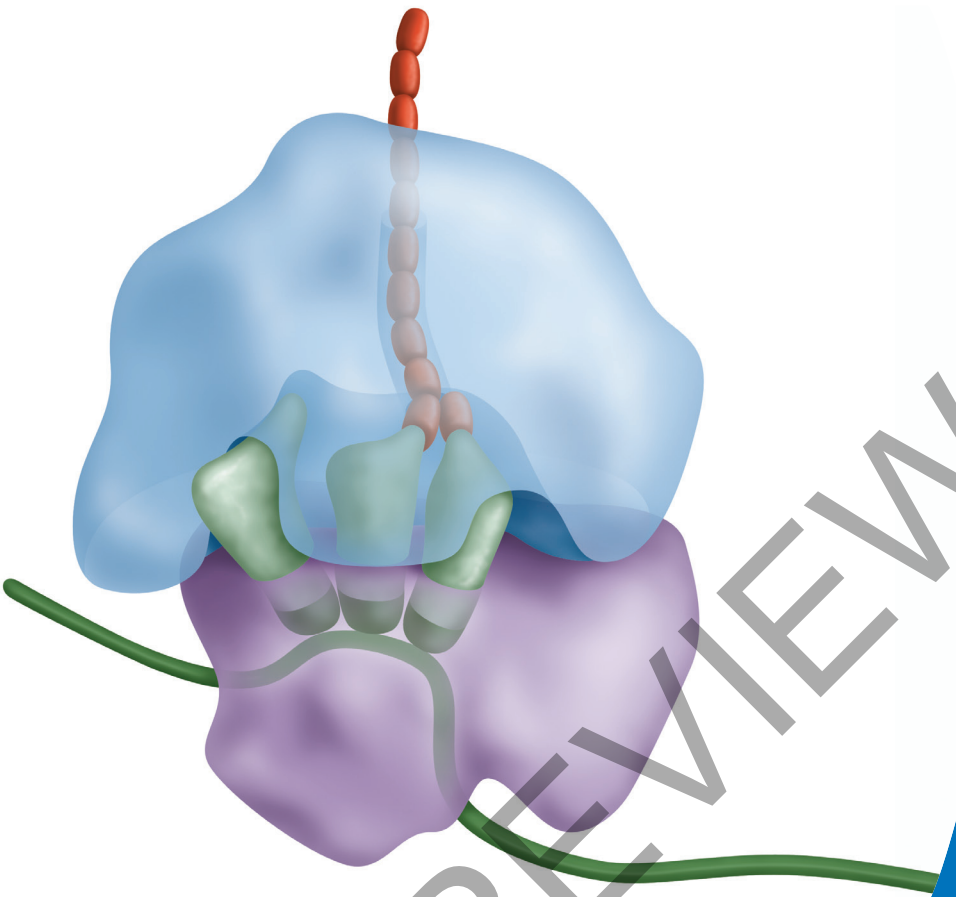
# The Molecular Biology of Translation

9



Ribosomes use codon sequences of messenger RNA to direct the assembly of polypeptides during translation. This rendering of a ribosome engaged in translation is based on recent crystal structure analysis and accurately shows the large subunit (top) and small subunit (bottom), the track of mRNA through the small subunit, the spaces for E, P, and A sites into which tRNAs fit, and the egress of the polypeptide through the large subunit.

CHAPTER OUTLINE


9.1 Polypeptides Are Composed of Amino Acid Chains That Are Assembled at Ribosomes
9.2 Translation Occurs in Three Phases
9.3 Translation Is Fast and Efficient
9.4 The Genetic Code Translates Messenger RNA into Polypeptide
9.5 Experiments Deciphered the Genetic Code
9.6 Translation Is Followed by Polypeptide Folding, Processing, and Protein Sorting


## ESSENTIAL IDEAS

- Translation is the cellular process of polypeptide production carried out by ribosomes under the direction of mRNA.
- Ribosomes assemble on mRNA and initiate translation at the start codon.
- Polypeptide elongation and termination are similar in bacteria and eukaryotes.
- Transfer RNA molecules carry amino acids to ribosomes, which assemble polypeptides with the aid of ribosomal proteins.
- A virtually universal genetic code comprising 64 mRNA codons directs polypeptide assembly.
- Polypeptides undergo posttranslational folding and processing, and in eukaryotes are sorted into vesicles for transport to cellular destinations or for secretion.

Long before the discovery that DNA is the hereditary molecule, biologists had established the relationship between genes and proteins. In 1902, Archibald Garrod was the first to explicitly draw this connection when he proposed that the human hereditary disorder alkaptonuria was caused by an inherited defect in the enzyme homogentisic acid oxidase (see Section 4.3 and Figure 4.17b). As Garrod and other biologists expanded their exploration of the gene–protein connection, they found evidence that hereditary variation was closely tied to variations in proteins. Principal

among the biologists who developed this connection were George Beadle and Edward Tatum, whose research established the "one gene–one enzyme" hypothesis (Chapter 5).

This chapter discusses translation, the mechanism by which the messenger RNA (mRNA) transcripts of genes are used to assemble amino acids into polypeptide strings that form proteins. Translation is carried out by ribosomes that bring together mRNA transcripts and transfer RNA (tRNA) molecules that carry amino acids and facilitate the assembly of polypeptides, strings of amino acids.

Polypeptides form the enzymes (catalytic proteins), structural proteins, transport proteins, signaling proteins, hormones, and other components that are assembled into cell structures and that perform biological activities in cells. Your body is composed of trillions of cells that collectively express and utilize tens of thousands of different polypeptides, all synthesized by translation.

The story of how polypeptides are produced by translation, and the story of how scientists came to understand the process, offers intriguing insight into the design of molecular genetic experiments. In this chapter, we describe some of these experiments and examine the molecular biology of translation. We look at the homology of proteins that are active in translation in organisms from the three domains of life and describe how this and other features of translation are evidence of a single origin of life and of the evolutionary relationships between bacteria, archaea, and eukaryotes. In the final chapter section, we discuss posttranslational processes that are instrumental in producing functional proteins and guiding them to their appropriate destinations in cells. The chapter concludes with a case study describing the action of commonly used antibiotics that interfere with bacterial translation.

## 9.1 Polypeptides Are Composed of Amino Acid Chains That Are Assembled at Ribosomes

Twenty different amino acids are the basic building blocks of polypeptides. All amino acids have features in common and features that are distinct. The distinctive features impart specific characteristics that allow the amino acid to participate in certain chemical reactions or behave in a hydrophilic or hydrophobic manner. In part, the common features allow amino acids to be joined into polypeptides by covalent bond formation between adjacent amino acids in the chain.

### Amino Acid Structure

The shared features of amino acids are a central carbon molecule known as the α-carbon, an amino ($NH_3$) group, and a carboxyl (COOH) group (**Figure 9.1**). Each amino and carboxyl group is joined to the α-carbon. During polypeptide assembly, an enzyme in the ribosome catalyzes the formation of a **peptide bond** between the carboxyl group of one amino acid and the amino group of the next amino acid in the chain. Each amino acid added in this way becomes a new monomer in the growing polymer that is the elongating polypeptide. The term **polypeptide** identifies a string of amino acids that are joined by peptide bonds. Each *protein* has a unique sequence of amino acids, may be composed of one or more polypeptide chains, and generally have a characteristic three-dimensional structure.

The distinctive portion of each amino acid is its side chain, known as an **R-group,** that is joined to the α-carbon. The R-groups range in complexity from a single hydrogen atom to ringed structures that in themselves contain multiple carbon atoms. Each R-group imparts specific characteristics as shown in **Table 9.1**. Ten of the amino acids have nonpolar R-groups, meaning that they have no charged atoms that can participate in formation of hydrogen bonds with other amino acids. Five other amino acids have polar R-groups that can carry partial

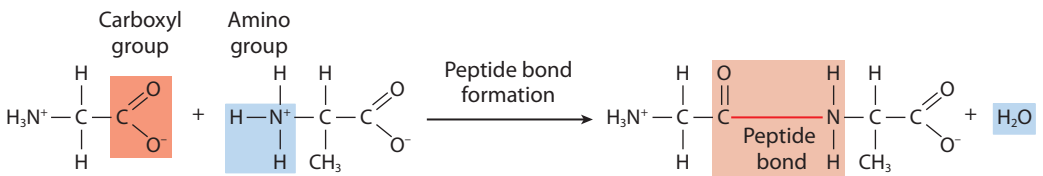



**Figure 9.1 Peptide bond formation.** The carboxyl group of one amino acid reacts with the amino group of a second amino acid to form a covalent peptide bond that joins amino acids in a polypeptide.

**Table 9.1** Amino Acids Grouped by Their Side Chain Properties

| | |
|---|---|
| Nonpolar side chains: Have no charged or electronegative atoms at pH 7.0 to form hydrogen bonds. | |
| Alanine (Ala or A) | Methionine (Met or M) |
| Cysteine (Cys or C) | Phenylalanine (Phe or F) |
| Glycine (Gly or G) | Proline (Pro or P) |
| Isoleucine (Ile or I) | Tryptophan (Trp or W) |
| Leucine (Leu or L) | Valine (Val or V) |
| Polar side chains: Have partial charges at pH 7.0 and can form hydrogen bonds. | |
| Asparagine (Asp or N) | Threonine (Thr or T) |
| Glutamine (Glu or Q) | Tyrosine (Tyr or Y) |
| Serine (Ser or S) | |
| Electrically charged side chains: At pH 7.0, can form hydrogen and ionic bonds. | |
| **Basic Side Chains** | **Acidic Side Chains** |
| Arginine (Arg or R) | Aspartate (Asp or D) |
| Histidine (His or H) | Glutamate (Glu or E) |
| Lysine (Lys or K) | |

charges and can participate in hydrogen bond formation with other amino acids. The five remaining amino acids have electrically charged R-groups: Three are basic and two are acidic. Electrically charged R-groups allow these amino acids to form ionic bonds and hydrogen bonds.

## Polypeptide and Transcript Structure

Polypeptide assembly is orchestrated by ribosomes, which are ribonucleoprotein "machines" containing multiple molecules of ribosomal RNA (rRNA) and dozens of proteins. Ribosomes of all organisms are composed of two subunits that assemble into a ribosome as translation begins. Ribosomes bind mRNA and provide an environment for complementary base pairing between mRNA codon sequences and the anticodon sequences of tRNA. (In Chapter 1 and Figure 1.11, we review these basic mechanical features of translation.) **Figure 9.2** encapsulates the essential elements of translation. Ribosomes translate mRNA in the 5′ → to 3′ direction, beginning with the start codon and ending with a stop codon. At each triplet codon, complementary base pairing between mRNA and tRNA determines which amino acid is added to the nascent (growing) polypeptide. The start codon and stop codon define the boundaries of the translated segment of mRNA. The resulting polypeptides have an N-terminal (amino-terminal) end corresponding to the 5′ end of mRNA and a C-terminal (carboxyl-terminal) end that corresponds to the 3′ end of mRNA (**Figure 9.3**).

Figure 9.3 identifies two segments of the mRNA transcript that do not undergo translation. Between the 5′ end of mRNA and the start codon is a segment known as the

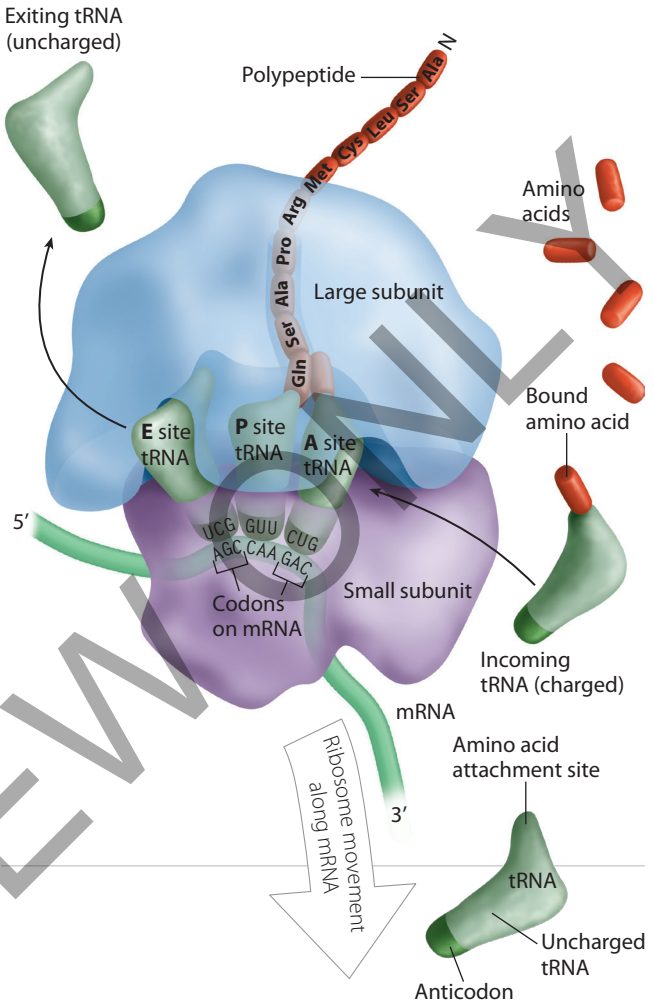



**Figure 9.2** **Translation overview.**

**5′ untranslated region,** abbreviated **5′ UTR.** The region between the stop codon and the 3′ end of the molecule is the **3′ untranslated region,** or **3′ UTR.** The 5′ UTR contains sequences that help initiate translation and the 3′ UTR contains sequences associated with transcription termination in almost all bacterial and eukaryotic mRNAs. By comparison,

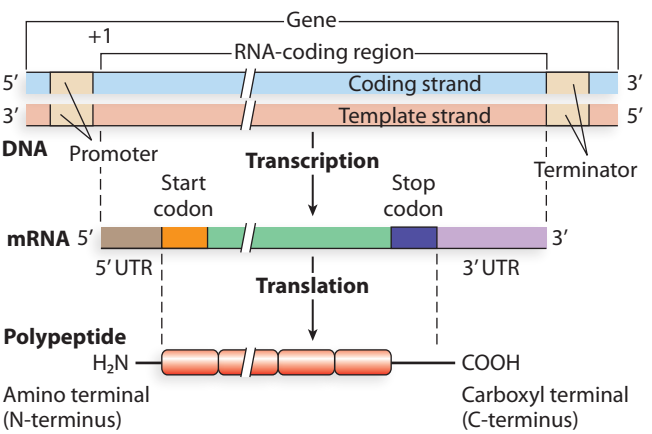



**Figure 9.3** **Alignment of DNA, mRNA, and polypeptide.**

relatively little is known about the roles of archaeal 5′ and 3′ UTRs. Many archaeal mRNAs have a 5′ UTR that functions similarly to those of bacteria and eukaryotes. However, a substantial proportion of archaeal mRNAs—some studies suggest 50% or more of them—do not have a 5′ UTR. These so-called "leaderless" mRNAs are still efficiently translated, but the details of the mechanism remain unclear. It has been proposed that archaeal leaderless mRNAs could perhaps be a relic of an ancestral mode of translation.

Polypeptides have four levels of organization that each describe an aspect of their underlying structure (Table 9.2). The polypeptide **primary structure** is the sequence of amino acids contained in the polypeptide. The order of amino acids and the length of a polypeptide (the number of amino acids it contains) are effectively limitless. There are billions of possible amino acid sequence options even among short polypeptides of 20 amino acids or less. The specific order of amino acids is, however, critical to the proper function of a polypeptide. The R-groups of amino acids affect the solubility and reactivity of amino acids, and therefore they affect the functional properties of the polypeptide.

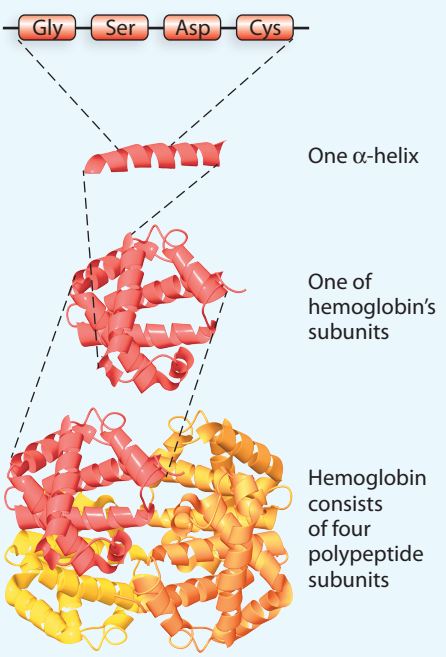
Polypeptide **secondary structure** is generated by hydrogen bonds that form between amino acids. Hydrogen bond formation requires that amino acids with polar R-groups align with one another. This is accomplished by bending or twisting the polypeptide in one of two possible structures. An **α-helix (alpha helix)** is a twisted coil of amino acids stabilized by hydrogen bonds between partially charged R-groups. A **β-pleated sheet (beta-pleated sheet)** is a 180-degree bend created when a segment of a polypeptide folds. The primary structure is critical to determining which, if either, of these secondary structures forms in a polypeptide.

A polypeptide's **tertiary structure** is the result of a variety of interactions involving the R-groups. Interactions such as hydrogen bonding, covalent bonding, ionic interactions, and hydrophobic interactions produce the overall shape of the protein. Tertiary structure is dependent on primary and secondary structure, and it should come as no surprise that protein shapes vary widely. These shapes form the binding, interaction, and catalytic domains that are responsible for the protein's action in the body. The tertiary structure of a protein may change in response to the presence of other chemical substances, including other protein molecules. For example, an enzyme may have a catalytically active tertiary structure under some circumstances and have an alternative, nonactive tertiary structure under others.

Primary, secondary, and tertiary structures describe different levels of organization of individual polypeptides. But some proteins contain two or more polypeptides, an organization described as **quaternary structure.** Proteins that have a quaternary structure contain distinct polypeptides that each have their own primary, secondary, and tertiary structures. Such proteins are often described as *multimers*. The individual polypeptides of a multimer may be identical or may be different. For example, a protein composed of four identical polypeptides can be called a *homotetramer*, and a four-polypeptide protein that contains two or more different polypeptides can be identified as a *heterotetramer*. Table 9.2 summarizes these four levels of polypeptide structure and illustrates the red blood cell

**Table 9.2 Polypeptide Structure**

| Level | Description | Stabilized by | Example: Hemoglobin |
|---|---|---|---|
| Primary | The sequence of amino acids in a polypeptide | Peptide bonds | |
| Secondary | Formation of α-helices and β-pleated sheets in a polypeptide (thus, depends on primary structures) | Hydrogen bonding between groups along the peptide-bonded backbone. | |
| Tertiary | Overall three-dimensional shape of a polypeptide (includes contribution from secondary structures) | Bonds and other interactions between R-groups, or between R-groups and the peptide-bonded backbone. | |
| Quaternary | Shape produced by combinations of polypeptides (each with its own tertiary structure) | Bonds and other interactions between R-groups, and between peptide backbones of different polypeptides. | |



protein hemoglobin—a heterotetramer—as an example of a protein with a quaternary structure. Hemoglobin and a specific variant of one of the polypeptides in this heterotetramer are the focus of discussion in Chapter 10.

## Ribosome Structures

The specific molecules composing bacterial, archaeal, and eukaryotic ribosomes differ, but the overall structures and functions of the ribosomes are similar, reflecting the fundamental nature of the translation process in all forms of life. In all three domains, ribosomes perform three essential tasks:

1. Bind messenger RNA and identify the start codon where translation begins.
2. Facilitate the complementary base pairing of mRNA codons and tRNA anticodons that determines amino acid order in the polypeptide.
3. Catalyze peptide bond formation between amino acids during polypeptide formation.

