AUGCUUGUUCCAAUCGUUAGCUGCGCAG-GAUCGCCCUGGG.....3′

When this mRNA is translated, what amino acid sequence will be specified by this portion of the mRNA?

12.31 A partial (5′ subterminal) nucleotide sequence of a prokaryotic mRNA is as follows:

5′-.....AGGAGGCUCGAACAUGUCAAUAUGC-UUGUUCCAAUCGUUAGCUGCGCAGGACCGU-CCCGGA.....3′

When this mRNA is translated, what amino acid sequence will be specified by this portion of the mRNA?

12.32 GO The following DNA sequence occurs in the nontemplate strand of a structural gene in a bacterium (the promoter sequence is located to the left but is not shown):

5′-GAATGTCAGAACTGCCATGCTTCATA↓TGAA-TAGACCTCTAG-3′

(a) What is the ribonucleotide sequence of the mRNA molecule that is transcribed from this piece of DNA?
(b) What is the amino acid sequence of the polypeptide encoded by this mRNA?
(c) If the nucleotide indicated by the arrow undergoes a mutation that changes T to A, what will be the resulting amino acid sequence following transcription and translation?

12.33 Alan Garen extensively studied a particular nonsense (chain-termination) mutation in the alkaline phosphatase gene of *E. coli*. This mutation resulted in the termination of the alkaline phosphatase polypeptide chain at a position where the amino acid tryptophan occurred in the wild-type polypeptide. Garen induced revertants (in this case, mutations altering the same codon) of this mutant with chemical mutagens that induced single base-pair substitutions and sequenced the polypeptides in the revertants. Seven different types of revertants were found, each with a different amino acid at the tryptophan position of the wild-type polypeptide (termination position of the mutant polypeptide fragment). The amino acids present at this position in the various revertants included tryptophan, serine, tyrosine, leucine, glutamic acid, glutamine, and lysine. Did the nonsense mutation studied by Garen contain a UAG, a UAA, or a UGA nonsense mutation? Explain the basis of your deduction.

12.34 The following DNA sequence occurs in a bacterium (the promoter sequence is located to the left but is not shown).

5′-CAATCATGGACTGCCATGCTTCATA↓TGAATAGTTGACAT-3′
3′-GTTAGTACCTGACGGTACGAAGTATACTTATCAACTGTA-5

(a) What is the ribonucleotide sequence of the mRNA molecule that is transcribed from the template strand of this

piece of DNA? Assume that both translational start and termination codons are present.

(b) What is the amino acid sequence of the polypetide encoded by this mRNA?

(c) If the nucleotide indicated by the arrow undergoes a mutation that causes this C:G base pair to be deleted, what will be the polypeptide encoded by the mutant gene?

Genomics on the Web at http://www.ncbi.nlm.nih.gov

The genetic code is degenerate, with two to six codons specifying each of the amino acids except for methionine and tryptophan.

1. Are all of the codons specifying a given amino acid used with equal frequency, or are some codons used more frequently than others? For example, the codons UUA, UUG, CUU, CUC, CUA, and CUG all specify leucine. Are these six leucine codons present with equal frequency in the coding regions of mRNAs?
2. Are the six codons specifying leucine used with equal frequency in mRNAs transcribed from human nuclear genes? From human mitochondrial genes? Are these codons used at the same frequency in nuclear and mitochondrial genes?
3. Are the leucine codons used at about the same frequencies in different species, for example, in humans and *E. coli* cells? Is there any bias in codon usage (preferred use of specific codons) related to the AT/GC content of the genomes of different species?

Hint: A search of the databases at the NCBI web site will yield an overwhelming amount of information. In this case, more accessible information can be obtained at the http://www.kazusa.jp/codon web site, which summarizes data on codon usage in 35,799 organisms (many viruses). These data are compiled from NCBI-GenBank File Release 160.0 (June 15, 2007). In the Query Box, type *Homo sapiens* and click "Submit." Your search will yield two results: (1) mitochondrion *Homo sapiens* and *Homo sapiens*. Clicking the first will give you a table of codon usage in human mitochondria, and clicking the second will give you a table of codon use in mRNAs encoded by nuclear genes. You can obtain codon usage data for *E. coli* and other species of interest by simply typing the species name in the Query Box.