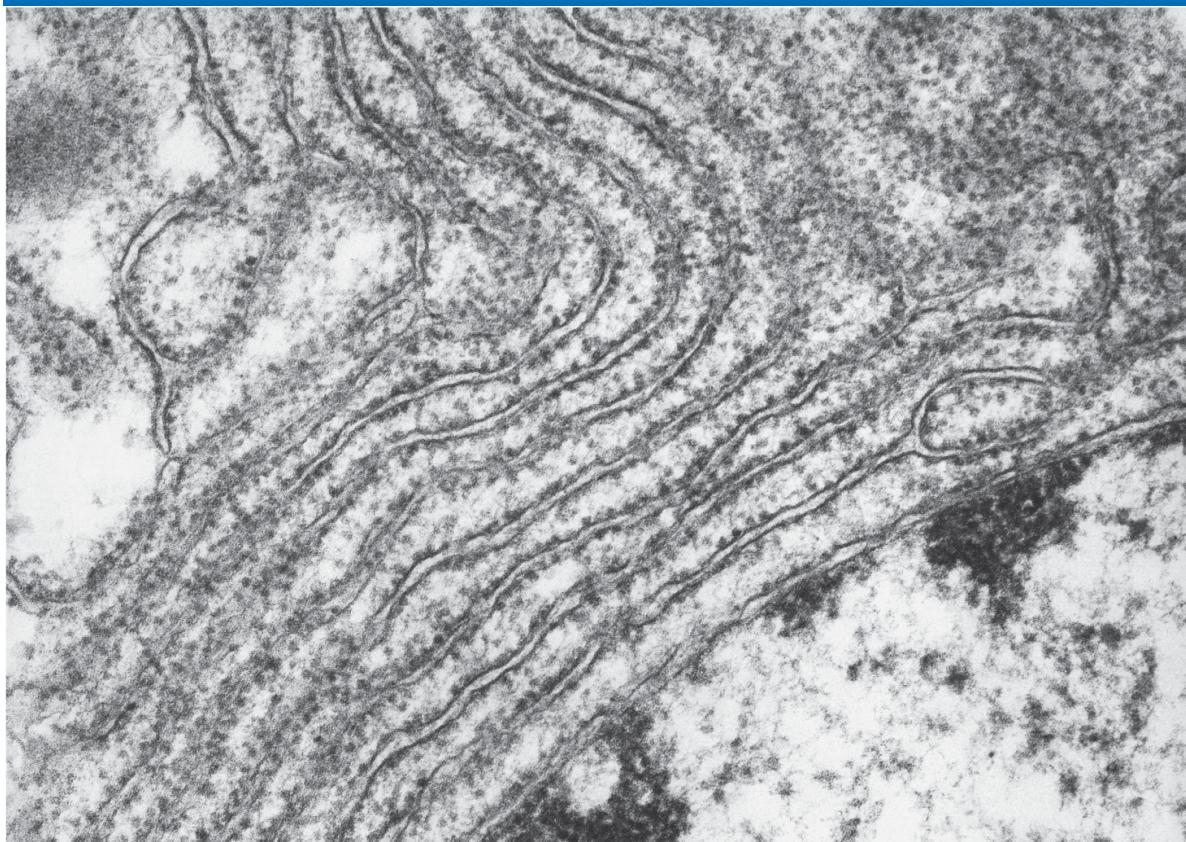


CHAPTER 9

Proteins and Their Synthesis



JOSEPH F. GENNARO JR./Science Source .

Ribosomes are RNA-protein machines that synthesize proteins in bacteria and eukaryotes. Many ribosomes (rows of black dots) associate with the endoplasmic reticulum in the cytoplasm of eukaryotic cells.

CHAPTER OUTLINE AND LEARNING OBJECTIVES

9.1 PROTEIN STRUCTURE

LO 9.1 Explain how the interactions of amino acids determine the structure of proteins.

9.2 THE GENETIC CODE

LO 9.2 Outline the experimental evidence supporting the rules of the genetic code.

LO 9.3 Describe features of the genetic code that minimize effects of point mutations on protein function.

9.3 tRNAs AND RIBOSOMES

LO 9.4 Explain how the structures of tRNAs and ribosomes determine how they function in protein synthesis.