Differences in ribosomal composition between bacteria, archaea, and eukaryotes include the number and sequence of rRNA molecules and the number and type of ribosomal proteins. Although the archaeal and bacterial ribosomes are similar in size, and somewhat smaller than the eukaryotic ribosomes, most of the archaeal ribosomal proteins (and the tRNAs and protein factors involved in translation) display homology to their eukaryotic counterparts. In all three domains, ribosomes display key structural similarities that are divided into two main subunits, called the **large ribosomal subunit** and the **small ribosomal subunit.** By convention, subunit size is measured in Svedberg units (S), which describe the velocity of their sedimentation when subjected to a centrifugal force. Named in honor of Theodor Svedberg, a 1926 Nobel Laureate in Chemistry and inventor of the ultracentrifuge, higher S values indicate faster sedimentation rates and larger molecules. It should be noted that Svedberg units are not additive when ribosomal subunits are combined because sedimentation is a composite property that is affected by multiple molecular factors, including size, shape, and hydration state.

The ribosomes of *E. coli* are the most thoroughly studied bacterial ribosomes and serve as a model for general ribosome structure (Figure 9.4a). The small subunit of bacterial ribosomes has a Svedberg value of 30S. It contains 21 proteins and a single 16S rRNA composed of 1541 nucleotides. The large subunit of the bacterial ribosome is a 50S particle composed of 32 proteins, a small 5S rRNA containing 120 nucleotides, and a large 23S rRNA containing 2904 nucleotides. When fully assembled, the intact bacterial ribosome has a Svedberg value of 70S.

Both the large and small subunits contribute to the formation of three regions that play important functional roles during translation: the **peptidyl site,** or **P site,** the **aminoacyl site,** or **A site** and, the **exit site,** or **E site.** The P site holds a tRNA to which the nascent polypeptide is attached. The A site binds a new tRNA molecule carrying the next amino acid to be added to the polypeptide. The E site provides an avenue of egress for tRNAs as they leave the ribosome after their amino acid has been added to the polypeptide chain. Ribosomes also form a channel through which the polypeptide emerges. In addition, there is a channel in the large subunit through which the nascent polypeptide is extruded from the ribosome (see Figure 9.2).

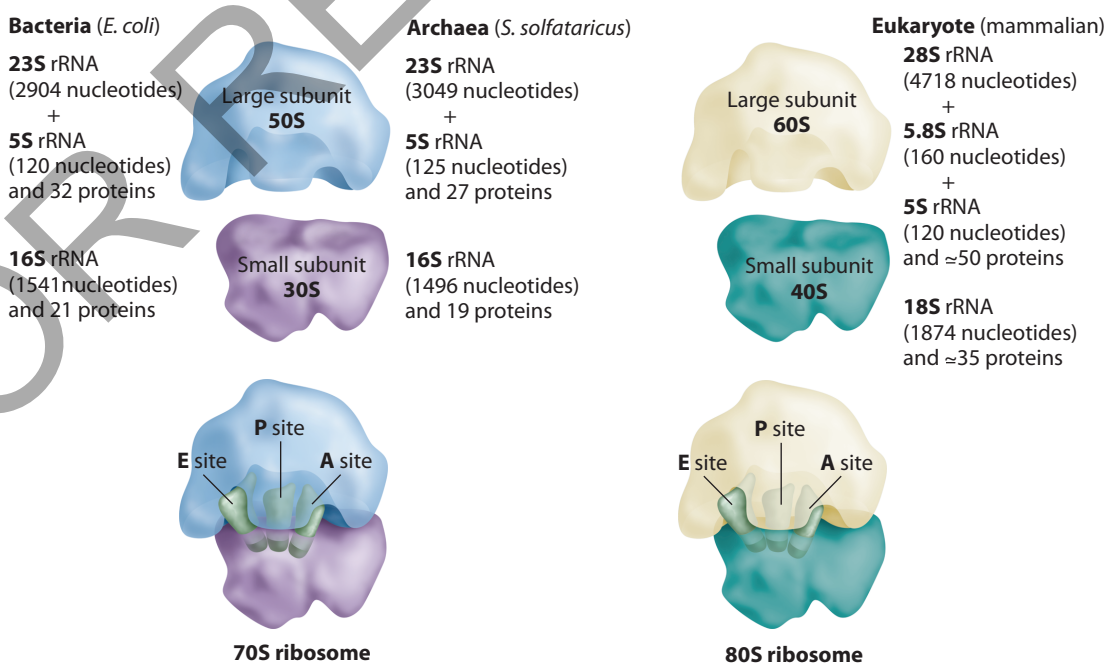



**Figure 9.4** **Ribosomes of bacteria, archaea, and eukaryotes.** **(a)** The best-studied bacteria ribosome is that of *E. coli*, and the best-described archaeal ribosome is that of *Haloarcula marismortui*. **(b)** The best-studied eukaryotic ribosomes are mammalian.

Among eukaryotes, mammalian ribosomes are the most fully characterized (**Figure 9.4b**). The small 40S ribosomal subunit contains approximately 35 proteins and a single 18S rRNA composed of 1874 nucleotides. The large mammalian ribosomal subunit has a Svedberg value of 60S and contains 45 to 50 proteins, along with three molecules of rRNA. The rRNA molecules have values of 5S (120 nucleotides), 5.8S (160 nucleotides), and 28S (4718 nucleotides). The intact mammalian ribosome has a Svedberg value of 80S. Like the bacterial ribosome, the intact mammalian ribosome possesses a P site, an A site, an E site, and a channel for polypeptide egress.

The ribosomes of archaeal species have not been studied nearly as fully as those of bacterial and eukaryotes, but some information is available. The first atomic crystal structure of the large ribosomal subunit of an archaeon was that of *Haloarcula marismortui*. This structure included a 23S and a 5S rRNA and 27 proteins. Follow-up analysis of the small subunit structure revealed a 16S rRNA and 19 proteins. This is the basis for the conclusion that archaeal ribosomes have an overall size and structure similar to that of the 70S bacterial ribosome. As we discuss later, however, archaeal tRNAs and translation proteins are similar to those in eukaryotes.

The proteins contained in ribosomal subunits can be separated from one another by a specialized type of electrophoresis called two-dimensional gel electrophoresis. The 21 proteins that are part of the small ribosomal subunit in *E. coli* and the 31 proteins found in the large ribosomal subunit are efficiently separated by this method. **Research Technique 9.1** describes how two-dimensional

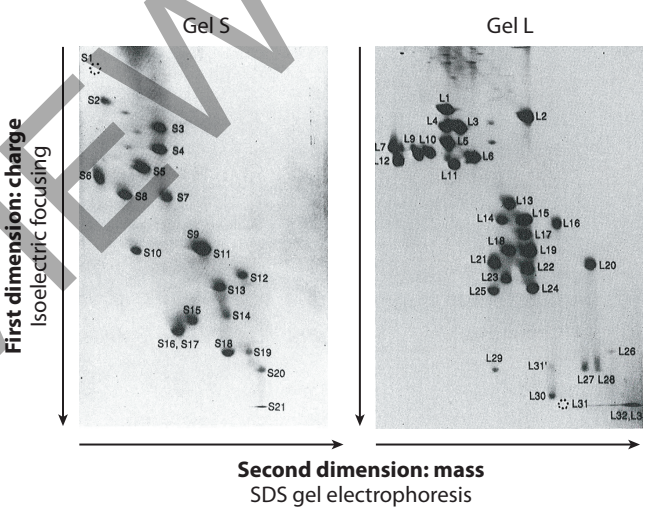
## Research Technique 9.1

### Two-Dimensional Gel Electrophoresis and the Identification of Ribosomal Proteins

**PURPOSE** All ribosomes are composed of two subunits that are each a complex mixture of rRNA and dozens of proteins. One approach to determining the number of proteins contained in each ribosomal subunit uses a method of electrophoresis known as two-dimensional gel electrophoresis to separate the proteins by their charge in the first dimension and then by their mass in the second dimension. Two-dimensional gel electrophoresis produces a distinctive "protein fingerprint" that distributes each ribosomal protein to a different location in the two-dimensional gel.

**MATERIALS AND PROCEDURES** Ribosomes are isolated from cells, the subunits are separated, and the subunits are treated to dissociate the proteins they contain. The mixture containing liberated ribosomal proteins is then separated in the first dimension by a version of gel electrophoresis known as isoelectric focusing. In this procedure, proteins are separated exclusively by their charge. In contrast to conventional gel electrophoresis, which uses a buffered solution to maintain constant pH throughout the gel, isoelectric focusing gels contain a pH gradient. A protein's pH environment affects its charge, and every protein has a pH—called the isoelectric point—at which it has neutral charge and cannot move in an electrical field. In isoelectric focusing, proteins migrate through the pH gradient to their isoelectric point, where they stop.

Once isoelectric focusing is complete, protein separation takes place in the second dimension, which uses SDS (sodium dodecyl sulfate) gel electrophoresis. SDS is a strong anionic detergent that denatures proteins by disrupting the interactions that keep them folded. Denatured proteins migrate through the gel at a rate determined by their mass, that is, by the number of amino acids they contain. In the SDS gel dimension of two-dimensional gel electrophoresis, each protein has a unique starting point corresponding to its isoelectric point. Proteins with large mass (more amino acids) migrate a short distance in the second dimension, whereas proteins with small mass (fewer amino acids) migrate a greater distance.




**DESCRIPTION** A pair of two-dimensional electrophoresis gels, one containing proteins of the small subunit of the *E. coli* ribosome (gel S) and the other containing proteins of the large subunit (gel L), reveal protein spots (the protein fingerprint) corresponding to the positions of proteins that make up each ribosomal subunit. Each spot identifies the location of a unique protein that differs from the other proteins in the gel by a combination of charge and mass. The proteins in gel S are identified as S1 to S21, and in gel L as L1 to L32.

**CONCLUSION** Two-dimensional gel electrophoresis identifies 21 proteins in the small subunit of the *E. coli* ribosome and 32 proteins in the large ribosomal subunit. Each protein obtained by two-dimensional electrophoresis can be subjected to additional biochemical examination to specifically identify the protein and investigate its role in translation.

gel electrophoresis is used to characterize the proteins found in *E. coli* ribosomal subunits.

### A Three-Dimensional View of the Ribosome

Ribosomes are so small—a mere 25 nanometers (nm) in diameter—that almost 10,000 of them can fit in the same space as the period at the end of this sentence. No one has ever "seen" a ribosome, but powerful molecular imaging techniques can resolve the three-dimensional configuration of ribosomes and ribosomal subunits, at levels of resolution that are measured in ångströms (Å). These structural analyses have clarified how ribosomal subunits fit together, and have produced a detailed understanding of ribosomal interactions with mRNA and tRNA.

Structural analysis of ribosomes and other molecular complexes in cells is made possible by a technique known as cryo-electron microscopy (cryo-EM), pioneered by Robert Glaeser in the 1970s and perfected by Jacques Dubochet in the 1980s. Cryo-EM uses liquid nitrogen or liquid ethane, with temperatures nearly −200°C, to instantaneously freeze macromolecules and thus preserve them in their native state. A frozen macromolecule is then placed on a microcaliper and scanned from various angles by electron beams that collect data analyzed by specialized software to create a three-dimensional picture of molecular structure. Cryo-EM creates exquisitely precise three-dimensional images of ribosome structure—much like CAT-scan imaging of the human body—revealing atomic-level details of ribosome structure (Figure 9.5). These images have identified the location and dimensions of the E, A, and P sites, for example, and have clarified the mechanical activities of ribosomes during translation. This work was recognized with the 2009 Nobel Prize in Chemistry awarded to Ada Yonath, Thomas Steitz, and Venki Ramakrishnan.

## 9.2 Translation Occurs in Three Phases

Translation occurs in three phases: initiation, elongation, and termination. The three phases are generally similar in bacteria, archaea, and eukaryotes, and yet they differ in several ways, particularly during translation initiation, where distinct mechanisms are used to identify the start codon.

### Translation Initiation

Translation initiation in all organisms begins when the small ribosomal subunit binds near the 5′ end of mRNA and identifies the start codon sequence. In the next stage, the **initiator tRNA,** the tRNA carrying the first amino acid of the polypeptide, binds to the start codon. In the final stage of initiation, the large subunit joins the small subunit to form an intact ribosome, and translation begins. During these stages, *initiation factor proteins* help control ribosome formation and binding of the initiator tRNA, and guanosine triphosphate (GTP) provides energy. The tRNAs used during translation each carry a specific amino acid and are identified as **charged tRNAs.** In contrast, a tRNA without an amino acid is **uncharged.** Specialized enzymes discussed in a later section are responsible for recognizing different tRNAs and charging each one with the correct amino acid.

Starting translation at the authentic (correct) start codon is essential for translation of the correct polypeptide.

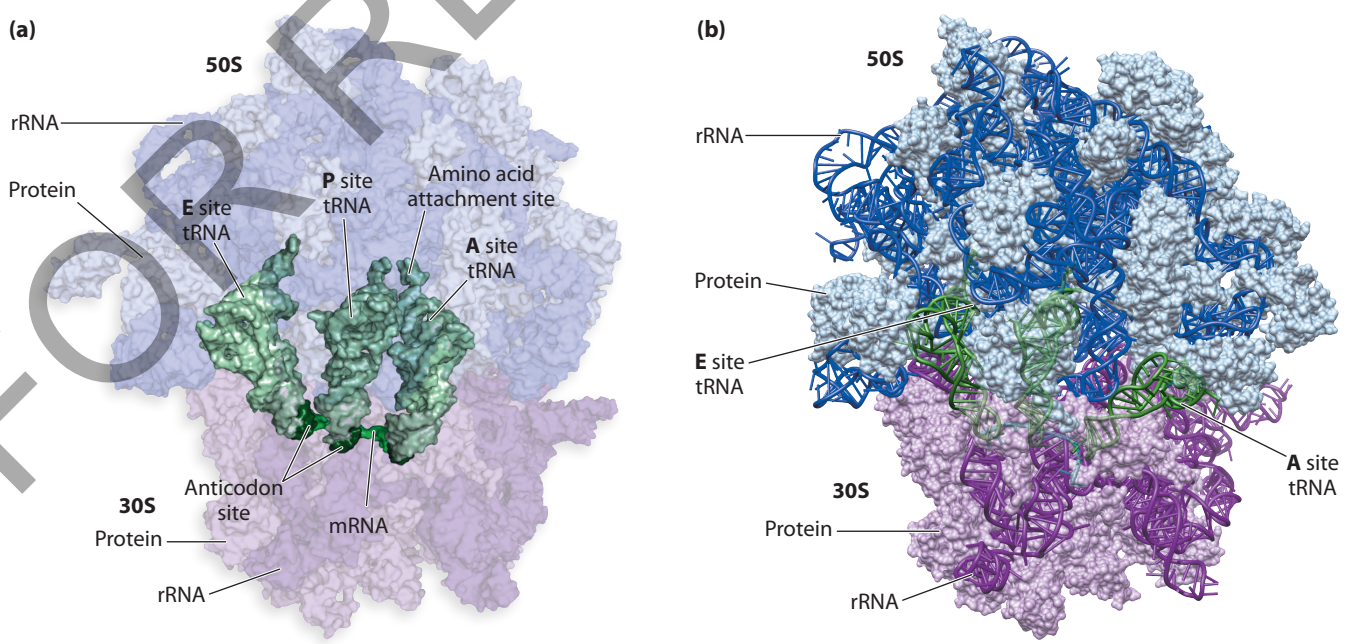



**Figure 9.5** **Three-dimensional computer interpretations of cryo-EM–generated data depict ribosome structure.**

Errant translation starting at the wrong codon, or even at the wrong nucleotide of the start codon, may produce an abnormal polypeptide and result in a nonfunctional protein. Thus, critical questions for biologists studying translation initiation were these: How does the ribosome locate the authentic start codon? And if more than one AUG (start codon) sequence occurs near the 5′ end of the mRNA, how is the authentic start codon identified? Bacteria and eukaryotes use different mechanisms to identify the authentic start codon.

**Bacterial Translation Initiation** In *E. coli*, six critical molecular components come together to initiate the translation process: (1) mRNA, (2) the small ribosomal subunit, (3) the large ribosomal subunit, (4) the initiator tRNA, (5) three essential initiation factor proteins, and (6) GTP.

For most of translation initiation in bacteria, the 30S ribosomal subunit is affiliated with an **initiation factor (IF)** protein called IF3, which facilitates binding between the mRNA and the 30S subunit. IF3 also prevents the 30S subunit from binding to the 50S subunit (Figure 9.6).

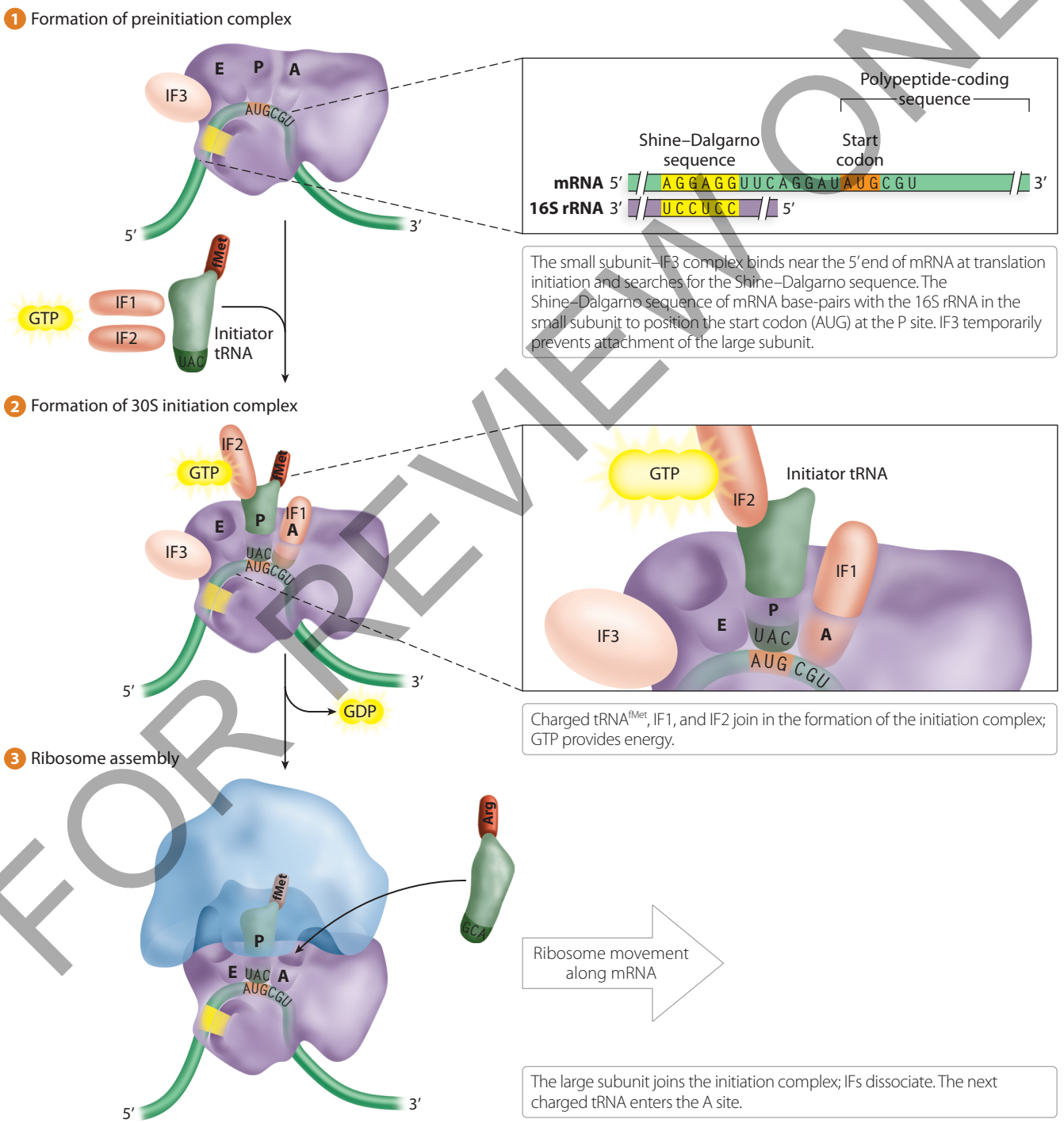



**Figure 9.6** **Initiation of bacterial translation.**

The small subunit–IF3 complex binds near the 5′ end of mRNA, searching for the AUG sequence that serves as the start codon. The **preinitiation complex** forms when the authentic start codon sequence is identified by base pairing that occurs between the 16S rRNA in the 30S ribosome and a short mRNA sequence located a few nucleotides upstream of the start codon in the 5′ UTR of mRNA (Figure 9.6, 1). John Shine and Lynn Dalgarno identified the location and sequence of this region in 1974, and it is named the **Shine–Dalgarno sequence** in recognition of their work.