9.4 TRANSLATION

LO 9.5 Outline the molecular events that take place during translation initiation, elongation, and termination.

9.5 TRANSLATIONAL AND POST-TRANSLATIONAL REGULATION

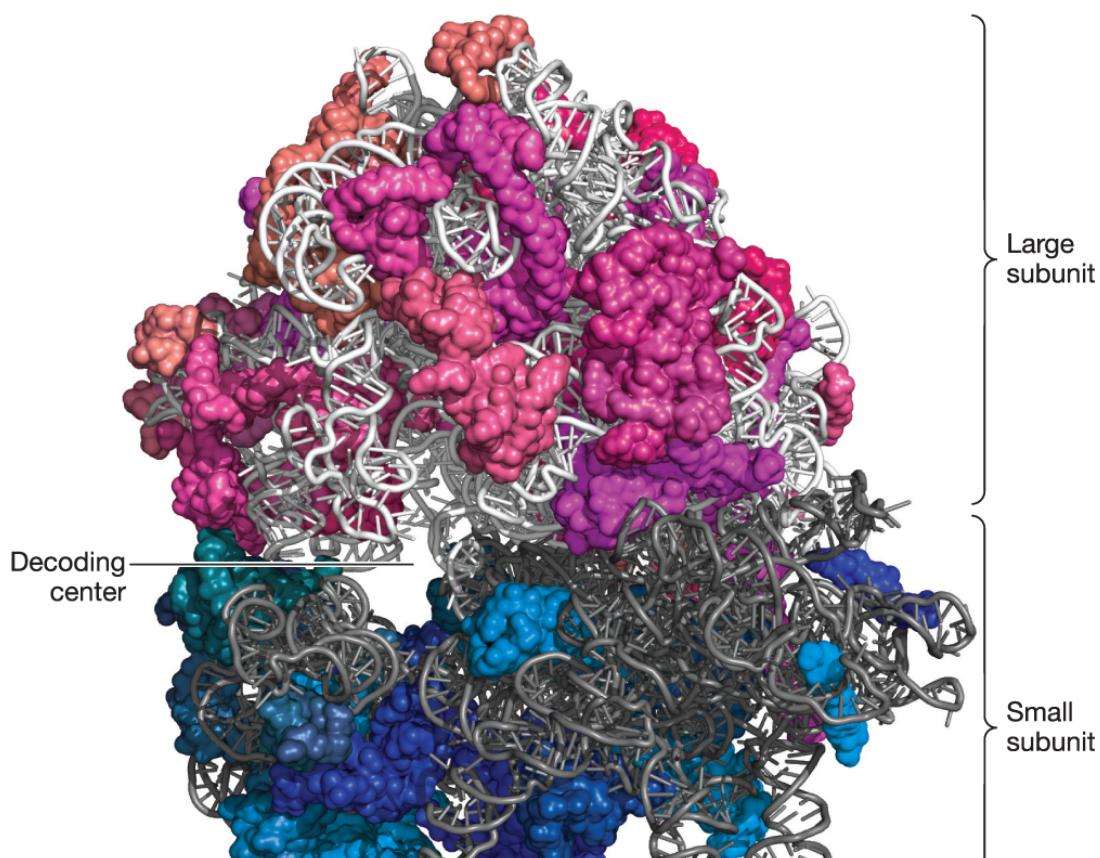
LO 9.6 Describe how protein synthesis and function are regulated.

CHAPTER OBJECTIVE

[Chapters 7](#) and [8](#) describe the first two stages of information transfer shown in [Figure 1-10](#): replication (the synthesis of a DNA copy of DNA) and transcription (the synthesis of an RNA copy of a segment of DNA). In this chapter, you will learn about the final stage of information transfer: translation (the synthesis of a protein copy of an RNA).

RNAs play several important roles in translation. As you learned in [Chapter 8](#), RNAs are classified as either messenger RNA (mRNA) or non-coding RNA (ncRNA). The majority of genes encode mRNAs, whose function is to serve as an intermediate in the synthesis of proteins. In contrast, ncRNAs are active as functional RNAs; they are never translated into proteins. ncRNAs involved in protein synthesis include transfer RNAs and ribosomal RNAs. [Transfer RNAs \(tRNAs\)](#) carry out the decoding work of translation, associating three-nucleotide sequences in an mRNA with their corresponding amino acids. This decoding occurs inside [ribosomes](#), which are composed of several types of [ribosomal RNAs \(rRNAs\)](#) and many different proteins ([Figure 9-1](#)). Ribosomes assemble on mRNAs and catalyze protein synthesis by chemically binding together the amino acids brought to the ribosome by tRNAs. Like tRNAs, ribosomes are general in function, in the sense that they can translate any mRNA.

The structure of the eukaryotic ribosome





PDB ID 4V7R.

FIGURE 9-1 In the yeast ribosome (shown), there are four individual rRNAs and about 80 proteins. RNAs are shown in white and gray in the large and small subunits, respectively. Proteins are shown in shades of red and shades of blue in the large and small subunits, respectively.

Although most genes encode mRNAs, ncRNAs make up the largest fraction of total cellular RNA.

In a typical actively dividing eukaryotic cell, rRNA and tRNA account for almost 95 percent of the total RNA, whereas mRNA accounts for only about 5 percent. Two factors explain the abundance of rRNAs and tRNAs. First, they are much more stable than mRNAs, so they remain intact much longer. Second, ribosomes are an abundant component of cells. There are tens of thousands of ribosomes in bacterial cells, about 200,000 ribosomes in yeast cells, and several million ribosomes in mammalian cells.

Components of the translational machinery and the process of translation are very similar in bacteria and eukaryotes. In addition to ribosomes, tRNAs, and mRNAs, each phase of translation involves a distinct set of protein regulatory factors; **initiation factors (IFs)** start translation at the beginning of the mRNA open reading frame (ORF), **elongation factors (EFs)** maintain translation through the ORF, and **termination factors**, also called **release factors (RFs)**, stop translation at the end of the ORF.

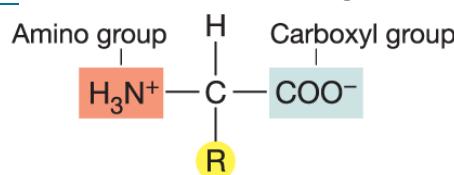
The major feature that distinguishes translation in bacteria from that in eukaryotes is the location where transcription and translation take place in the cell: the two processes take place in the same compartment in bacteria, whereas they are physically separated in eukaryotes by the nuclear membrane. After processing, eukaryotic mRNAs are exported from the nucleus for translation by ribosomes that reside in the cytoplasm. In contrast, transcription and translation are coupled in bacteria: translation of an RNA begins at its 5' end while the rest of the mRNA is still being transcribed.

KEY CONCEPT Translation occurs within ribosomes and requires three types of RNAs: mRNAs carry the sequence information from DNA to ribosomes, tRNAs decode mRNA nucleotide sequences into amino acids, and rRNAs are structural and functional components of ribosomes.