The Shine–Dalgarno sequence is a purine-rich sequence of about six nucleotides located three to nine nucleotides upstream of the start codon. A complementary pyrimidine-rich segment containing the sequence UCCUCC is found near the 3′ end of 16S rRNA, and it pairs with the Shine–Dalgarno sequence to position the mRNA on the 30S subunit (see Figure 9.6). The Shine–Dalgarno sequence is another example of a consensus sequence. Like the consensus sequences we describe for promoters (Chapter 8) the Shine–Dalgarno sequence has a characteristic nucleotide composition and a precise position relative to the start codon, but its exact nucleotide sequence varies slightly from one mRNA to another (Figure 9.7).

In the next step of translation initiation (Figure 9.6, 2), the initiator tRNA binds to the start codon at what will be part of the P site after ribosome assembly. The amino acid on the initiator tRNA is a modified methionine called **N-formylmethionine (fMet);** thus, the charged initiator tRNA is abbreviated **tRNA$^{fMet}$.** This tRNA has a 3′-UAC-5′ anticodon sequence that is a complementary mate to the start codon sequence. An initiation factor (IF) protein designated IF2 and a molecule of GTP are bound at the P site to facilitate binding of tRNA$^{fMet}$. Initiation factor 1 (IF1) also joins the complex to forestall attachment of the 50S subunit. At this point, the **30S initiation complex,** consisting of mRNA bound to the 30S subunit, tRNA$^{fMet}$ located at the start codon, three initiation factors, and a molecule of GTP, has been formed.


**Figure 9.7 The Shine–Dalgarno consensus binding sequence.** The AUG start codon sequence (orange) is near the Shine–Dalgarno region (gold), which binds to the 3′ end of 16S rRNA.

In the final step of initiation (Figure 9.6, 3), the 50S subunit joins the 30S subunit to form the intact ribosome. The energy for the union of the two subunits is derived from hydrolysis of GTP to GDP (guanosine diphosphate). The dissociation of IF1, IF2, and IF3 accompanies the joining of subunits that creates the **70S initiation complex.** This complex is a fully active ribosome with a P site, an A site, an E site, and a channel for exit of the polypeptide. The first tRNA (tRNA$^{fMet}$) is already paired with mRNA at the P site, and the open A site contains the second codon and is awaiting the next charged tRNA.

**Eukaryotic Translation Initiation** The eukaryotic 40S ribosomal subunit complexes with three **eukaryotic initiation factor (eIF)** proteins eIF1, eIF1A, and eIF3 to form the preinitiation complex (Figure 9.8, 1). In step, 2 the preinitiation complex joins with the initiator tRNA and eIF5.

The **initiation complex** is formed by binding of the mRNA. This initiates the process called **scanning** (Figure 9.8, 3), in which the small ribosomal subunit moves along the 5′ UTR in search of the start codon. About 90% of eukaryotic mRNAs use the first AUG encountered by the initiation complex as the start codon, but the remaining 10% use the second or, in some cases, the third AUG as the start codon. The initiation complex is able to accurately locate the authentic start codon because the codon is embedded in a consensus sequence that reads

5′-ACCAUGG-3′

(the start codon itself is shown in bold). This consensus sequence is called the **Kozak sequence** after Marilyn Kozak, who discovered it in 1978.

Locating the start codon leads to recruitment of the 60S subunit to the complex, using energy derived from GTP hydrolysis. This final step 4 in the formation of the 80S ribosome is accompanied by joining of the two subunits and dissociation of the eIF proteins. In the 80S ribosome, the initiator tRNA$^{Met}$ is located at the P site; the A site is vacant, awaiting arrival of the second tRNA (Genetic Analysis 9.1).

**Archaeal Translation Initiation and Its Implications for Evolution** Archaeal ribosome subunits are composed of rRNAs that are more similar in size to those of bacteria than of eukaryotes. However, the ribosomal RNAs that make up the central structure of the subunits are distinct in each domain. Indeed the archaeal domain was only discovered after Carl Woese sequenced and compared rRNAs from many organisms and found that their sequences clustered into the three domains of life depicted in Figure 1.3.

Despite the similarity in size of archaeal and bacterial ribosomes, the process of translation initiation

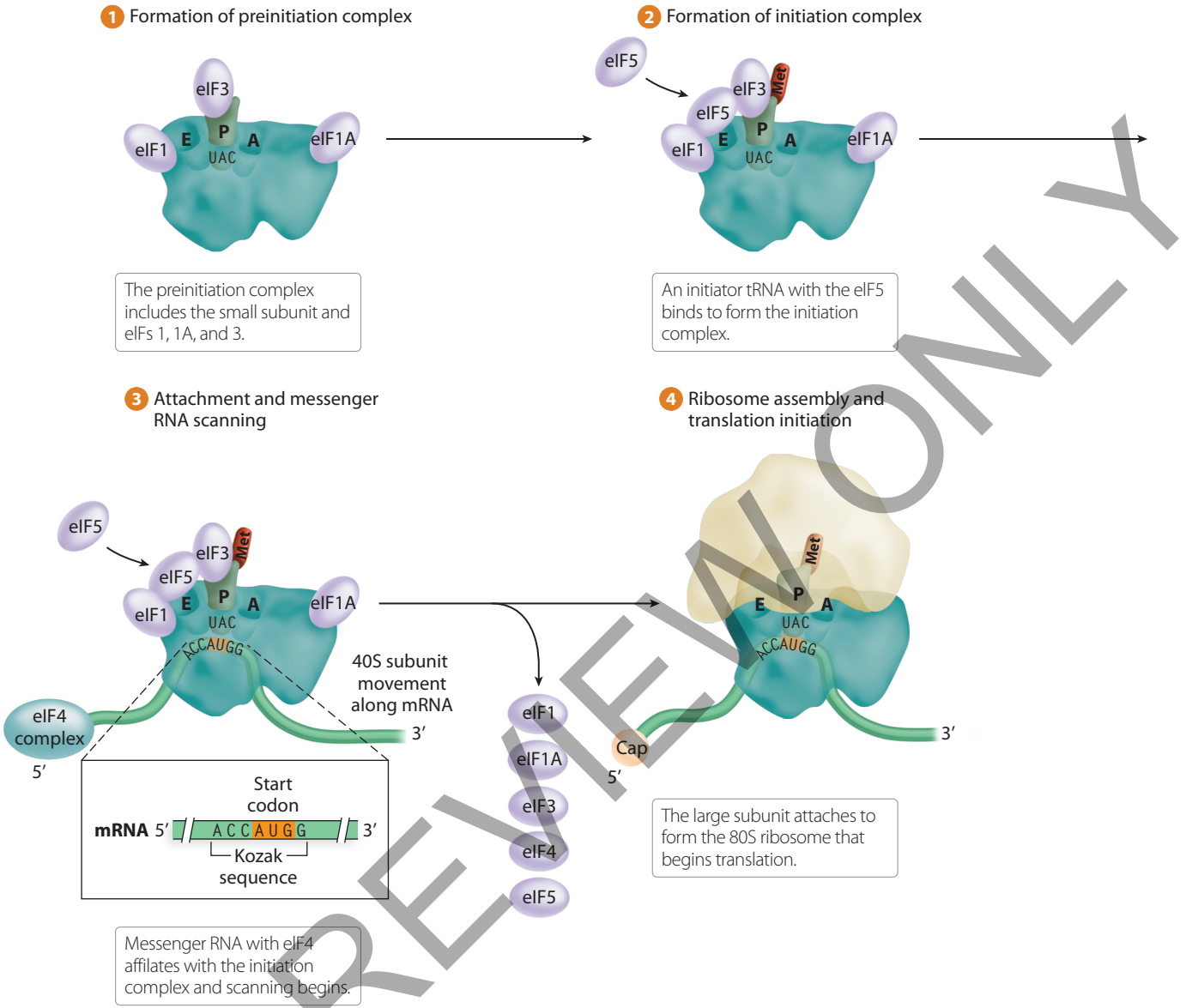



**Figure 9.8** **Initiation of eukaryotic translation.**

in archaea is decidedly eukaryote-like. One example of this similarity is the archaeal use of methionine as the common first amino acid of polypeptide chains. This is like eukaryotes and unlike bacteria, which use N-fromyl-methionine. A second aspect of archaeal translation initiation concerns the presence of Shine–Dalgarno sequences. These are relatively common in archaeal species that either do not produce leaderless mRNAs or produce very few. In contrast, archaeal species that produce a high proportion of leaderless mRNA, Shine–Dalgarno sequences are not as common, although they have been detected.

More significantly from an evolutionary perspective, **Table 9.3** lists archaeal translation initiation factor proteins and identifies their homologies to eukaryotic and bacterial proteins. Recall from our discussion in Section 1.4 that amino acid or nucleic acid sequences (proteins, DNA, or RNA) that are homologous have a common ancestral origin. As a consequence, proteins that have greater degrees of homology have more recent common ancestral history than do proteins with lower levels of homology. If proteins do not share a common ancestral history, they will not reveal homology.

Based on the homologous protein information in Table 9.3, it is clear that translation initiation in archaea is more complex than in bacteria and that known **archaeal initiation factor** proteins (**aIFs**) are homologous in structure and function to eIFs. This comparison of critical translational proteins also indicates striking similarity of translation initiation across the three domains of life. Translation in all forms of life has a common origin. Evolution has acted to conserve the key protein

**Table 9.3** Translation Initiation Factor Homologs

| Function | Bacterial Homolog[a] | Archaeal Homolog[b] | Eukaryotic Homolog[c] |
|---|---|---|---|
| mRNA binding; start codon fidelity | IF3 (in some phyla only) | aIF1 | eIF1 |
| mRNA binding | IF1 | aIF1a | eIF1A/eIF4 |
| tRNA P site binding | IF2 | aIF2/5 | eIF5 |
| $tRNA^{Met}$ binding | No homolog | aIF3 | eIF3 |

[a] The absence of a homologous protein is identified as "No homolog."
[b] Archaeal proteins are identified by the letter *a*.
[c] Eukaryotic proteins are identified by the letter *e*.

components of translation, with each domain acquiring its own specific features of translation.

The archaea have multiple mechanisms of mRNA–ribosome interaction at translation initiation. This is most apparent at the 5′ mRNA end where certain archaeal species have a large percentage—some studies say more than 50% of their mRNAs—that appear not to have a 5′ UTR. Those mRNAs lacking a 5′ UTR are said to be leaderless mRNAs and are apparently missing all or most of the translation initiating segments, including the Shine–Dalgarno sequence in some cases. The mechanism through which leaderless mRNA translation is initiated is not yet known. Archaeal species producing mRNAs with 5′ UTRs typically have Shine–Dalgarno sequences to aid translation intiation.

Analysis of experimental in vitro translation (translation in a test tube using ribosomes and translationally active proteins) testing the ability of bacterial and eukaryotic ribosomes and translational proteins to translate leaderless mRNAs from archaea finds that translation works efficiently in both in vitro systems. Leaderless mRNAs are very rare in bacteria or in eukaryotes, yet they are efficiently translated in vitro. This finding does not suggest a translational mechanism, but it has led to speculation that the leaderless mRNA state may be ancestral to the state featuring 5′ UTRs. In other words, it is possible that the last universal common ancestor (LUCA) of bacteria, archaea, and eukaryotes produced leaderless mRNAs and that the mRNAs with 5′ UTRs are a more recent development. In this context, archaeal translation may be something of a relic reminiscent of the situation in the LUCA.

## Polypeptide Elongation

Elongation, the second phase of translation, begins with the recruitment of **elongation factor (EF)** proteins into the initiation complex. Elongation factors facilitate three steps of polypeptide synthesis:

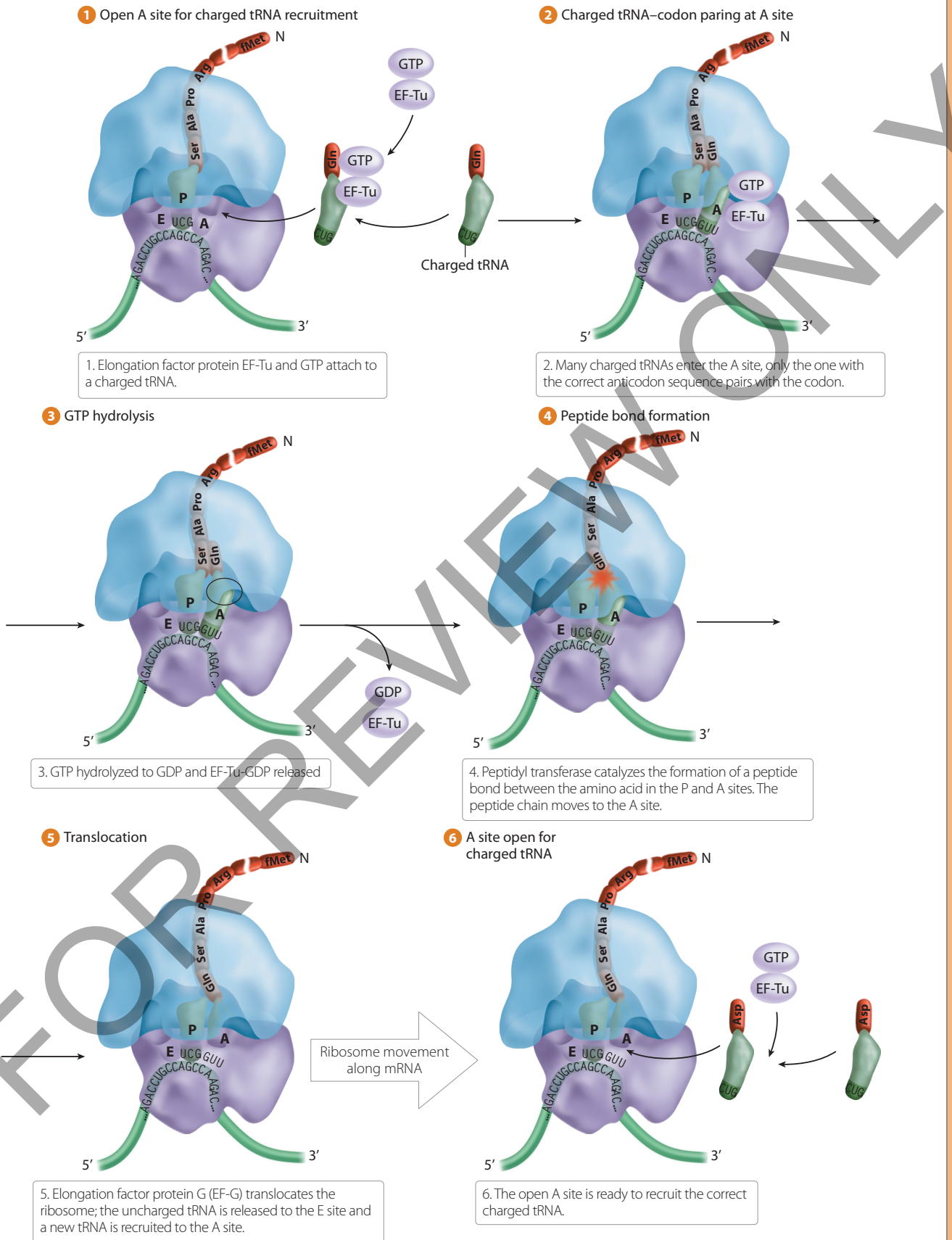
1. Recruitment of charged tRNAs to the A site
2. Formation of a peptide bond between sequential amino acids
3. Translocation of the ribosome in the 3′ direction along mRNA

GTP cleavage provides the energy for each step of elongation in bacteria, archaea, and eukaryotes (**Foundation Figure 9.9**). Moreover, the steps in the elongation process are the same in all three types of organisms: although the elongation factors differ, the ribosomal P, A, and E sites of all three organisms serve nearly identical functions. The rates of elongation are also similar; bacteria add about 20 new amino acids per second to a nascent polypeptide chain, and eukaryotes elongate the polypeptide at a rate of 15 amino acids per second. The elongation rate in archaea has not been established. Lastly, numerous studies indicate high fidelity of translation in all organisms. An error rate of approximately one amino acid in each 10,000 added to polypeptides is estimated for bacteria.

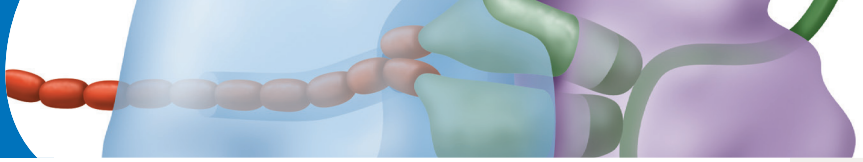
**Polypeptide Elongation in Bacteria** Different elongation factor proteins (EFs) and other ribosomal proteins carry out elongation in a series of steps depicted in Foundation Figure 9.9, while specifically describing translation in bacteria, is generally accurate for all organisms. The energy required for these steps is generated by hydrolysis, the cleavage of one phosphate molecules from guanosine triphosphate molecules (GTP). Hydrolysis releases energy and converts nucleotide triphosphates to nucleotide diphosphates (i.e., GTP → GDP). In step 1 a charged tRNAs is bound by the elongation factor EF-Tu and GTP. In step 2, the tRNA affiliates with the correct anticodon sequence enters the A site. In step 3 tRNA pairs with the mRNA codon and hydrolysis of GTP releases EF-Tu-GDP from tRNA. In step 4, the enzyme peptidyl transferase catalyzes peptide bond formation between the amino acid at the P site and the newly recruited amino acid at the A site. This elongates the polypeptide and transfers the polypeptide to the tRNA at the A site. The tRNA at the P site departs the ribosome through

## FOUNDATION FIGURE 9.9

# Bacterial Translation Elongation



# GENETIC ANALYSIS 9.1



**PROBLEM** In an investigation designed to identify the consensus sequence containing the AUG codon that initiates translation of eukaryotic mRNA, Marilyn Kozak (1986) compared the amounts of protein produced from 10 mutant mRNA molecules having different single-base substitutions flanking the AUG. Protein production was gauged by the optical density (OD) of protein bands in electrophoretic gels. Higher OD values indicated more protein produced. In the two tables shown, AUG, the start codon, is highlighted and its adenine (A) is labeled the +1 nucleotide of the translated region. Kozak examined six single-base mutants at nucleotide −3 and +4. These are identified by number (1 to 6) in Table A. She also examined four single-base mutants of positions −2 and −1. These are numbered 7 to 10 in Table B. The OD for protein production by each mutant was measured and is given below the mutant in the table. Use the OD values to determine answers to the problem questions.

**BREAK IT DOWN:** The Kozak consensus sequence, 5′-ACCAUGG-3′, includes the AUG start codon sequence and several surrounding mRNA nucleotides and is critical to ribosome recognition of the authentic start codon (p. 313).

**BREAK IT DOWN:** Efficient translation of mRNA produces more protein and is indicated by higher OD values for mutants possessing that capability (p. 313).

**Table A** Six Position −3 and +4 Mutants

| Mutant number | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| −3 | G | A | U | C | G | A |
| −2 | C | C | C | C | C | C |
| −1 | C | C | C | C | C | C |
| +1 | A | A | A | A | A | A |
| +2 | U | U | U | U | U | U |
| +3 | G | G | G | G | G | G |
| +4 | U | U | G | G | G | G |
| OD | 0.7 | 2.6 | 0.9 | 0.9 | 3.1 | 5.0 |

**Table B** Four Position −2 and −1 Mutants

| Mutant number | 7 | 8 | 9 | 10 |
|---|---|---|---|---|
| −3 | A | A | A | A |
| −2 | C | C | G | G |
| −1 | A | A | A | A |
| +1 | A | A | A | A |
| +2 | U | U | U | U |
| +3 | G | G | G | G |
| OD | 3.3 | 1.8 | 1.9 | 2.0 |

**a.** Looking just at the nucleotides in positions −3 and +4 for the six mutants in Table A, decide which nucleotides give the highest level of protein production.

**b.** Describe the impact of each nucleotide (A, T, C, and G) in the −3 position.

**c.** Looking just at nucleotides at position −2 and −1 for the four mutants in Table B, decide which nucleotides give the highest level of protein production.

**d.** Why did Kozak use only A in the −3 position to test the effects of nucleotides at positions −2 and −1?

**e.** Putting together data from both Table A and Table B, give the sequence of the mRNA region from −3 to +4 that produces the highest level of translation.

## Solution Strategies

### Evaluate

1. Identify the topic this problem addresses and the nature of the requested answer.
2. Identify the critical information given in the problem.

**TIP:** Notice that AUG is the start codon sequence in all mutants tested. As a consequence, differences in OD result from differences among the surrounding nucleotides.

### Deduce

3. Identify the constant and variable nucleotides displayed in Table A.
4. Identify the constant and variable nucleotides shown in Table B.

## Solution Steps

1. This problem involves examination and interpretation of the effects that sequence differences surrounding the mRNA start codon have on translation. The answer requires identifying the effects of base substitutions on translation and identifying the mRNA sequence corresponding to the highest translation level.
2. Two tables provide mRNA sequence for different sequence variants. For each variant, an OD value describes the approximate level of protein produced by translation of the sequence. Higher OD values correspond to more protein production.

3. In Table A, the nucleotide C is constant at positions −1 and −2, and position +3 is always G. Nucleotide variability is limited to positions −3 and +4.
4. In Table B, only the nucleotide at the −2 position varies; all other nucleotides are constant.

**Solve**

5. Specify the nucleotides in the −3 and +4 positions (Table A) that give the highest OD.

6. Assess how each nucleotide in the −3 position affects OD.

7. Evaluate how nucleotide differences at the −1 and −2 positions (Table B) affect OD.

8. Explain the decision to base Table B evaluations only on sequences with A in the −3 position.

TIP: Compare OD values and nucleotide differences from both tables to determine the most efficient consensus sequence.

9. Identify the start codon consensus sequence that results in the highest level of translation.

Answer a

5. In Table A, the presence of A in position −3 and G in position +4 produces the highest OD value. At the +4 position, G produces two high OD values and two low ODs, and T produces one high and one low OD.

Answer b

6. At position −3, A produces the highest and the third-highest OD values; G produces the second-highest and the lowest OD; T and C produce the same low OD value.

Answer c

7. In Table B, a C in position −2 and an A in position −1 produce the highest OD. Considering only the variable position −2, C produces higher OD values than does G.

Answer d

8. Adenine is selected as the nucleotide in position −3 for Table B evaluations based on the high average OD value for this nucleotide in comparison to other nucleotides. The average OD for A in the −3 position is $\frac{(5.0 + 2.6)}{2} = 3.8$ versus the next-highest average of $\frac{(3.1 + 0.7)}{2} = 1.9$ for G in the −3 position.

Answer e

9. Data from the two tables combined identify the sequence ACCAUGG (start codon in bold) as the most efficient consensus sequence for the start codon. For the nucleotide positions immediately surrounding the start codon, A is most efficient at −3, C is more efficient than G at −2, C is more efficient than A at −1, and G is more efficient than U at +4.

For more practice, see Problems 32, 33, and 34.

Visit the Study Area to access study tools. MasteringGenetics™

the E site. In step 5 elongation factor EF-G uses GTP hydrolysis to, EFs translocate the ribosome by moving it in the 3′ direction on mRNA. This translocation step is exactly one codon in length, that is, three nucleotides. Translocation moves the tRNA formerly at the A site to the P site, and opens the A site for binding by a charged tRNA with the correct anticodon sequence. In step 6 the next charged tRNA is ready to enter the A site.

### Elongation of Eukaryotic and Archaeal Polypeptides

Evolution has acted to strongly conserve the basic biochemistry of polypeptide elongation in all three domains of life. The elongation factors that carry out polypeptide elongation in eukaryotes and archaea are shown in Table 9.4. All organisms use two elongation factors to carry out polypeptide elongation, and the illustration of polypeptide elongation in Figure 9.9 is an equally accurate portrayal of the process in eukaryotes and archaea. Based on sequence comparisons, the archaeal and eukaryotic elongation factor homologs are more alike than are archaeal and bacterial EFs. This sequence analysis supports the initial assessment of Carl Woese that eukaryotes and archaea are more closely related to one another than either is to bacteria (see Section 1.1).

## Translation Termination

The elongation cycle continues until one of the three stop codons, UAG, UGA, or UAA, enters the A site of the ribosome. There are no tRNAs with anticodons complementary to stop codons, so the entry of a stop codon into the A site is a translation-terminating event. All organisms use **release factors (RF)** to bind a stop codon in the A site (Figure 9.10). The catalytic activity of RFs releases the polypeptide bound to tRNA at the P site. Polypeptide release causes ejection of the RF from the P site and leads to the separation of the ribosomal subunits.

In bacteria, two release factors, RF1 and RF2, recognize stop codons. RF1 recognizes UAG and UAA, and

**Table 9.4 Translation Elongation Factor Homologs**

| Function | Bacterial Homolog | Archaeal Homolog | Eukaryotic Homolog |
|---|---|---|---|
| Adjusts tRNA in A site | EFT | aEF1 | eEF1 |
| Promotes translocation | EFG | aEF2 | eEF2 |

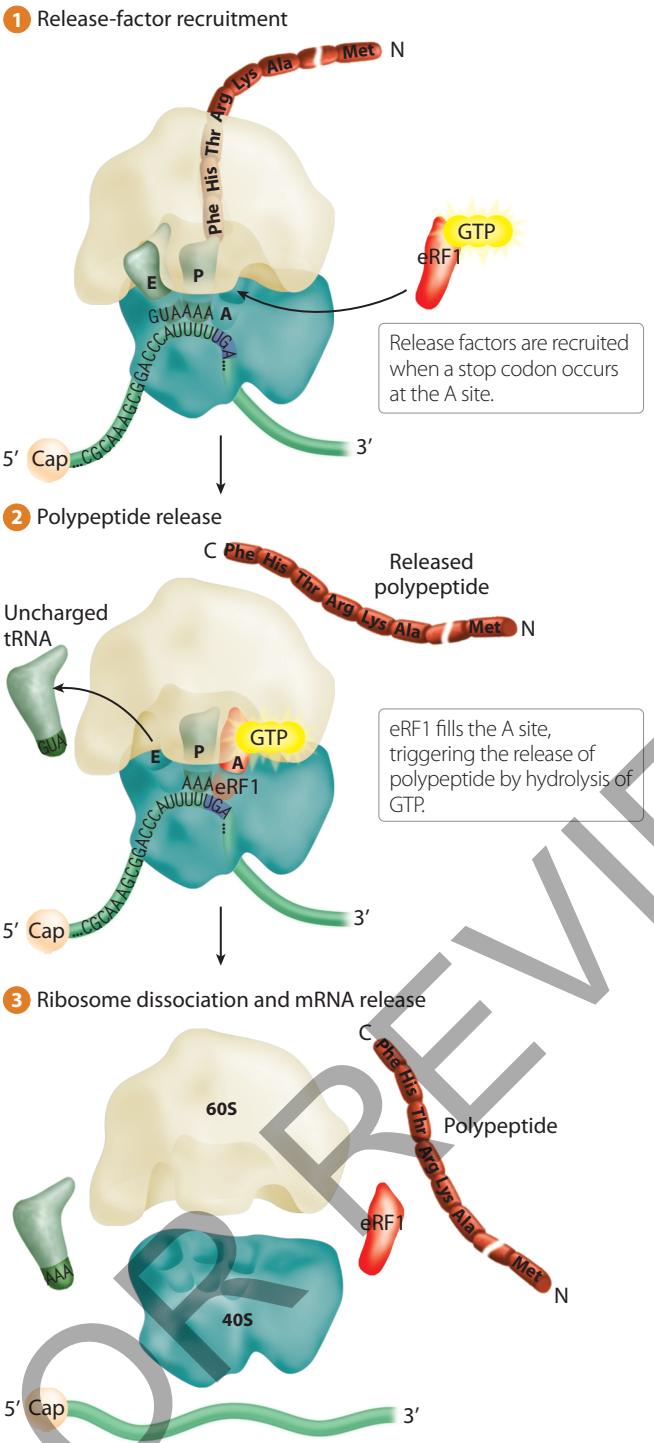



**Figure 9.10 Termination of translation by release factor (eRF) proteins.** A similar process terminates bacterial and archaeal translation.

RF2 recognizes UAA and UGA. A third bacterial release factor, RF3, is active in recycling RF1. Eukaryotic and archaeal translation are terminated by the action of a single release factor, identified as eRF1 in eukaryotes and aRF1 in archaea, that recognizes all three stop codons in organisms of both of these domains. Eukaryotes have a second RF that, like RF3 of bacteria, participates in recycling eRF1. The currently available information on sequence and function of RFs suggests that archaea and eukaryotes have RFs that are more like one another than either is to bacterial RFs (Table 9.5).

**Table 9.5 Translation Termination Factor Homologs**

| Function | Bacterial Homolog | Archaeal Homolog | Eukaryotic Homolog |
|---|---|---|---|
| Stop codon recognition | RF1 and RF2 | aRF1 | eRF1 |
| Recycling RF1 and eRF1 | RF3 | No homolog | eRF3 |
| Ribosome recycling | RRF | No homolog | No homolog |

## 9.3 Translation Is Fast and Efficient

With mRNA transcripts of hundreds to thousands of genes in cells, translation is an active and ongoing process that must efficiently initiate, elongate, and terminate polypeptide synthesis. In recent decades, research has uncovered several aspects of the translation machinery that help explain the speed, accuracy, and efficiency of polypeptide production.

### The Translational Complex

Cell biologists estimate that each bacterial cell contains about 20,000 ribosomes, collectively constituting nearly one-quarter of the mass of the cell. The number of ribosomes per eukaryotic cell is variable, but it too is in the tens of thousands. Given these numbers, it is not surprising that translation is almost never a matter of a solitary ribosome translating a single mRNA. Rather, electron micrographs reveal structures called **polyribosomes,** a busy translational complex containing multiple ribosomes that are each actively translating the same mRNA (Figure 9.11). Each ribosome in the polyribosome structure independently synthesizes a polypeptide, markedly increasing the efficiency of utilization of an mRNA.

In bacteria, the coupling of transcription and translation (Chapter 8) allows ribosomes to engage in translation of the 5′ region of mRNAs whose 3′ end is still under construction by RNA polymerase. This coupling is observed in Figure 9.11. Transcription occurs along DNA in the left-hand to right-hand direction. Translation of the mRNA transcripts begins before transcription is complete. In eukaryotes, however, transcription and translation are uncoupled. Transcription takes place in the nucleus, where pre-mRNA is processed to form mature mRNA. Translation occurs in the cytoplasm after release of mature mRNA.

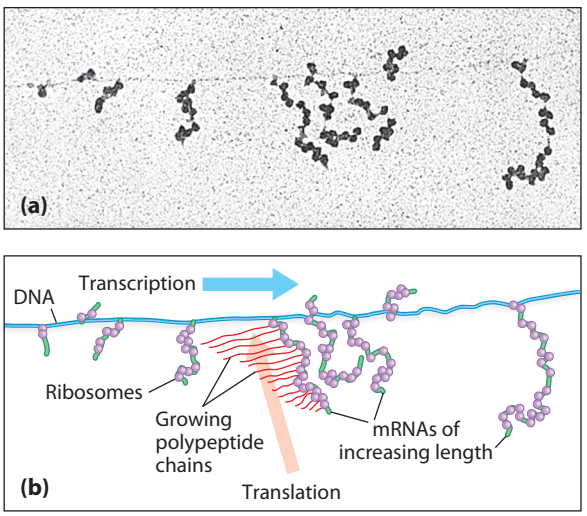



**Figure 9.11** **Polyribosomes.** **(a)** Electron micrograph of a polyribosome shows multiple ribosomes simultaneously translating a single mRNA molecule. Ribosomes that are closest to the stop codon have the longest polypeptides. **(b)** Artist rendition of the polyribosome electron micrograph. Transcription and translation are coupled in bacteria, and the translation direction is indicated.

## Translation of Polycistronic mRNA

Each polypeptide-producing gene in eukaryotes produces monocistronic mRNA, meaning mRNA that directs the synthesis of a single kind of polypeptide. The scanning model for translation described earlier for eukaryotes implies that a single start codon is identified in eukaryotic mRNA to initiate synthesis of one kind of polypeptide chain. In contrast, groups of bacterial and archaeal genes often share a single promoter, and the resulting mRNA transcript contains information that synthesizes several different polypeptides. These **polycistronic mRNAs** are produced as part of operon systems that regulate the transcription of sets of bacterial genes functioning in the same metabolic pathway (a form of regulation we discuss in Chapter 15).

Polycistronic mRNAs consist of multiple polypeptide-producing segments—multiple cistrons—that each contain sequence information for translation initiation. In the case of bacteria, and in all but the leaderless mRNAs in archaea, the translation-initiating region contains a Shine–Dalgarno sequence and start and stop codons. An intercistronic spacer sequence that is not translated separates the cistrons of polycistronic mRNA and contains the Shine–Dalgarno sequences (Figure 9.12).

Bacterial intercistronic spacers are variable in length: Some are just a few nucleotides long, although most are 30 to 40 nucleotides long. If the intercistronic spacer is a few nucleotides in length, it is, short enough to be spanned by a ribosome. In such systems, the ribosome remains intact after completing synthesis of one polypeptide, and it translates the other genes encoded in the polycistronic mRNA as well. On the other hand, for longer intercistronic spacers, the initial ribosome dissociates and new translation initiation must occur to translate the next polypeptide encoded by the polycistronic mRNA.

## 9.4 The Genetic Code Translates Messenger RNA into Polypeptide

Nucleic acids and amino acids are chemically very different compounds, and there is no *direct* mechanism by which mRNA could synthesize a polypeptide. Nevertheless, the genetic information carried in the nucleotide sequences of mRNA does provide a means by which the amino acid sequences of polypeptides can be specified. The "genetic code" is the name used to describe the correspondence between mRNA codon sequences and individual amino acids.

Converting the sequence of mRNA into a polypeptide depends on transfer RNA (tRNA) to carry amino acids to the ribosome. At ribosomes, tRNA pairs with mRNA by complementary base pairing between mRNA codon nucleotides and tRNA anticodon nucleotides. Once the correct tRNA is bound by a codon, it transfers its amino acid to the end of a growing polypeptide chain. Transfer RNA molecules facilitate the translation of genetic information from one chemical language (nucleic acid) to another (amino acid). That is, tRNA is an adaptor molecule that interprets and then acts on the information carried in mRNA.

Our review of translation and the genetic code in Chapter 1 depicts a triplet genetic code: Groups of three consecutive mRNA nucleotides form codons that each correspond to one amino acid. The genetic code contains 64 different codons, more than enough to encode the 20 common amino acids used to construct polypeptides. The greater number of codons than amino acids

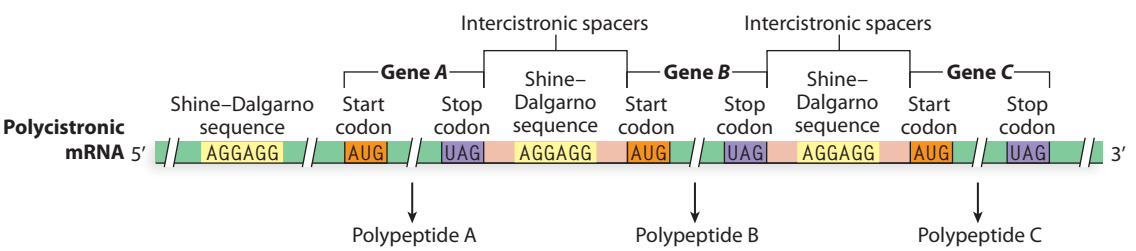



**Figure 9.12** **Polycistronic mRNA.** A polycistronic mRNA is a transcript of multiple genes and will produce a polypeptide from each gene.

leads to *redundancy* of the genetic code, as evidenced by the observation that single amino acids are specified by from one to as many as six different codons. This redundancy is explained by aspects of the base-pairing interactions between tRNA anticodons and mRNA codons.

## The Genetic Code Displays Third-Base Wobble

The triplet genetic code is a biological example of Ockham's razor, the principle that the simplest hypothesis is the most likely to be correct: During the late 1950s, arithmetic logic led many researchers to conclude that the genetic code was most likely triplet. This simple solution to the question of how amino acid sequences could be coded by nucleic acid sequences posits that a doublet genetic code (two nucleotides per codon) could produce just 16 ($4^2$) combinations of codons, which is not enough different combinations to specify 20 amino acids. On the other hand, a quadruplet genetic code would generate $4^4$, or 256, different combinations of codons—far too many for the needs of genomes. In contrast, a triplet genetic code, yielding $4^3$, or 64 different codons, provides enough variety to encode 20 amino acids with some, but not excessive, redundancy (**Figure 9.13** and genetic code information inside the front cover of the book). Among the 64 codons, 61 specify amino acids, and the remaining 3 are the stop codons that terminate translation. Only two amino acids, methionine (Met)—with the codon AUG—and tryptophan (Trp)—with the codon UGG—are encoded by single codons. The other 18 amino acids are specified by two to six codons. Codons that specify the same amino acid are called **synonymous codons.**

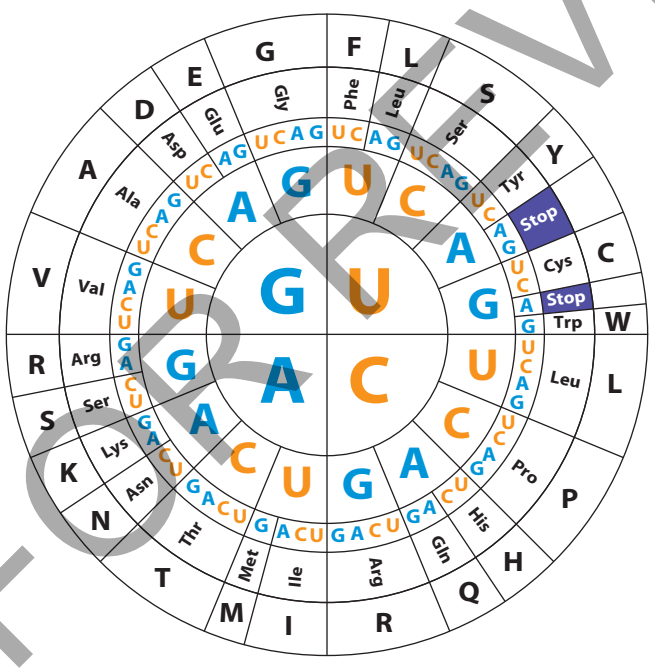



**Figure 9.13** **The genetic code.** To read this circular table of the genetic code, start with the inner ring, which contains the nucleotide in the first position (5′ nucleotide) of a codon. The second-position nucleotide is in the second ring, and the third-position nucleotide is in the third ring. Three-letter and one-letter abbreviations for the corresponding amino acids occupy the outermost rings.