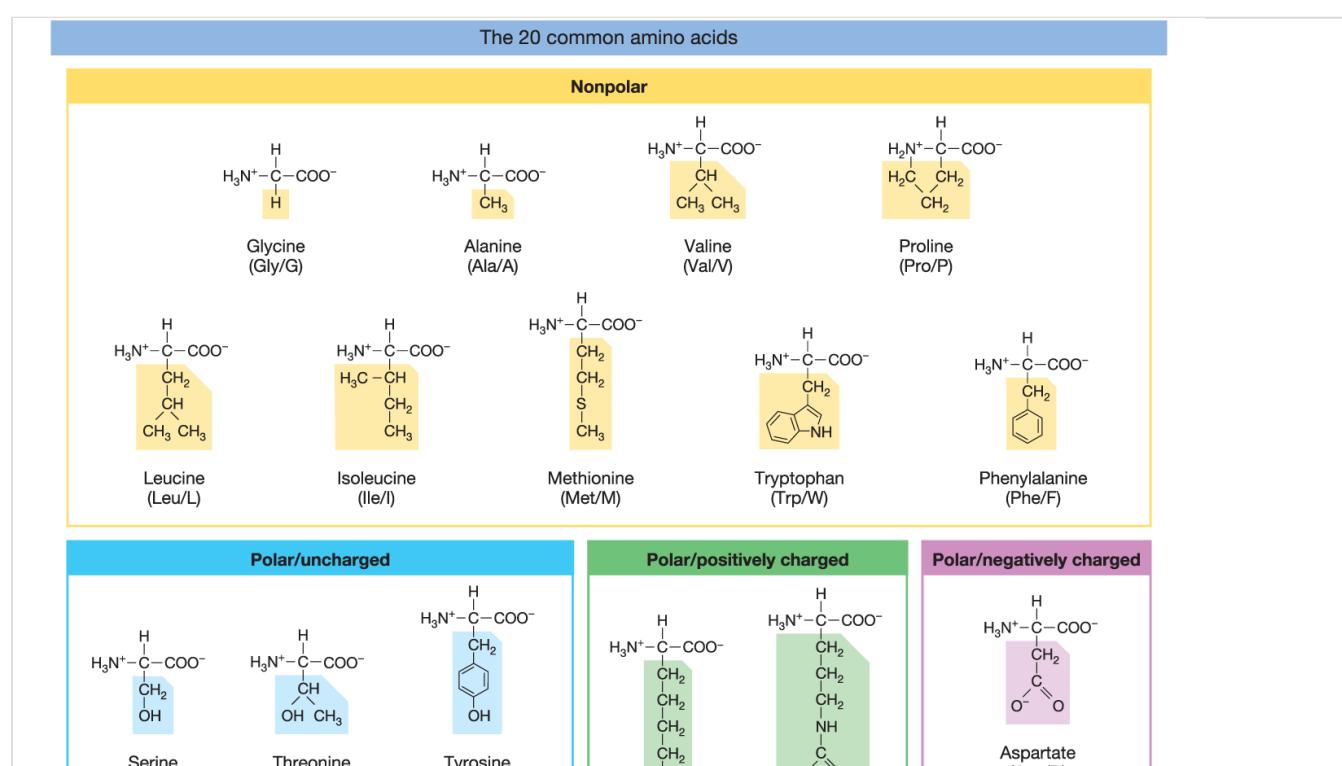
9.1 PROTEIN STRUCTURE

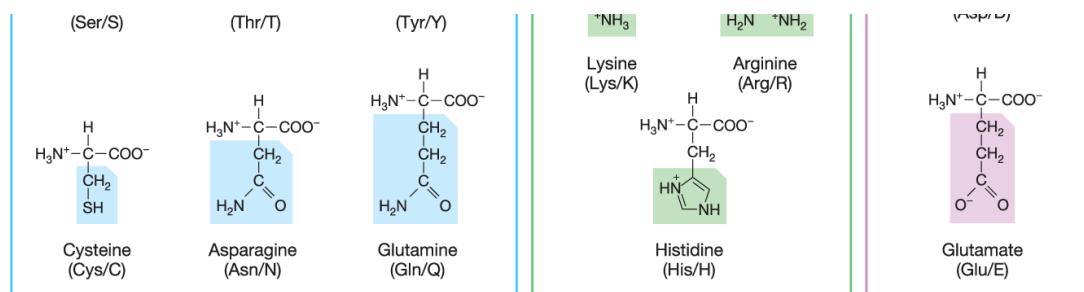
LO 9.1 Explain how the interactions of amino acids determine the structure of proteins.

Before considering how proteins are made, let's start with a discussion of the structure of proteins. Proteins are polymers composed of building blocks called **amino acids**. In other words, a protein is a chain of amino acids. Because amino acids were once called peptides, a chain is sometimes referred to as a **polypeptide**. Amino acids have the general formula



All amino acids have two functional groups (an amino group and a carboxyl group, shown above) bonded to the same carbon atom (called the α carbon). Also attached to the α carbon is a hydrogen (H) atom and a **side chain**, known as an **R (reactive) group**. There are 20 common amino acids that can make up proteins, each amino acid having a different R group that gives it unique properties (**Figure 9-2**). The side chains are categorized into four groups based on their chemical properties: nonpolar, polar/uncharged, polar/positively charged, and polar/negatively charged. As described in this chapter, the chemical properties of side chains play a role in determining the folded structures of proteins.





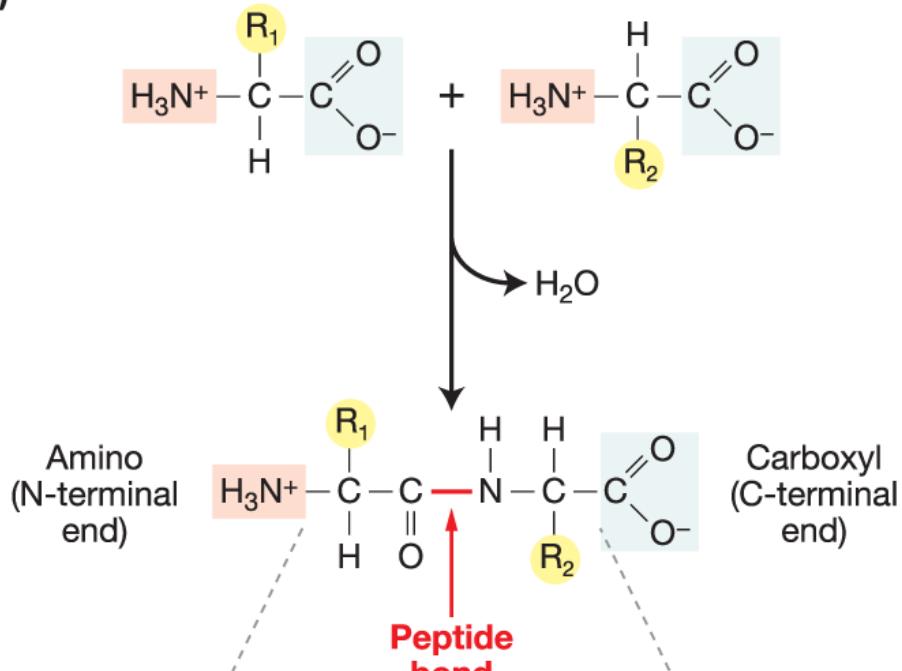
Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 9-2 The chemical structures of the side chains (R groups) of the 20 common amino acids are grouped based on their polarity and charge. Each amino acid is labeled with its full name, abbreviation, and single letter designation such as Glycine, Gly, and G, respectively.

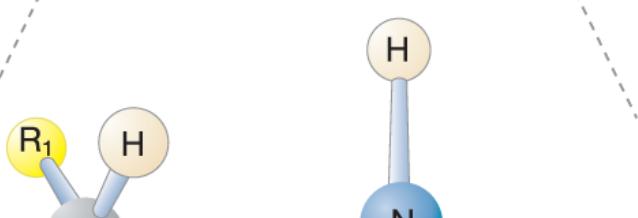
In proteins, amino acids are linked together by covalent bonds called **peptide bonds**. A peptide bond is formed by linkage of the **amino group** (NH_3^+) of one amino acid with the **carboxyl group** (COO^-) of another amino acid (Figure 9-3). One water molecule is removed during the reaction. Because of the way in which peptide bonds form, a polypeptide chain always has an amino end (N-terminal end) and a carboxyl end (C-terminal end).

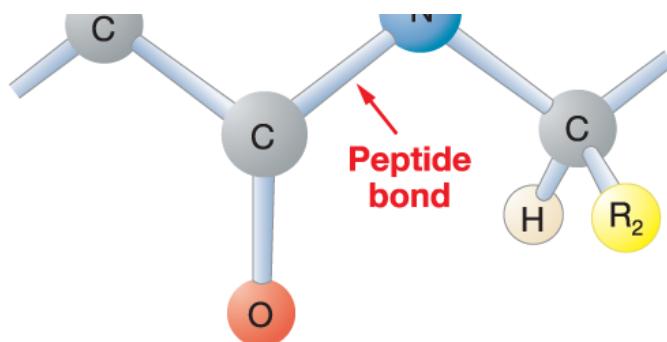
The peptide bond

(a)



(b)





Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

FIGURE 9-3 (a) Peptide bonds form between amino acids by the removal of water. R₁ and R₂ represent R groups (side chains) of amino acids. (b) Peptide bonds are rigid planar units with R groups projecting out from the carbon-nitrogen (C-N) backbone.