Each transfer RNA molecule carries a particular amino acid to the ribosome, where complementary base pairing between each mRNA codon sequence and the corresponding anticodon sequence of a correct tRNA takes place. Note that this complementary base pairing requires antiparallel alignment of the mRNA and tRNA strands. Consider the codon sequence for aspartic acid (Asp), 5′-GAC-3′. Base-pairing rules predict that the tRNA anticodon sequence is 3′-CUG-5′ (**Figure 9.14**). Asp is also specified by a synonymous codon, 5′-GAU-3′, that pairs with tRNA carrying the anticodon sequence 5′-CUA-3′. Transfer RNA molecules with different anticodon sequences for the same amino acid are called **isoaccepting tRNAs.**

Does the presence of synonymous codons and isoaccepting tRNAs mean that a genome must provide 61 different tRNA genes and transcribe a tRNA molecule to match each codon? The answer is no. In fact, most genomes have 30 to 50 different tRNA genes. How does a genome that encodes fewer than 61 different tRNA molecules recognize all 61 functional codons? The answer lies in relaxation of the strict complementary base-pairing rules at the third base of the codon. The mechanics of translation provide for flexibility in the pairing of the third base, the 3′-most nucleotide, of the codon. **Third-base wobble** is the name given to the mechanism that relaxes the requirement for complementary base pairing between the third base of a codon and the corresponding nucleotide of its anticodon.

How does third-base wobble work? The answer is found in the chemical structures of nucleotides that hydrogen bond in base-pairing reactions. A careful look at synonymous codons reveals a pattern to the chemical structure of the third bases in cases of wobble. With the exception of the AUA codon for isoleucine (Ile) and the UGG codon for tryptophan (Trp), synonymous codons can be grouped into pairs that have the same two nucleotides in the first and second positions and differ only at the third base, where

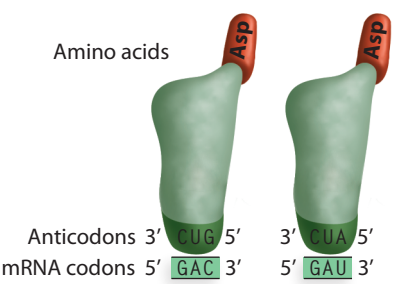



**Figure 9.14** **Codon–anticodon pairing.** A pair of isoaccepting aspartic acid tRNAs illustrates complementary antiparallel base-pairing of codon and anticodon sequences.

the synonymous codons either both carry a purine (A or G) or both carry a pyrimidine (C or U). For example, consider the synonymous pairs of codons for histidine (His) and glutamine (Gln; see Figure 9.13). The first two bases of each of these codons are C and A. Both His codons have a pyrimidine at the third position, whereas the Gln codons have a purine in the third position. As you look at other pairs of synonymous codons in the genetic code information inside the book front cover, notice that they also differ only by carrying the alternative purine or pyrimidine nucleotide at the third position.

Amino acids specified by four synonymous codons, such as alanine (Ala), valine (Val), and glycine (Gly), display an analogous pattern: Each amino acid is represented by two pairs of synonymous codons, and the members of each pair differ in the third position only, by carrying the alternate purine or pyrimidine. The pattern continues in arginine (Arg), serine (Ser), and leucine (Leu), each of which is specified by six synonymous codons. These sets of codons each consist of three pairs, each pair having the same nucleotides in the first two positions and differing by having the alternate purine or pyrimidine in the third position.

Third-base wobble occurs through flexible base pairing between the wobble nucleotide—that is, the 3′ nucleotide of a codon—and the 5′ nucleotide of an anticodon. At the wobble position, base pairing between the nucleotides of the codon and the anticodon need not be complementary. They must, however, involve a purine and a pyrimidine. Third-base wobble pairings are summarized in Table 9.6. The wobble nucleotides in different anticodons include all the RNA nucleotides and also the modified nucleotide **inosine (I).** Inosine is structurally similar to G but lacks the amino group attached to guanine's 2 carbon. Because of this difference, inosine base-pairs with either purines or pyrimidines. Figure 9.15 shows three examples of third-base wobble, in which three tRNA molecules collectively recognize seven different codons.

## Charging tRNA Molecules

Transfer RNA molecules are transcribed from tRNA genes. Recall the three-dimensional structure of tRNAs (see Figure 8.28) and the CCA terminus at the 3′ end of tRNA molecules as the site of attachment of an amino acid. Each tRNA carries only one of the 20 amino acids, and correct charging of each tRNA is crucial for the integrity of the genetic code.

**Table 9.6** Third-Base Wobble Pairing between Codon and Anticodon Nucleotides

| 3′ Nucleotide *of Codon* | 5′ Nucleotide *of Anticodon* |
|---|---|
| A or G | U |
| G | C |
| U | A |
| U or C | G |
| U, C, or A | I |

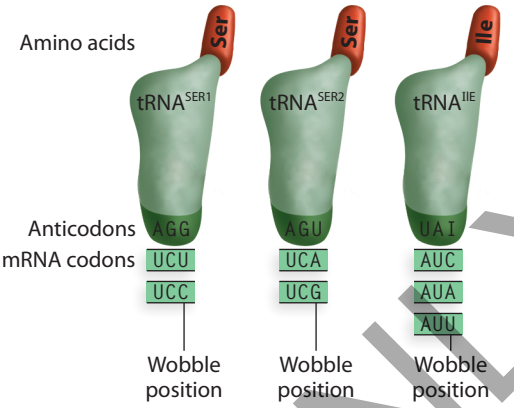



**Figure 9.15 Effect of wobble.** Wobble base pairing reduces the number of different tRNAs required during translation. In this example, two different tRNAs, each carrying serine, each use wobble to recognize a different pair of serine codons. A single isoleucine-carrying tRNA uses wobble to recognize three isoleucine codons.

The charging of tRNAs is catalyzed by enzymes called **aminoacyl-tRNA synthetases** or, more simply, **tRNA synthetases.** There are 20 different tRNA synthetases, one for each of the amino acids. To charge an uncharged tRNA, a tRNA synthetase catalyzes a two-step reaction that forms a bond between the carboxyl group of the amino acid and the 3′ hydroxyl group of adenine in the CCA terminus. Experimental analysis reveals that the recognition of isoaccepting tRNAs by tRNA synthetase is a complex process that does not follow a single set of rules. Mutations in any of the four arms of tRNA, or in the anticodon sequence itself, render a tRNA unrecognizable to its tRNA synthetase.

Studies of structural interactions between tRNA synthetases and their tRNAs show tRNA synthetase to be a large molecule that contacts several parts of a tRNA as part of the recognition process. These contact points can include the anticodon sequence and the other arms and loops of the tRNA (Figure 9.16). Once in contact with tRNA synthetase, the tRNA acceptor stem fits into an active site of tRNA synthetase. The active site contains the amino acid that will be added to the tRNA acceptor stem and ATP that provides energy for amino acid attachment.

Familiarize yourself with Figure 9.13 and the genetic code information inside the front cover by using them to decipher the mutations shown in Genetic Analysis 9.2.

## 9.5 Experiments Deciphered the Genetic Code

A remarkable set of experiments performed over less than 4 years in the early 1960s deciphered the genetic code and opened the way for biologists to understand

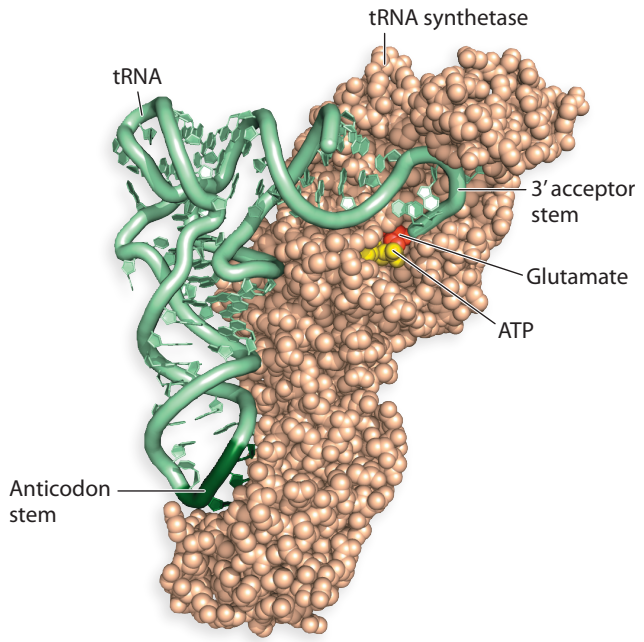



**Figure 9.16 Interaction of aminoacyl-tRNA synthetase with tRNA.** Aminoacyl-tRNA synthetase contacts multiple points on tRNA. ATP and the 3′ acceptor stem of tRNA fit in a cleft that also accommodates the amino acid.

the molecular processes that convert a messenger RNA nucleotide sequence into a polypeptide. At the time, biologists knew *what* the hereditary material was (DNA), and they knew *what* molecule conveyed the genetic message to ribosomes for translation (mRNA), but they did not know *how* the protein-coding information carried by messenger RNA was deciphered during the assembly of polypeptides. Several questions had to be answered about the structural nature of the genetic code before the code itself could be deciphered. The three most important questions, listed here, are examined in the sections below:

1. Do neighboring codons overlap one another, or is each codon a separate sequence?
2. How many nucleotides make up a messenger RNA codon?
3. Is the polypeptide-coding information of messenger RNA continuous, or is coding information interrupted by gaps?

## No Overlap in the Genetic Code

Consider the partial messenger RNA sequence

…ACUAAG…

If the genetic code is triplet and nonoverlapping (recall that a doublet code does not provide enough codons to specify 20 amino acids, and a quadruplet code provides far too many), this partial sequence produces two codons, each specifying an amino acid:

| codon | 1 | 2 |
|---|---|---|
| | …ACU | AAG… |
| amino acid | 1 | 2 |

In an overlapping triplet genetic code, on the other hand, these six nucleotides would spell out four complete codons and two partial codons. The sequence would fully encode four amino acids and contribute to the coding of two others:

| | …ACUAAG… |
|---|---|
| amino acid 1 | ACU |
| 2 | CUA |
| 3 | UAA |
| 4 | AAG |
| 5 | AG… |
| 6 | G… |

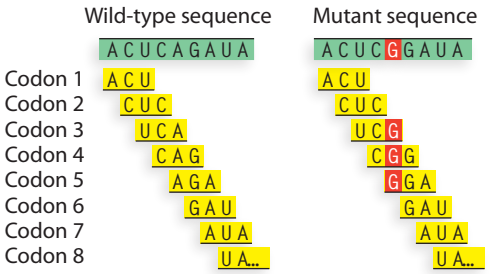
In 1957, based on his analysis of the available information on amino acid sequences of proteins, Sidney Brenner became convinced that an overlapping triplet genetic code was impossible because it was too restrictive. To test his hypothesis, Brenner examined the upstream neighbor of each AAG lysine in a large number of proteins and found 17 different amino acids in that position. He concluded that an overlapping genetic code restricted evolutionary flexibility and was unsupported by biochemical observations.

Conclusive evidence of a nonoverlapping genetic code came from a 1960 study of single-nucleotide substitutions induced by the mutation-producing compound nitrous oxide. Heinz Fraenkel-Conrat and his colleagues studied the effect of nitrous oxide on the coat protein of tobacco mosaic virus (TMV). Nitrous oxide causes mutations by inducing single base-pair substitutions in DNA that lead to mutant mRNA molecules with one nucleotide base change compared to wild-type mRNA. Asingle base change in mRNA would alter *three consecutive codons* if the genetic code were overlapping, but just a *single codon* if the genetic code were nonoverlapping (**Figure 9.17a**). Fraenkel-Conrat's mutation analysis revealed that only single amino acid changes occurred as a result of mutation by nitrous oxide. This result is consistent with that predicted for a nonoverlapping genetic code, and it is inconsistent with the prediction for an overlapping genetic code.

## A Triplet Genetic Code

Proof of a triplet genetic code came in 1961 when Francis Crick, Leslie Barnett, Sidney Brenner, and R. J. Watts-Tobin used the compound proflavin to create mutations in a gene called *rII* in T4 bacteriophage. Proflavin causes mutations by inserting or deleting single base pairs from DNA. This deletion leads to the absence of single nucleotides from

**(a)** An overlapping genetic code would change three consecutive codons with each base mutation.




**(b)** A nonoverlapping genetic code would change one codon with each base mutation.

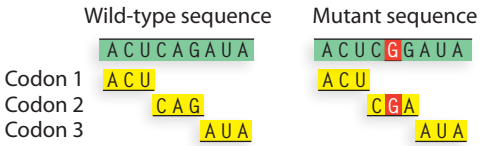



**Figure 9.17** **Proof that the genetic code is nonoverlapping.** The sequence of the last 10 amino acids at the C-terminal end of a TMV protein contained a single amino acid change following the induction of base-substitution mutation. This result conforms to the prediction of the nonoverlapping model of the genetic code.

mRNA, thus changing the reading frame of the mRNA. **Reading frame** refers to the specific codon sequence as determined by the point at which the grouping of nucleotides into triplets begins. The addition or deletion of nucleotides changes the reading frame and produces a mutation called a **frameshift mutation.**

The following analogy illustrates the impact of frameshift mutations. Single-letter additions or deletions garble the translated message by changing the reading frame:

wild-type: YOUMAYNOWSIPTHETEA ("you may now sip the tea")
mutant (addition): YOUMA C YNOWSIPTHETEA ("you ma c yno wsi pth ete a")
mutant (deletion): YOUMAYNO | | SIPTHETEA ("you may nos ipt het ea")

Frameshift mutations can be reverted (i.e., the correct reading frame can be restored) if a second mutation in a different location within the same gene restores the reading frame. This second mutation, called a **reversion mutation,** counteracts ("reverses") the reading frame disruption by inserting a nucleotide, if the initial mutation was a deletion, or by deleting a nucleotide, if the initial mutation was an insertion. For example, here is how the two frameshift mutations shown above might be reverted:

mutant (addition): YOUMA C YNOWSIPTHETEA (you mac yno wsi pth ete a)
reversion mutant (deletion): YOUMA c YNO | | SIPTHETEA ("you mac yno sip the tea")
mutant (deletion): YOUMAYNO | | SIPTHETEA ("you may nos ipt het ea")
reversion (addition): YOUMAYNOSIP R THE TEA ("you may nos ipr the tea")

Crick and his colleagues analyzed numerous bacteriophage proflavin-induced *rII*-gene mutants, designating each addition mutant as a (+) and each deletion mutation as a (−). They *guessed* that the first *rII*-gene mutant they examined, a mutation designated FC 0, resulted from insertion ("FC" stands for Francis Crick). Designating FC 0 as a (+) mutation turned out to be a correct guess. Based on their assumptions that (1) the genetic code is a nonoverlapping triplet and (2) FC 0 is an insertion (+) mutation, the data reported by Crick and colleagues supported the notion that the genetic code is based on nucleotide triplets.

Data on several mutants is displayed in **Table 9.7**. Each mutant is designated either (+) or (−). Any combination of a (+) mutant and a (−) mutant generates a wild-type revertant. In each case, the initial mutation causes a frameshift mutation, and the reversion mutation restores the reading frame. The triplet structure of the genetic code is demonstrated by the observation that the reading frame is restored by the presence of *three* (+) mutations or *three* (−) mutations. For example, the total of three insertions restores the reading frame in the following sentence after the position of the third insertion:

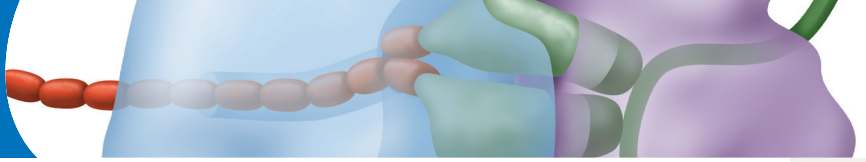
triple mutant (addition):
YOUMA C YNOW T S L IPTHETEA ("you ma c yno w t s l ip the tea")

## No Gaps in the Genetic Code

In their 1961 research, Crick and colleagues also suggested that the genetic code is read as a continuous string of mRNA nucleotides uninterrupted by any kind of gap,

**Table 9.7** **Phenotypes Resulting from Various Combinations of Proflavin-Induced Base-Pair Insertion (+) and Deletion (−) Mutations at the *rII* Locus of Bacteriophage T4**

| Combined Mutations | +/− Designations | Result |
|---|---|---|
| FC 0, FC 1 | +− | Wild-type revertant |
| FC 0, FC 21 | +− | Wild-type revertant |
| FC 40, FC 1 | +− | Wild-type revertant |
| FC 58, FC 1 | +− | Wild-type revertant |
| FC 0, FC 40, FC 58 | +++ | Wild-type revertant |
| FC 1, FC 21, FC 23 | −−− | Wild-type revertant |
| FC 0, FC 40 | ++ | *rII* mutant |
| FC 0, FC 58 | ++ | *rII* mutant |
| FC 1, FC 21 | −− | *rII* mutant |
| FC 1, FC 23 | −− | *rII* mutant |



**PROBLEM** A portion of an mRNA encoding C-terminal amino acids and the stop codon of a wild-type polypeptide is

5′-…CAACUGCCUGACCCACACUUAUCACUAAGUAGCCUAGCAGUCUGA…-3′

The wild-type amino acid sequence encoded by this portion of mRNA contains the amino acid Asn encoded by the codon 5′-CAA-3′. The remainder of the amino acids are encoded in the same reading frame.

**BREAK IT DOWN:** The mRNA sequence is complementary to the DNA template strand and differs from the DNA coding strand only by having uracil instead of thymine (p. 270).

N…Asn-Cys-Leu-Thr-His-Thr-Tyr-His-C

The C-terminal ends of three independently obtained mutant proteins produced by this gene are as follows.

Mutant 1: N…Asn-Cys-Leu-Thr-His-Thr-C
Mutant 2: N…Asn-Cys-Leu-Thr-His-Thr-Tyr-His-Lys-C
Mutant 3: N…Asn-Cys-Leu-Thr-His-Thr-Tyr-His-Tyr-Ser-Ser-Leu-Ala-Val-C

Identify the mutational events that produce each of the mutant proteins.

**BREAK IT DOWN:** Mutations occur at the level of DNA. Comparison of each mutant DNA and amino acid sequences with the wild-type sequence will reveal how the DNA sequence is changed (p. 321).

| Solution Strategies | Solution Steps |
|---|---|
| **Evaluate** | |
| 1. Identify the topic this problem addresses and the nature of the requested answer. | 1. This problem concerns evaluation of the C-terminal end of a wild-type protein sequence and the mRNA segment that encodes it and comparison of the wild-type protein to three mutant proteins to determine the alteration producing each mutant. The answers require the identification of specific mRNA sequence changes leading to each mutant protein. |
| 2. Identify the critical information given in the problem. | 2. In this problem the C-terminal end of a wild-type protein and the mRNA sequence that encodes it are given. Also given are the C-terminal sequences of three mutant proteins encoded by mutant mRNA sequences derived by alteration of the wild-type sequence. |
| **Deduce** | |
| 3. Use the genetic code to identify the codons corresponding to wild-type amino acids and to identify the stop codon. | 3. Two codons, AAC and AAU, encode asparagine (Asn). If we skip the 5′-most nucleotide of the mRNA sequence and begin reading at the A in the second position, the first codon is AAC followed by UGC-CUG-ACC-CAC-ACU-UAU-CAC-UAA. These codons encode the wild-type amino acids, and UAA is the stop codon. |
| 4. Compare each mutant polypeptide to the wild type and determine which codon contains the mutation. | 4. Mutant 1—The polypeptide sequence is truncated two amino acids short of the normal stop codon. The Tyr codon (UAU) appears to have changed to a stop codon.<br>Mutant 2—The wild-type sequence is extended by the addition of lysine (Lys), indicating that mutation changed the stop codon to a codon specifying Lys and is now followed immediately by a new stop codon.<br>Mutant 3—The wild-type sequence is extended by six amino acids. This suggests another mutation affected the stop codon. |
| **Solve** | |
| 5. Identify the mutation and its consequence for translation in Mutant 1. | 5. Two different base substitutions altering the tyrosine (Tyr) codon UAU to a stop codon could cause Mutant 1. The wild-type UAU codon was most likely altered by base substitution to form either a UAA or a UAG stop codon. |
| 6. Identify the mutation and its consequence in Mutant 2. | 6. Lysine (Lys), which was added to the mutant polypeptide, is encoded by AAA or AAG. Deletion of the U from the wild-type stop codon would produce an AAG codon followed by UAG, a stop codon. |
| 7. Identify the mutation and its consequence in Mutant 3. | 7. Tyrosine, specified by codons UAU and UAC, is found in place of the normal stop codon. This is followed by a serine codon (UCN or AGU/C), rather than the GUA (Val) that follows the "in-frame" stop codon in the wild type. A base-pair insertion that adds a U or a C into the third position of the normal UAA stop codon forms a UAU or a UAC tyrosine (Tyr) codon. The altered reading frame from that point would then read AGU (Ser), followed by AGC (Ser), CUA (Leu), GCA (Ala), GUC (Val), and UGA (stop). |

**TIP:** Any of three stop codons (UAG, UGA, or UAA) terminates translation immediately after the codon specifying the amino acid at the C terminus of a polypeptide.

**TIP:** Examine the wild-type nucleotide sequence at the place where mutation is expected to have occurred, and identify ways in which base substitution, insertion, or deletion could have had the observed effect on the amino acid sequence.

For more practice, see Problems 5, 11, 16, and 29.

space, or pause. If a gap or spacer were present between mRNA codons, the mRNA transcript might be represented as follows (*x* indicates the gap between codons):

YOUxMAYxNOWxSIPxTHExTEAx ("you may now sip the tea")

If the genetic code were structured in some such way, with each codon set off from its neighbors, insertion or deletion of a nucleotide would not cause the kind of frameshift mutation that Crick and colleagues had observed. Instead, insertion or deletion of nucleotides could be expected to alter the affected codon but not the identity of adjoining codons. For example, consider the following insertion mutation, where the separation between codons confines the alteration to a single word:

YOUx,MA T Yx,NOWx,SIPx,THEx,TEAx, ("you ma t y now sip the tea")

## Deciphering the Genetic Code

The genetic code was deciphered in a series of experiments performed between 1961 and 1965. This remarkable 4-year period in biology was highlighted by extensive collaborative and competitive international research that culminated in the assembly of a simple table containing the instructions shared by all organisms for translating mRNA nucleotide sequences into polypeptide sequences. Deciphering the genetic code was a milestone in establishing the mechanism of the central dogma of biology (DNA → RNA → protein) and laying the molecular foundation for modern genetic research. This triumph of deductive reasoning was instantly recognized for its profound significance, and it resulted in the awarding of a Nobel Prize in Physiology or Medicine to Har Gobind Khorana and Marshall Nirenberg in 1968.

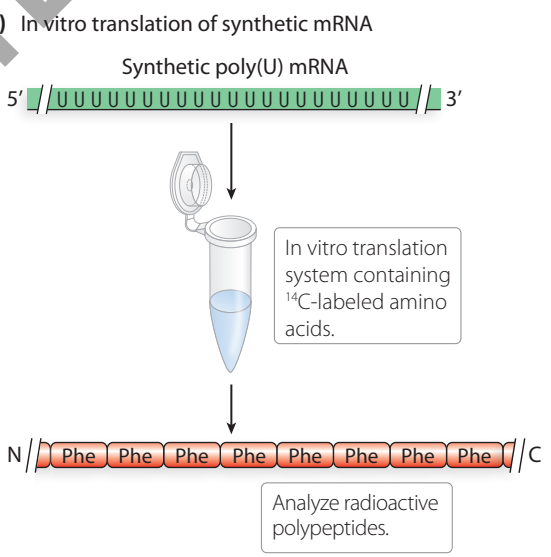
Once it had been established that the genetic code consists of triplets, researchers sprang to the task of establishing which triplets are associated with each amino acid in the process of translation. Nirenberg and Johann Heinrich Matthaei performed a simple experiment in 1961 that laid the groundwork for later experiments in deciphering the genetic code. Their experimental design was straightforward: Construct synthetic strings of repeating nucleotides, and use an in vitro translation system to translate the sequence into a polypeptide. For example, Nirenberg and Matthaei synthesized an artificial mRNA containing only uracils, known as a poly(U). They devised an in vitro translation system composed of the known cellular components of bacterial translation—ribosomes, charged transfer RNA molecules, and essential translational proteins. Regardless of where translation might begin along the poly(U) mRNA, the only possible codon it contained was UUU. The researchers were therefore hoping to determine which amino acid corresponds to the UUU codon.

Twenty separate in vitro translations of poly(U) mRNA were carried out, each time using a pool of 19 unlabeled amino acids and one amino acid labeled with radioactive carbon ($C^{14}$). To determine which amino acid is encoded by poly(U) mRNA, Nirenberg and Matthaei used a different radioactive amino acid in each translation. They detected production of a highly radioactive polypeptide after conducting translation in a system containing radioactively labeled phenylalanine (**Figure 9.18**). The radioactive polypeptide was poly-phenylalanine (poly-Phe). Since the only possible triplet codon in the mRNA was UUU, Nirenberg and Matthaei reasoned that 5′-UUU-3′ codes for phenylalanine. They went on to construct poly(A), poly(C), and poly(G) synthetic mRNAs and identified 5′-AAA-3′ as a codon for lysine (Lys), 5′-CCC-3′ as a proline (Pro) codon, and 5′-GGG-3′ as a codon for glycine (Gly) (**Table 9.8**).

Khorana adapted the experimental strategy of Nirenberg and Matthaei to synthesize mRNA molecules that contained di-, tri-, and tetranucleotide repeats. His construction of repeat-sequence mRNAs allowed him to define many additional codons (see Table 9.8). For example, Khorana used the dinucleotide repeat UC to form a synthetic mRNA with the sequence

5′-UCUCUCUCUCUCUCUC-3′

This mRNA can be translated in either a reading frame that begins with uracil or a reading frame that begins with cytosine. In both cases, the reading frame produces




**(b)** Incorporation of $^{14}$C-labeled phenylalanine into polypeptides

| Synthetic mRNA | Radioactivity (counts/min) |
|---|---|
| None | 44 |
| Poly(U) | 39,800 |
| Poly(A) | 50 |
| Poly(C) | 38 |

**Figure 9.18 Use of synthetic mRNAs to determine genetic code possibilities.** **(a)** Synthetic poly(U) mRNA is translated in vitro in the presence of individual $^{14}$C-labeled amino acids. A polypeptide consisting of phenylalanine is formed. **(b)** These radioactivity counts demonstrate that only poly(U) synthetic mRNA incorporates radioactive phenylalanine into a polypeptide.

**Table 9.8** Example Polypeptide Production from Synthetic mRNAs

| Synthetic mRNA | mRNA Sequence | Polypeptides Synthesized | Observation |
|---|---|---|---|
| Repeating nucleotides | Poly-U UUUU… | Phe- Phe- Phe… | Polypeptides have one amino acid. |
| | Poly-C CCCC… | Pro- Pro- Pro | |
| Repeating dinucleotides | Poly-UC UCUC… | Ser-Leu-Ser-Leu | Polypeptides have two alternating amino acids. |
| | Poly-AG AGAG… | Arg-Glu-Arg-Glu | |
| Repeating trinucleotides | Poly-UUC UUCUUCUUC… | Phe-Phe…and Ser-Ser…and Leu-Leu… | Three polypeptides have one amino acid each. |
| | Poly-AAG AAGAAGAAG… | Lys-Lys…and Arg-Arg… and Glu-Glu | |
| Repeating tetranucleotides | Poly-UAUC UAUCUAUC… | Tyr-Leu-Ser-Ile-Tyr-Leu-Ser-Ile | Some polypeptides have four repeating amino acids. Others identify stop codons. |
| | Poly-GUAA GUAAGUAA… | None (UAA stop codon) | |

*Note:* Data adapted from Khorana (1967).

alternating UCU-CUC codons. Khorana identified the amino acids of the resulting polypeptide and found it contained alternating serine (Ser) and leucine (Leu).

When Khorana used mRNA containing trinucleotide repeats, most of these mRNAs produced three different polypeptides that each consisted of only one kind of amino acid. For example, the reading frame for poly-UUC can begin with either of the uracils or with cytosine. Messenger RNA is read as consecutive UUC codons if the first uracil initiates the reading frame, as UCU if the second uracil begins the reading frame, or as CUU if cytosine is at the start of the reading frame. Although the different reading frames each produced a polypeptide containing one amino acid, Khorana was again unsure which codon specified which amino acid.

Nirenberg and Philip Leder contributed the final piece of the genetic code puzzle in 1964 when they devised an experiment to resolve the ambiguities of codon identity remaining from Khorana's experiments. They synthesized many different mini-mRNAs that were each just three nucleotides in length (**Figure 9.19**). The tiny mRNAs were added individually to in vitro translation systems containing ribosomes, along with 19 unlabeled amino acids and 1 $^{14}$C-labeled amino acid, all attached to different transfer RNA molecules. The mRNA formed a complex with the ribosome and the tRNA charged with the corresponding amino acid. Each in vitro mixture was then poured through a filter that captured the large ribosome–mRNA–tRNA complexes but permitted noncomplexed molecules of mRNA or tRNA to pass through. The filter was subsequently tested to determine if the three-nucleotide mRNA sequence bound a transfer RNA with the radioactive amino acid. Nirenberg and Leder tested all 64 combinations of nucleotides with their tiny mRNA system and were able to identify codon–amino acid correspondences for the entire genetic code. In addition, they identified the nucleotide composition of the three stop codons, UAA, UAG, and (Use **Genetic Analysis 9.3** to test your skill at interpreting the genetic code).

## The (Almost) Universal Genetic Code

In astonishing testimony to a single origin of life on Earth and to the power of evolution to maintain virtually complete uniformity over hundreds of millions of years, every living organism uses the same genetic code to synthesize polypeptides. In all living things, from bacteria to humans, the hereditary script carried by any given mRNA is translated by a similar mechanism and produces the same polypeptide. The universality of the genetic code makes it possible to use bacterial systems to express biologically important protein products found in plants or animals. The production of human insulin to treat diabetes and of factor VIII protein to treat hemophilia are two of numerous examples of recombinant human gene cloning that are possible in part because bacteria and humans use the same genetic code for translation.

As with most general rules, however, there are a few exceptions to the universality of the genetic code; thus, biologists characterize the genetic code as *almost* universal. The exceptions are found principally in mitochondria, which are specially adapted to life within plant and animal cells, but two exceptions occur in free-living organisms as well (**Table 9.9**). The near universality of the genetic code presents two important evolutionary questions. First, why has the genetic code remained essentially unchanged in living organisms; and second, why have changes evolved mostly in mitochondria? The answer to the first question is that natural selection pressure against codon change is intense. A single codon change would dramatically alter the composition of almost every polypeptide an organism produces.

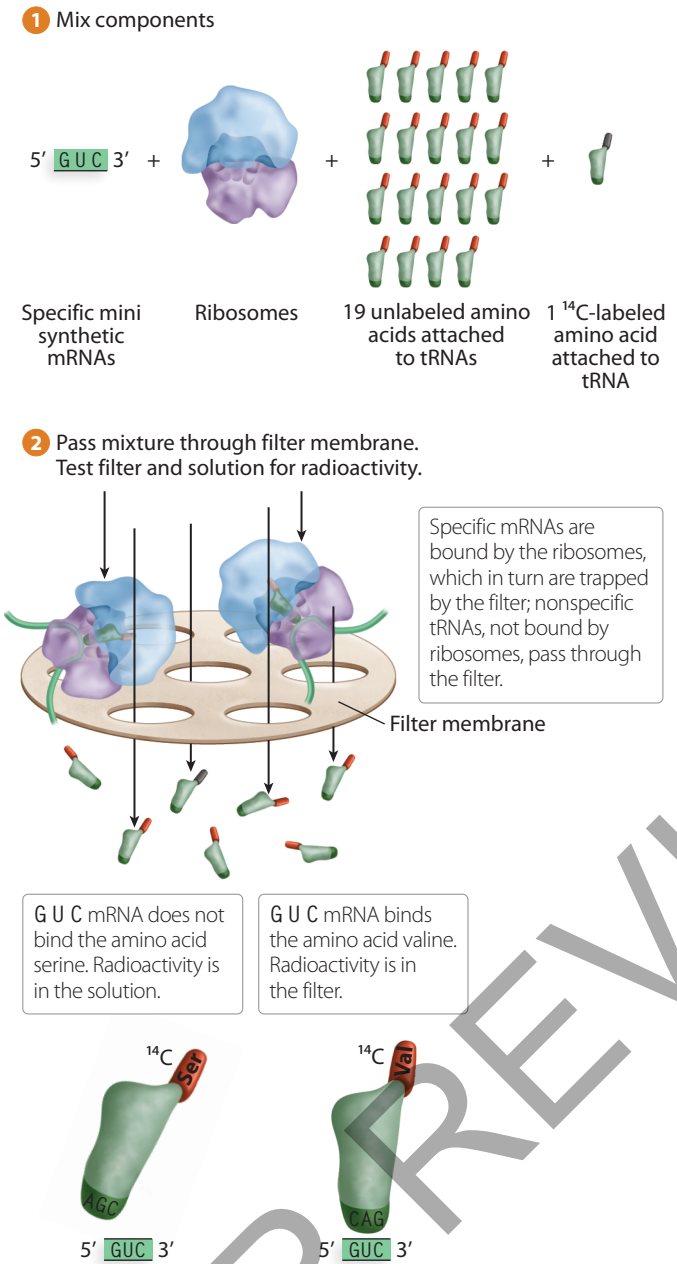



**Figure 9.19** **Deciphering the genetic code with synthetic mini mRNAs.** For the synthetic mini mRNA GUC, a $^{14}$C-labeled serine tRNA does not hybridize within the ribosome to form a complex, and radioactivity is located in the pass-through solution. $^{14}$C-labeled valine tRNA does hybridize to the GUC mini mRNA within the ribosome. The mRNA–ribosome–tRNA complex is caught by the filter membrane, where radioactivity is detected.

Countless evolutionary examples tell us that nearly all of the changes that occur would be deleterious, and many would be lethal. Simply stated, a change in the genetic code would alter the rules of the game of life, and natural selection prevents such changes.

The answer to the second question is that natural selection appears to be less intensive on the mitochondrial genetic code than on the genetic code for nuclear genes. The genomes of mitochondria found in plant and animal cells are small compared to nuclear genomes, and any disruption caused by a change in the mitochondrial genetic code is likely to be limited, since the number of genes affected is so small. In addition, there are many mitochondria per cell, providing "backup copies" of the mitochondrial genome. If a change in the genetic code severely disrupts the function of one mitochondrion, others are present in the cell to carry out normal activities.

**Table 9.9** Genomes Using Modifications of the Universal Genetic Code

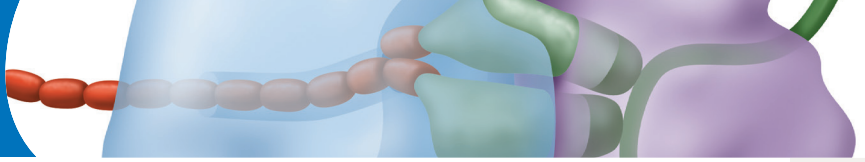
| Codon | Universal Code | Unusual Code | Genome |
|---|---|---|---|
| AGA, AGG | Arg | Stop | Mitochondria in plants, animals, and yeast |
| AUA, AUU | Ile | Met | Mitochondria in plants, animals, and yeast |
| UGA | Stop | Trp | Mitochondria in plants, animals, and yeast, and in *Mycoplasma* species |
| CUN[a] | Leu | Thr | Mitochondria in yeast |
| UAA, UAG | Stop | Gln | Green algae, protozoa |
| UGA | Stop | Cys | Protozoa |

N[a] = any third-position nucleotide.

## Transfer RNAs and Genetic Code Specificity

In our discussion of the genetic code and polypeptide assembly at the ribosome, we describe the specific base-pair interaction between the anticodon sequence of charged tRNA and the codon sequence of mRNA as the key to incorporating the correct amino acid into the polypeptide. But how did biologists determine that the specificity of the genetic code resides in the tRNA–mRNA interaction and not in the recognition of the amino acid carried by tRNA?

The answer came from a simple and clever experiment by Francois Chapeville and several colleagues in 1962. The researchers began by preparing normal cysteine-charged tRNAs. This complex is designated Cys-tRNA$^{Cys}$. The researchers then treated Cys-tRNA$^{Cys}$ with the compound Raney nickel that removes an SH group from cysteine and converts it to alanine. This treatment produces Ala-tRNA$^{Cys}$ in which alanine rather than cysteine is attached to tRNA$^{Cys}$. When Chapeville and colleagues used Ala-tRNA$^{Cys}$ in an in vitro translation reaction, the polypeptide contained alanine rather than cysteine in amino acid positions that would normally carry cysteine. In other words, Ala-tRNA$^{Cys}$ efficiently paired with mRNA codons specifying cysteine and deposited alanine in the nascent polypeptide, even though the mRNA sequence specified cysteine.



The following segment of DNA encodes a polypeptide containing six amino acids. DNA triplets encoding the start codon (AUG) and a stop codon are included in the sequence.

**BREAK IT DOWN:** The DNA coding strand differs from mRNA by the presence of T in DNA in place of the U in RNA (p. 270).

```
5′-...CCCAGCCTAGCCTTTGCAAGAGGCCATATCGAC...-3′
3′-...GGGTCGGATCGGAAACGTTCTCCGGTATAGCTG...-5′
```

**a.** Identify the sequence and polarity of the mRNA encoded by this gene.

**b.** Determine the amino acid sequence of the polypeptide, and identify the N- and C-terminal ends of the polypeptide.

**BREAK IT DOWN:** The genetic code (see inside the front cover or Figure 9.13) is used for translation (p. 321).

**c.** Base-substitution mutation changes the first transcribed G of the template strand to an A. How does this alter the polypeptide?

**BREAK IT DOWN:** A base substitution on the template DNA strand also requires that the nucleotide on the coding strand be changed to the complementary nucleotide (p. 321).

| Solution Strategies | Solution steps |
|---|---|
| **Evaluate** | |
| 1. Identify the topic this problem addresses and the nature of the requested answer. | 1. This problem concerns the identification of DNA coding and template strands, the protein encoded by DNA, and an evaluation of a mutation of the DNA sequence. The answer requires identification of the DNA strands, identification of start and stop codons, and determination of the amino acid sequence of wild-type and mutant proteins. |
| 2. Identify the critical information given in the problem. | 2. DNA sequence that includes a start (AUG) codon and a stop codon is given. |
| **Deduce** | |
| 3. Identify the start codon by inspecting both DNA strands for 3′-TAC-5′ that potentially encodes a start (AUG) codon on the template strand. | 3. Scanning both DNA strands in their 3′ to 5′ direction identifies a single 3′-TAC-5′ sequence. The sequence is on the upper strand of the sequence beginning with the seventh nucleotide from the right. |
| **TIP:** The AUG start codon is the most common codon for translation initiation and is encoded by the DNA triplet 3′-TAC-5′. | |
| 4. Survey the putative template strand identified in the previous step and determine if DNA triplets 3′-ATC-5′, 3′-ACT-5′, and 3′-ATT-5′ encoding possible stop codons occur as the seventh codon of an mRNA sequence. | 4. Since just one DNA triplet encoding a start codon is present, a scan of the strand at the correct distance from the start codon does find a 3′-ATC-5′ triplet sequence encoding a UAG stop codon:<br>5′-CCCAGC CTA GCCTTTGCAAGAGGC CAT ATCGAC-3′ |
| **TIP:** The stop codons UAG, UGA, and UAA are encoded by DNA triplets 3′-ATC-5′, 3′-ACT-5′, and 3′-ATT-5′. | |
| **Solve** | |
| 5. Identify the mRNA sequence encoding the six amino acids of the polypeptide. | Answer a<br>5. The mRNA sequence is<br>5′-AUG GCC UCU UGC AAA GGC UAG-3′ |
| **TIP:** The mRNA sequence can be determined from either the coding strand or the template strand of DNA. | |
| **TIP:** Substituting U for T on the coding strand produces mRNA sequence. Alternatively, arranging RNA nucleotides complementary to the template strand and assigning antiparallel polarity produces mRNA. | |
| 6. List the amino acid sequence of the polypeptide. | Answer b<br>6. The polypeptide sequence is<br>N-Met-Ala-Ser-Cys-Lys-Gly-C |
| 7. Identify the effect of the G ⟶ A base substitution on the polypeptide. | Answer c<br>7. Substituting the first transcribed G ⟶ A alters the second codon of mRNA by changing GCC ⟶ GUC and substitutes valine (Val) for alanine (Ala) in the second position of the polypeptide sequence. |

For more practice, see Problems 1, 28, 30, and 31.

Visit the Study Area to access study tools. MasteringGenetics™

Two important conclusions come from this experiment. First, the genetic code derives its specificity through the complementary base-pair interaction of tRNA and mRNA. The amino acid carried by charged tRNA does not play a role in determining which amino acids are incorporated into polypeptides. Rather, tRNA alone—acting through the base-pairing interaction of its anticodon with the codon of mRNA—gives specificity to the genetic code. Second, these findings show the importance of the fidelity with which aminoacyl-tRNA synthetases correctly recognize their cognate tRNAs and charge them with the proper amino acid.

## 9.6 Translation Is Followed by Polypeptide Folding, Processing, and Protein Sorting

Translation produces polypeptides, but the production of functional proteins is not complete until the polypeptides are folded into their functional tertiary or quaternary structures. Recall from Section 9.1 that these steps involve the formation of ionic or covalent bonds, and they may also involve specific chemical modifications of amino acids in polypeptides. In addition, two other categories of posttranslational events provide further modifications and sort the proteins for transport to their destinations.

### Posttranslational Polypeptide Processing

The removal of one or more amino acids from a polypeptide is a common form of **posttranslational polypeptide processing.** Earlier in the chapter, we identified AUG as the usual start codon and noted that it encodes the modified amino acid N-formylmethionine (fMet) in bacterial cells and methionine in eukaryotes. Yet fMet is never found in functional bacterial proteins, and amino acids other than methionine are frequently the first amino acid of polypeptides in eukaryotes. The absence of fMet from functional bacterial proteins is the result of posttranslational cleavage of fMet from each bacterial polypeptide (Figure 9.20a). Similarly, methionine is usually removed as part of posttranslational processing in eukaryotes, and the new N-terminal amino acid is acetylated as part of the process.

In addition to N-terminal amino acids, other amino acid residues can be chemically modified as well. One of the most common modifications of individual amino acids is performed by enzymes known as kinases that carry out phosphorylation of proteins by adding a phosphate group to individual amino acids (Figure 9.20b). This is an important regulatory process that can switch a protein from an inactive to an active form, or vice versa. Other enzymes may add methyl groups, hydroxyl groups, or acetyl groups to individual amino acids of polypeptides. The addition of carbohydrate side chains to polypeptides to form a glycoprotein is another important kind of posttranslational modification. For example, in one kind of

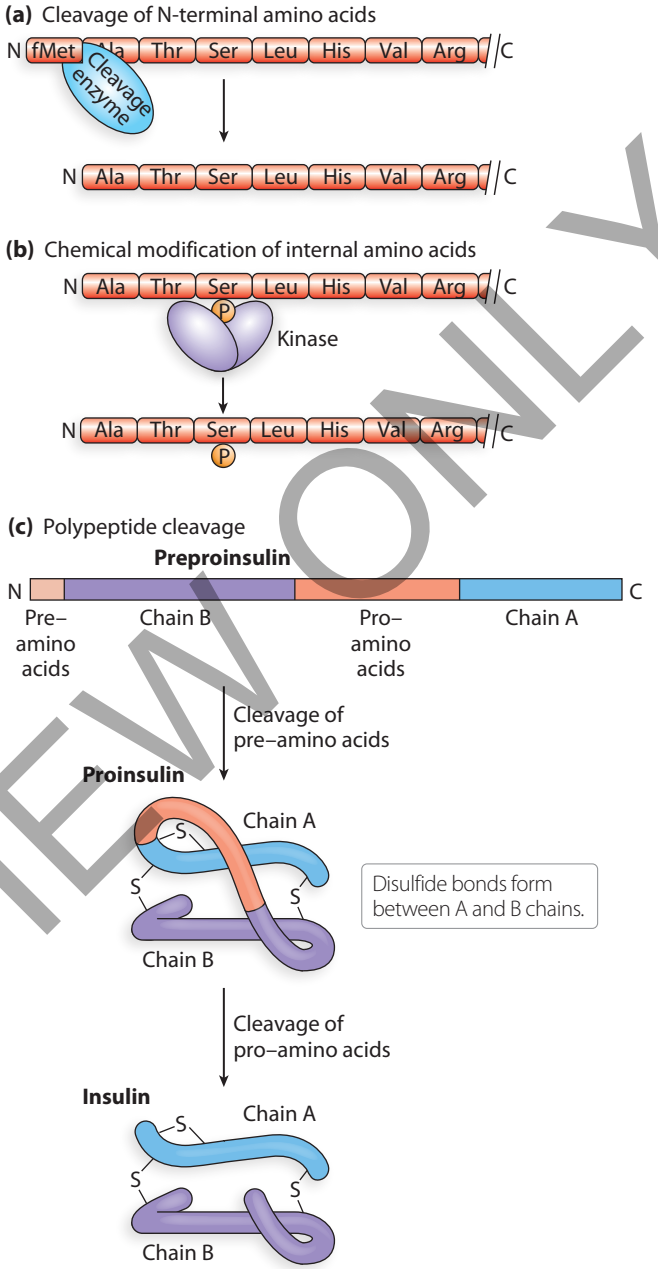



**Figure 9.20** **Examples of posttranslational processing.**

posttranslational modification, the H substance is altered by the protein products of the $I^A$ and $I^B$ alleles of the ABO blood group gene (see Section 4.1).

Posttranslational processing may also include the cleavage of a polypeptide into multiple segments that each form functional proteins or that aggregate after elimination of one or more segments to form a functional protein. Production of the hormone insulin, which facilitates transport of glucose into cells, includes two posttranslational modification steps that remove segments of the original polypeptide (Figure 9.20c). The polypeptide product translated from the insulin gene is called preproinsulin. It is an inactive protein that contains a leader segment, called the pre–amino acid segment, at

the N-terminal end and a connecting segment, called the pro–amino acid segment, that separates the A-chain segment and the B-chain segment, the two functional pieces of the polypeptide. During posttranslational processing of preproinsulin, the pre–amino acids of the signal sequence are removed, after the polypeptide is transported through the cell membrane, to form proinsulin. Three disulfide bonds form within and between the A-chain and B-chain segments, followed by polypeptide cleavage that removes the pro–amino acid segment. What results is a functional insulin molecule consisting of 20 amino acids in the A-chain segment and 31 amino acids in the B-chain segment.

## The Signal Hypothesis

Like the passengers in a busy airline terminal, the proteins produced in a cell have different destinations, to which they travel with the aid of a "ticket" that tells the cell where to transport them. The destination is often an organelle or the cell membrane; in certain cases, the polypeptide is destined for transport out of the cell. The ticket that communicates the destination of a polypeptide is a **signal sequence** of 15 to 20 or so amino acids at the N-terminal end.

First articulated in the early 1970s by Gunther Blobel, the **signal hypothesis** proposes that the first 15 to 20 amino acids of many polypeptides contain an "address label" in the form of a signal sequence that designates the protein's destination in the cell. Blobel's hypothesis proposed that the signal sequence directs proteins to the endoplasmic reticulum (ER), where they are sorted for their cellular destinations.

Blobel's signal hypothesis is now a widely accepted model for the identification of the cellular destinations of proteins. In fact, follow-up research has identified the mechanism by which proteins are processed and packaged for export from a cell. While proteins destined to remain in a cell are typically translated at "free" ribosomes (ribosomes that float freely in the cytoplasm), large numbers of ribosomes are attached to the rough endoplasmic reticulum (rough ER) where proteins destined for intercellular transport are translated. Figure 9.21 illustrates the translation of polypeptides

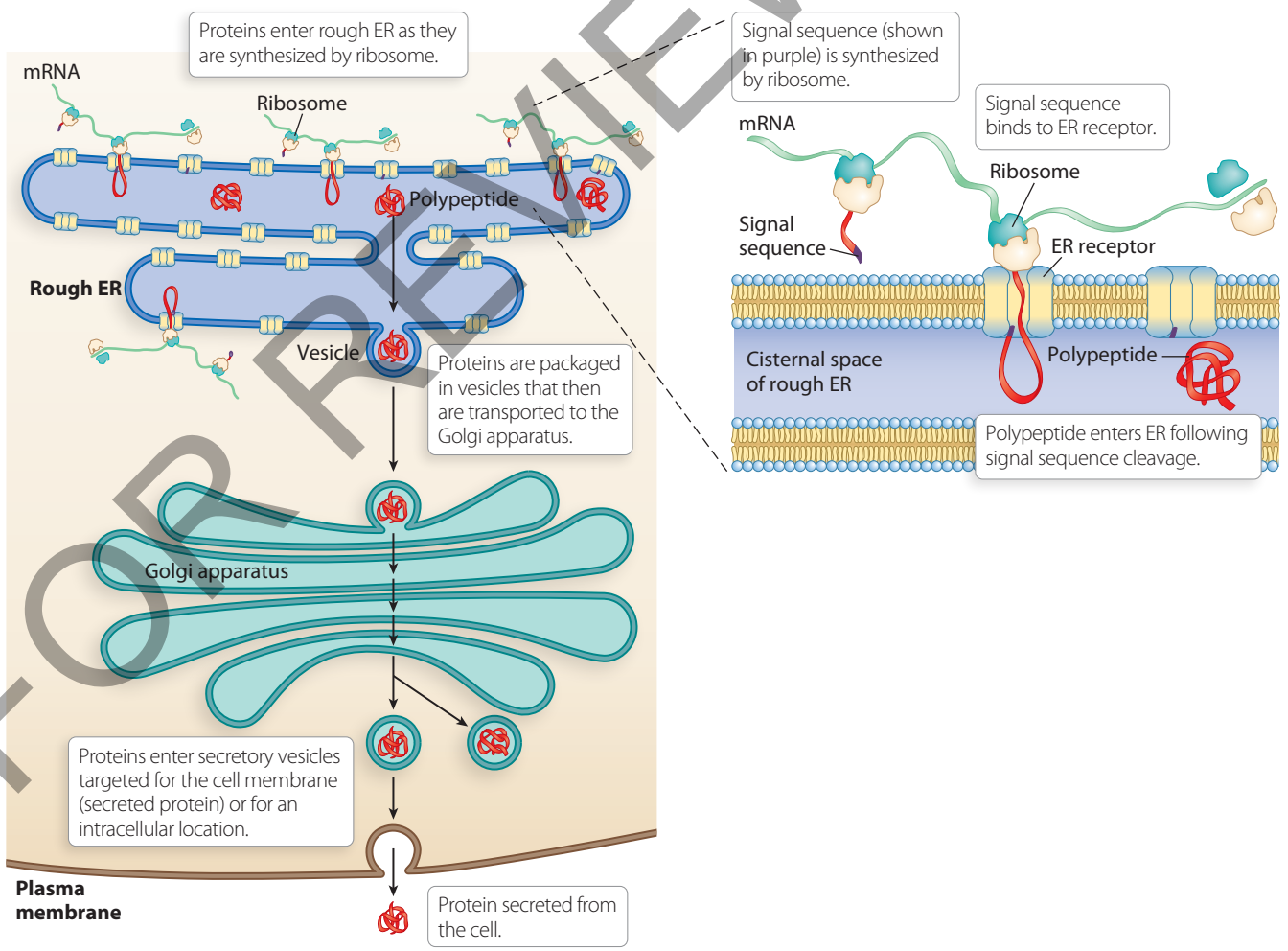



**Figure 9.21 Proteins enter the endoplasmic reticulum (ER).** Translated proteins enter the cisternal space of the ER through receptors that cleave the signal sequences to begin the protein-sorting process.

into the cisternal space of the rough ER where the polypeptides are processed and packaged for transport to the Golgi apparatus. In the Golgi apparatus additional protein processing takes place and the proteins are packaged into vesicles for transport to the intercellular destinations.

## CASE STUDY

### Antibiotics and Translation Interference

We have all taken antibiotics at various times during our lives to counteract a painful or persistent microbial infection. As a result of the efficiency of these compounds, we have experienced rapid relief of symptoms and elimination of the infection. These beneficial effects are accomplished by selective cell death or through blocking cell proliferation. Specifically, the antibiotic kills microorganisms without harming our own cells in the process or they act to prevent further microbial cell growth. What is the biochemical basis of antibiotic action? How do antibiotic compounds specifically target microbial cells for destruction?

**PROTEIN SYNTHESIS INHIBITION BY ANTIBIOTIC COMPOUNDS** You will probably not be surprised to learn that different antibiotics target different aspects of microbe biology to inhibit. But you may be surprised to learn that many different antibiotics target microbial translation as their mode of action (**Table 9.10**). Familiar antibiotics such as tetracycline, streptomycin, and chloramphenicol target different stages of microbial translation, as do less familiar antibiotics such as erythromycin, puromycin, and cycloheximide. Each antibiotic contains a different active compound that takes advantage of unique features of bacterial translation to disrupt the production of bacterial proteins while not interfering with the translation of proteins in our cells.

**Table 9.10** Antibiotic Inhibitors of Protein Synthesis

| Antibiotic | Inhibitory Action |
|---|---|
| Chloramphenicol | Blocks polypeptide formation by inhibiting peptidyl transferase in the 70S ribosome (antibacterial action) |
| Erythromycin | Blocks translation by binding to 50S subunit and inhibiting polypeptide release (antibacterial action) |
| Streptomycin | Inhibits translation initiation and causes misreading of mRNA by binding to the 30S subunit (antibacterial action) |
| Tetracycline | Binds to the 30S subunit and inhibits binding of charged tRNAs (antibacterial action) |
| Cycloheximide | Blocks polypeptide formation by inhibiting peptidyl transferase activity in the 80S ribosome (antieukaryote action) |
| Puromycin | Causes premature termination of translation by acting as an analog of charged tRNA (antibacterial and antieukaryote action) |

**TRANSLATION DISRUPTION BY AMINOGLYCOSIDES** *Streptomycin* is one of several antibiotics in a class of biochemical compounds called *aminoglycosides*. Streptomycin inhibits bacterial translation by interfering with binding of N-formylmethionine tRNA to the ribosome, thus preventing the initiation of translation. Streptomycin can also cause misreading of mRNA during translation by generating mispairing between codons and anticodons. For example, the codon UUU normally specifies phenylalanine, but streptomycin induces pairing between a UUU codon and the tRNA carrying isoleucine, whose codon is AUU. This error leads to amino acid changes in proteins and potentially to defective protein activity. Other aminoglycosides, such as neomycin, kanamycin, and gentamycin, also cause mispairing between codons and anticodons and can generate defective proteins. *Erythromycin* also impairs bacterial translation, but it does so in a very different way. It binds to the 50S (large) subunit in the tunnel from which the newly synthesized polypeptide emerges. In this manner, erythromycin blocks the passage of the polypeptide out of the ribosome. This causes the ribosome to stall on mRNA, bringing translation to a halt. Table 9.10 provides details about these and other actions of antibacterial agents.

**TRANSLATION BLOCKAGE BY ANTIFUNGAL COMPOUNDS** Single-celled eukaryotic microorganisms, such as fungi, can also cause human infections. To fight these infections, antibiotics such as puromycin and cycloheximide that target translational activities of eukaryotic cells are used. *Puromycin* has a three-dimensional structure similar to that of the 3′ end of a charged tRNA. It stops translation of bacterial and eukaryotic mRNAs by binding at the ribosomal A site and acting as an analog of charged tRNA. When puromycin is bound at the A site, its amino group forms a peptide bond with the carboxyl group of the P-site amino acid. However, puromycin does not contain a carboxyl group. This difference prevents formation of any additional peptide bonds and puts an end to translation. *Cycloheximide* exclusively blocks eukaryotic translation by binding to the 60S subunit and inhibiting peptidyl transferase activity, much like chloramphenicol does to bacterial peptidyl transferase.

## SUMMARY

MasteringGenetics™ For activities, animations, and review quizzes, go to the Study Area.

### 9.1 Polypeptides Are Composed of Amino Acid Chains That Are Assembled at Ribosomes

- Polypeptides contain 20 kinds of amino acids that carry side chains, giving them specific properties.
- Translation takes place at the ribosome, where mRNA codons are coupled to transfer RNA anticodons by complementary base pairing.
- Polypeptides have four structural levels: the amino acid order (primary), intrachain folding (secondary), three-dimensional functional folding (tertiary), and multimeric protein structure (quaternary).
- Polypeptides have an N-terminal (amino) end and a C-terminal (carboxyl) end.
- Ribosomes are composed of two subunits that each consist of ribosomal RNA and numerous proteins.
- Ribosomes have three functional sites of action: the P site, where the polypeptide is held; the A site, where tRNA molecules bind to add their amino acid to the end of the polypeptide; and the E site, which provides an exit point for uncharged tRNAs.

### 9.2 Translation Occurs in Three Phases

- Bacterial translation is initiated with the binding of the Shine–Dalgarno sequence on the 5′ mRNA end to a complementary sequence of nucleotides on the 3′ end of the 16S rRNA in the small ribosomal subunit. The nearby start codon is the site where translation commences.
- In eukaryotic mRNA, the 5′ cap is the binding site for eukaryotic initiation factors that cause the small ribosomal subunit to begin scanning in search of the start codon, which is part of the Kozak sequence.
- Archaea carry multiple translation-initiation factors that are homologous to eukaryotic initiation factors, but they also produce a high proportion of leaderless mRNAs that have an unknown translation-initiation mechanism.
- During polypeptide synthesis, charged tRNAs enter the A site, and peptidyl transferase catalyzes peptide bond formation, transferring the polypeptide from the A-site tRNA to the P-site tRNA. Elongation factor proteins translocate the ribosome, shifting the tRNA–polypeptide complex from the A site to the P site and opening the A site for the next charged tRNA.
- Translation terminates when a stop codon enters the A site. Release factor proteins, rather than tRNA, bind to stop codons. Release factors cause release of the polypeptide and lead to the dissociation of the ribosome from mRNA.

### 9.3 Translation Is Fast and Efficient

- An mRNA undergoes simultaneous translation by several ribosomes that attach to it sequentially to form a polyribosome.
- Usually, a ribosome will dissociate from mRNA upon encountering a stop codon, but the small size of some intercistronic spacers in bacterial polycistronic mRNAs permits a ribosome to translate two or more polypeptides sequentially from the mRNA before dissociating.
- The evolutionary evidence derived from homologies among translationally active proteins of members of the three domains of life suggests that archaea are more closely related to eukaryotes than they are to bacteria.

### 9.4 The Genetic Code Translates Messenger RNA into Polypeptide

- The genetic code is redundant, meaning that most amino acids are specified by more than one codon. Redundancy of the genetic code is made possible by third-base wobble that relaxes the strict complementary base-pairing requirements at the third base of the codon.
- Specialized enzymes called aminoacyl-tRNA synthetases catalyze the addition of a specific amino acid to each tRNA.

### 9.5 Experiments Deciphered the Genetic Code

- In vitro experimental analysis demonstrates that the genetic code is triplet and does not contain gaps or overlaps.
- Each mRNA codon is composed of three consecutive nucleotides. Of the 64 codons contained in the genetic code, 61 specify amino acids and 3 are stop codons.
- The genetic code was deciphered by analysis of in vitro translation of synthetic messenger RNA.
- The genetic code is essentially universal among living organisms. The few exceptions to the genetic code are found mainly in mitochondria.
- Properly charged tRNAs play the central role in converting mRNA sequence into polypeptide sequence.

### 9.6 Translation Is Followed by Polypeptide Folding, Processing, and Protein Sorting

- Formation of functional proteins occurs after translation is completed and may be aided by ribosome-associated proteins or by separate protein complexes.
- Proteins in eukaryotic cells are sorted to their cellular destinations by signal sequences at their N-terminal ends. Signal sequences are removed from polypeptides in the ER, and polypeptides destined for different sites in the cell are differentially glycosylated before being packaged for transport to the Golgi apparatus.
- In the Golgi apparatus, polypeptides are packaged into transport vesicles for shipment to their cellular destinations.

## KEYWORDS

3′ untranslated region (3′ UTR) *(p. 307)*
5′ untranslated region (5′ UTR) *(p. 307)*
30S initiation complex *(p. 313)*
70S initiation complex *(p. 313)*
aminoacyl site (A site) *(p. 309)*
α-helix (alpha helix) *(p. 308)*
aminoacyl-tRNA synthetase (tRNA synthetase) *(p. 322)*
archaeal initiation factor (aIF) *(p. 314)*
β-pleated sheet (beta-pleated sheet) *(p. 308)*
charged tRNA *(p. 311)*
elongation factor (EF) *(p. 315)*
eukaryotic initiation factor (eIF) *(p. 313)*
exit site (E site) *(p. 309)*
frameshift mutation *(p. 324)*
initiation complex *(p. 313)*
initiation factor (IF) *(p. 312)*
initiator tRNA *(p. 311)*
inosine (I) *(p. 322)*
isoaccepting tRNA *(p. 321)*
Kozak sequence *(p. 313)*
large ribosomal subunit *(p. 309)*
N-formylmethionine (fMet; $tRNA^{fMet}$) *(p. 313)*
peptide bond *(p. 306)*
peptidyl site (P site) *(p. 309)*
polypeptide *(p. 306)*
polycistronic mRNA *(p. 320)*
polyribosome *(p. 319)*
posttranslational polypeptide processing *(p. 330)*
preinitiation complex *(p. 313)*
primary structure *(p. 308)*
quaternary structure *(p. 308)*
R-group *(p. 306)*
release factor (RF) *(p. 318)*
reading frame *(p. 324)*
reversion mutation *(p. 324)*
scanning *(p. 313)*
secondary structure *(p. 308)*
Shine–Dalgarno sequence *(p. 313)*
signal hypothesis *(p. 331)*
signal sequence (leader sequence) *(p. 331)*
small ribosomal subunit *(p. 309)*
synonymous codon *(p. 321)*
tertiary structure *(p. 308)*
third-base wobble *(p. 321)*
uncharged tRNA *(p. 311)*

## PROBLEMS

MasteringGenetics™ Visit for instructor-assigned tutorials and problems.

### Chapter Concepts

*For answers to selected even-numbered problems, see Appendix: Answers.*

1. Some proteins are composed of two or more polypeptides. Suppose the DNA template strand sequence 3′-TACGTAGGCTAACGGAGTAAGCTAACT-5′ produces a polypeptide that joins in pairs to form a functional protein.
   a. What is the amino acid sequence of the polypeptide produced from this sequence?
   b. What term is used to identify a functional protein like this one formed when two identical polypeptides join together?
2. In the experiments that deciphered the genetic code, many different synthetic mRNA sequences were tested.
   a. Describe how the codon for phenylalanine was identified.
   b. What was the result of studies of synthetic mRNAs composed exclusively of cytosine?
   c. What result was obtained for synthetic mRNAs containing AG repeats, that is,

      AGAGAGAG…?

   d. Predict the results of experiments examining GCUA repeats.
3. Several lines of experimental evidence pointed to a triplet genetic code. Identify three pieces of information that supported the triplet hypothesis of genetic code structure.
4. Outline the events that occur during initiation of translation in *E. coli*.
5. A portion of a DNA template strand has the base sequence

   5′-…ACGCGATGCGTGATGTATAGAGCT…-3′

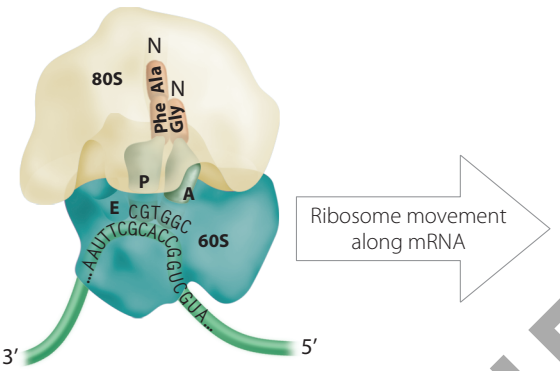
   a. Identify the sequence and polarity of the mRNA transcribed from this fragmentary template strand sequence.
   b. Determine the amino acid sequence encoded by this fragment. Identify the N- and C-terminal directions of the polypeptide.
   c. Which is the third amino acid added to the polypeptide chain?
6. Describe three features of tRNA molecules that lead to their correct charging by tRNA synthetase enzymes.
7. Identify the amino acid carried by tRNAs with the following anticodon sequences.
   a. 5′-UAG-3′
   b. 5′-AAA-3′
   c. 5′-CUC-3′
   d. 5′-AUG-3′
   e. 5′-GAU-3′
8. For each of the anticodon sequences given in the previous problem, identify the other codon sequence to which it could potentially pair using third base wobble.
9. What is the role of codons UAA, UGA, and UAG in translation? What events occur when one of these codons appears at the A site of the ribosome?
10. Compare and contrast the composition and structure of bacterial and eukaryotic ribosomes, identifying at least three features that are the same and three features that are unique to each type of ribosome.

11. Consider translation of the following mRNA sequence:

    5′-…AUGCAGAUCCAUGCCUAUUGA…-3′

    a. Diagram translation at the moment the fourth amino acid is added to the polypeptide chain. Show the ribosome; label its A, P, and E sites; show its direction of movement; and indicate the position and anticodon triplet sequence of tRNAs that are currently interacting with mRNA codons.
    b. What is the anticodon triplet sequence of the next tRNA to interact with mRNA?
    c. What events occur to permit the next tRNA to interact with mRNA?

12. The diagram of a eukaryotic ribosome shown below contains several errors.




    a. Examine the diagram carefully, and identify each error.
    b. Redraw the diagram, and correct each error using the mRNA sequence shown.

13. Third-base wobble allows some tRNAs to recognize more than one mRNA codon. Based on this chapter's discussion of wobble, what is the *minimal* number of tRNA molecules necessary to recognize the following amino acids?

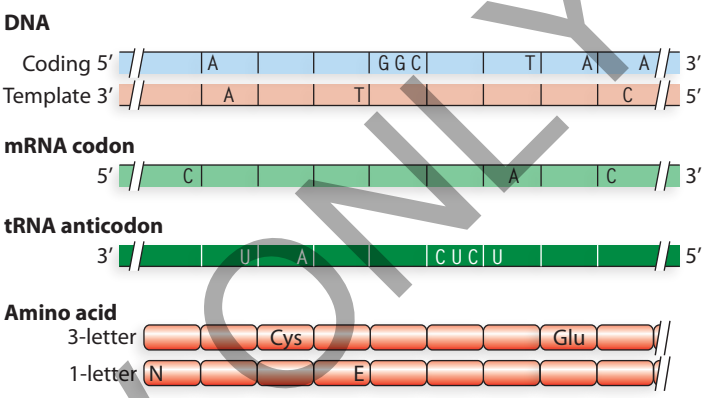
    a. leucine
    b. arginine
    c. isoleucine
    d. lysine

14. The genetic code contains 61 codons to specify the 20 common amino acids. Many organisms carry fewer than 61 different tRNA genes in their genomes. These genomes take advantage of isoaccepting tRNAs and the rules governing third-base wobble to encode fewer than 61 tRNA genes. Use these rules to calculate the *minimal* number of tRNA genes required to specify all 20 of the common amino acids.

15. The three major forms of RNA (mRNA, tRNA, and rRNA) interact during translation.

    a. Describe the role each form of RNA performs during translation.
    b. Which of the three types of RNA might you expect to be the least stable? Why?
    c. Which form of RNA is least stable in eukaryotes? Why is this form least stable?
    d. Compared to the average stability of mRNA in *E. coli*, is mRNA in a typical human cell more stable or less stable? Why?

16. The figure below contains sufficient information to fill in every row. Use the information provided to complete the figure.




17. The line below represents a mature eukaryotic mRNA. The accompanying list contains many sequences or structures that are part of eukaryotic mRNA. A few of the items in the list, however, are not found in eukaryotic mRNA. As accurately as you can, show the location, on the line, of the sequences or structures that belong on eukaryotic mRNA; then, separately, list the items that are not part of eukaryotic mRNA.

    5′ _____ 3′

    a. stop codon
    b. poly-A tail
    c. intron
    d. 3′ UTR
    e. promoter
    f. start codon
    g. AAUAAA
    h. 5′ UTR
    i. 5′ cap
    j. termination sequence

18. After completing Problem 17, carefully draw a line below the mRNA to represent its polypeptide product in accurate alignment with the mRNA. Label the N-terminal and C-terminal ends of the polypeptide. Carefully draw two lines above and parallel to the mRNA, and label them "coding strand" and "template strand." Locate the DNA promoter sequence. Identify the locations of the +1 nucleotide and of a transcription termination sequence.

19. Define and describe the differences in the primary, secondary and tertiary structures of a protein.

20. Describe the roles and relationships between

    a. tRNA synthetases and tRNA molecules.
    b. tRNA anticodon sequences and mRNA codon sequences.

## Application and Integration

*For answers to selected even-numbered problems, see Appendix: Answers.*

21. In an experiment to decipher the genetic code, a poly-AC mRNA (ACACACAC…) is synthesized. What pattern of amino acids would appear if this sequence were to be translated by a mechanism that reads the genetic code as
    a. a doublet without overlaps?
    b. a doublet with overlaps?
    c. a triplet without overlaps?
    d. a triplet with overlaps?
    e. a quadruplet without overlaps?
    f. a quadruplet with overlaps?

22. Identify and describe the steps that lead to the secretion of proteins from eukaryotic cells.

23. The amino acid sequence of a portion of a polypeptide is

    N…Cys-Pro-Ala-Met-Gly-His-Lys…C.

    a. What is the mRNA sequence encoding this polypeptide fragment? Use N to represent any nucleotide, Pu to represent a purine, and Py to represent a pyrimidine. Label the 5′ and 3′ ends of the mRNA.
    b. Give the DNA template and coding strand sequences corresponding to the mRNA. Use the N, Pu, and Py symbols as placeholders.

24. Har Gobind Khorana and his colleagues performed numerous experiments translating synthetic mRNAs. In one experiment, an mRNA molecule with a repeating UG dinucleotide sequence was assembled and translated.
    a. Write the sequence of this mRNA and give its polarity.
    b. What is the sequence of the resulting polypeptide?
    c. How did the polypeptide composition help confirm the triplet nature of the genetic code?
    d. If the genetic code were a doublet code instead of a triplet code, how would the result of this experiment be different?
    e. If the genetic code was overlapping rather than nonoverlapping, how would the result of this experiment be different?

25. An experiment by Khorana and his colleagues translated a synthetic mRNA containing repeats of the trinucelotide UUG.
    a. How many reading frames are possible in this mRNA?
    b. What is the result obtained from each reading frame?
    c. How does the result of this experiment help confirm the triplet nature of the genetic code?

26. The human β-globin polypeptide contains 146 amino acids. How many mRNA nucleotides are required to encode this polypeptide?

27. The mature mRNA transcribed from the human β-globin gene is considerably longer than the sequence needed to encode the 146–amino acid polypeptide. Give the names of three sequences located on the mature β-globin mRNA but not translated.

28. Figure 9.7 contains several examples of the Shine–Dalgarno sequence. Using the seven Shine–Dalgarno sequences from *E. coli*, determine the consensus sequence and identify its location relative to the start codon.

29. Figure 9.20 shows three posttranslational steps required to produce the sugar-regulating hormone insulin from the starting polypeptide product preproinsulin.
    a. A research scientist is interested in producing human insulin in the bacterial species *E. coli*. Will the genetic code allow the production of human proteins from bacterial cells? Explain why or why not.
    b. Explain why it is not feasible to insert the entire human insulin gene into *E. coli* and anticipate the production of insulin.
    c. Recombinant human insulin (made by inserting human DNA encoding insulin into *E. coli*) is one of the most widely used recombinant pharmaceutical products in the world. What segments of the human insulin gene are used to create recombinant bacteria that produce human insulin?

30. A DNA sequence encoding a five–amino acid polypeptide is given below.

    …ACGGCAAGATCCCACCCTAATCAGACCGTACCATTCACCTCCT…
    …TGCCGTTCTAGGGTGGGATTAGTCTGGCATGGTAAGTGGAGGA…

    a. Locate the sequence encoding the five amino acids of the polypeptide, and identify the template and coding strands of DNA.
    b. Give the sequence and polarity of the mRNA encoding the polypeptide.
    c. Give the polypeptide sequence, and identify the N-terminus and C-terminus.
    d. Assuming the sequence above is a bacterial gene, identify the region encoding the Shine–Dalgarno sequence.
    e. What is the function of the Shine–Dalgarno sequence?

31. A portion of the coding strand of DNA for a gene has the sequence

    5′-…GGAGAGAATGAATCT…-3′

    a. Write out the template DNA strand sequence and polarity as well as the mRNA sequence and polarity for this gene segment.
    b. Assuming the mRNA is in the correct reading frame, write the amino acid sequence of the polypeptide using three-letter abbreviations and, separately, the amino acid sequence using one-letter abbreviations.

32. A eukaryotic mRNA has the following sequence. The 5′ cap is indicated in italics (*CAP*), and the 3′ poly(A) tail is indicated by italicized adenines.

    5′-*CAP*CCAAGCGUUACAUGUAUGGAGAGAAUGAAACUG-
    AGGCUUGCCACGUUUGUUAAGCACCUAUGCUACCG*AAAAAAA*
    *AAAAAAAAAAAAAAAAA*-3′

    a. Locate the start codon and stop codon in this sequence.
    b. Determine the amino acid sequence of the polypeptide produced from this mRNA. Write the sequence using the three-letter and one-letter abbreviations for amino acids.

33. Diagram a eukaryotic gene containing three exons and two introns, the pre-mRNA and mature mRNA transcript of the gene, and a partial polypeptide that contains the following sequences and features. Carefully align the nucleic acids, and locate each sequence or feature on the appropriate molecule.
    a. the AG and GU dinucleotides corresponding to intron–exon junctions
    b. the +1 nucleotide
    c. the 5′ UTR and the 3′ UTR
    d. the start codon sequence
    e. a stop codon sequence
    f. a codon sequence for the amino acids Gly-His-Arg at the end of exon 1 and a codon sequence for the amino acids Leu-Trp-Ala at the beginning of exon 2

34. The following table contains DNA-sequence information compiled by Marilyn Kozak (1987). The data consist of the percentage of A, C, G, and T at each position among the 12 nucleotides preceding the start codon in 699 genes from various vertebrate species, and as the first nucleotide after the start codon. The start codon occupies positions +1 to +3, and the +4 nucleotide occurs immediately after the start codon. Use the data to determine the consensus sequence for the 13 nucleotides (−12 to −1 and +4) surrounding the start codon in vertebrate genes.

| Position | −12 | −11 | −10 | −9 | −8 | −7 | −6 | −5 | −4 | −3 | −2 | −1 | [start] +4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Percent A | 23 | 26 | 25 | 23 | 19 | 23 | 17 | 18 | 25 | 61 | 27 | 15 | [AUG] 23 |
| Percent C | 35 | 35 | 35 | 26 | 39 | 37 | 19 | 39 | 53 | 2 | 49 | 55 | [AUG] 16 |
| Percent G | 23 | 21 | 22 | 33 | 23 | 20 | 44 | 23 | 15 | 36 | 13 | 21 | [AUG] 46 |
| Percent T | 19 | 18 | 18 | 18 | 19 | 20 | 20 | 20 | 7 | 1 | 11 | 9 | [AUG] 15 |

35. The following table lists α-globin and β-globin gene sequences for the 12 nucleotides preceding the start codon and the first nucleotide following the start codon. The data are for 16 vertebrate globin genes reported by Kozak (1987). The sequences are written from −12 to +4 with the start codon sequence in capital letters.

| | Gene Sequence |
|---|---|
| | −12 start +4 |
| ***α-Globin Family*** | |
| Human adult | agagaacccaccATGg |
| Human embryonic | caccctgccgccATGt |
| Baboon | ccagcgcgggcATGg |
| Mouse adult | caggaagaaaccATGg |
| Rabbit adult | gaaggaaccaccATGg |
| Goat embryonic | tcagctgccaccATGt |
| Duck adult | ggagctgcaaccATGg |
| Chicken embryonic | ctctcctgcacaATGg |
| ***β-Globin Family*** | |
| Human fetal | agtccagacgccATGg |
| Human embryonic | aggcctggcatcATGg |
| Rabbit adult | aaacagacagaATGg |
| Rabbit embryonic | agaccagacatcATGg |
| Chicken adult | ccaaccgccgccATGg |
| Chicken embryonic | cccgctgccaccATGg |
| *Xenopus* adult | tcaactttggccATGg |
| *Xenopus* larval | tctacagccaccATGg |

Use the data in this table to
    a. Determine the consensus sequence for the 16 selected α-globin and β-globin genes.
    b. Compare the consensus sequence for these globin genes to the consensus sequence derived from the larger study of 699 vertebrate genes in Problem 34.

36. The six nucleotides preceding the start codon and the first nucleotide after the start codon in eukaryotes exhibit strong sequence preference as determined by the percentages of nucleotides in the −6 to −1 positions and the +4 position. Use the data given in the table for Problem 35 to determine the seven nucleotides that most commonly surround the start in vertebrates.

37. In terms of the polycistronic composition of mRNAs and the presence or absence of Shine–Dalgarno sequences, compare and contrast bacterial, archaeal, and eukaryotic mRNAs.

38. Organisms of all three domains of life usually use the mRNA codon AUG as the start codon.
    a. Do organisms of the three domains use the same amino acid as the initial amino acid in translation? Identify similarities and differences.
    b. Despite AUG being the most common start codon sequence, very few proteins have methionine as the first amino acid. Why is this the case?