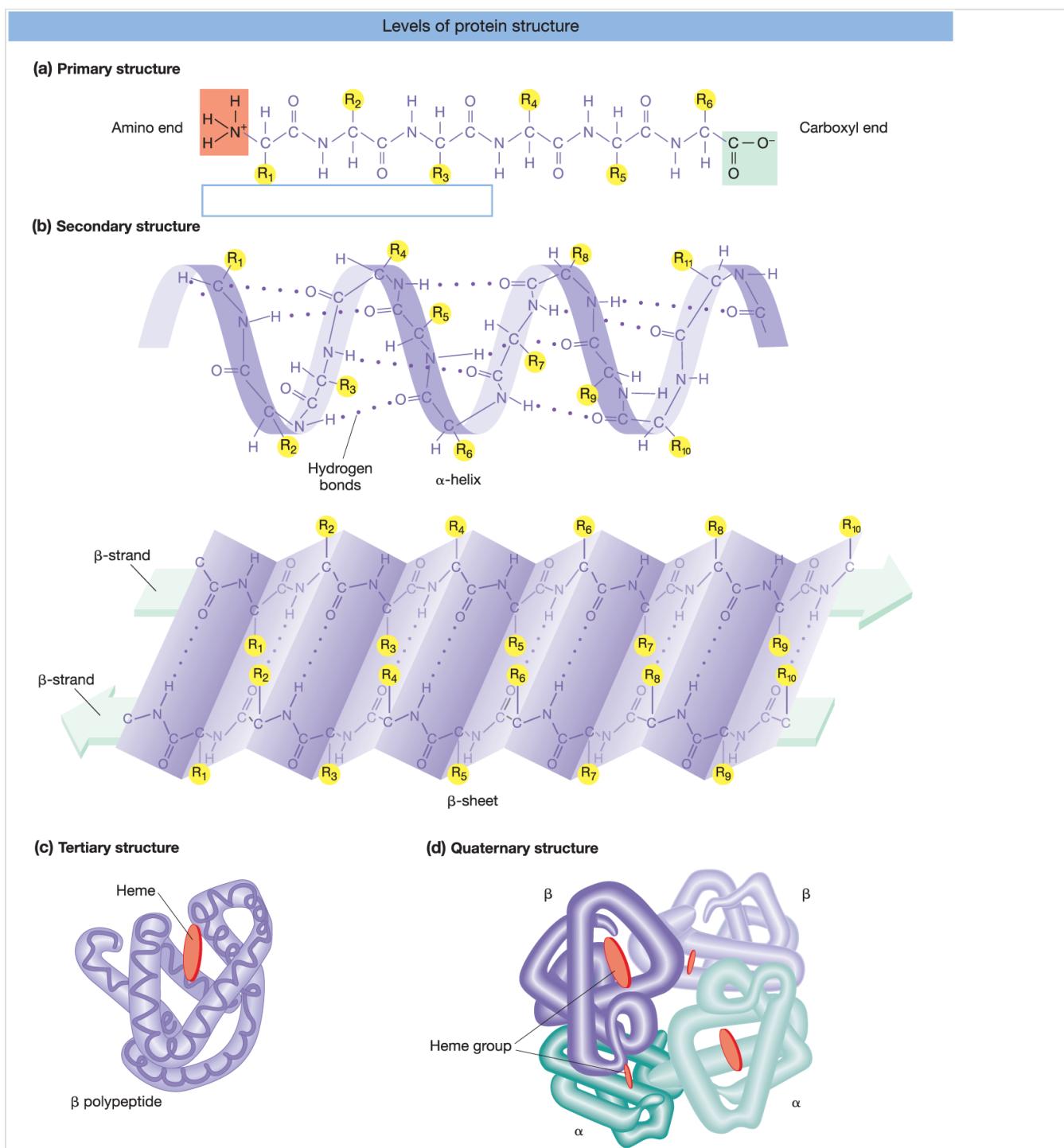
Pal ANIMATED ART Sapling Plus

Peptide bond formation

KEY CONCEPT Each of the 20 common amino acids has an amino group and a carboxyl group as well as a side chain (R group) whose different chemical and physical properties determine the structure and function of proteins. In polypeptides, amino acids are linked together by a peptide bond between the carboxyl group of one amino acid and the amino group of the next.

Protein structures have four levels of organization, illustrated in [Figure 9-4](#). The linear sequence of amino acids in a protein constitutes the [primary structure](#). Local regions of the protein fold into specific shapes, called [secondary structures](#). Each shape arises from bonding forces between amino acids, including several types of non-covalent interactions, notably electrostatic forces such as hydrogen bonds, van der Waals forces (a type of electrostatic interaction involving dipoles), and hydrophobic effects (i.e., the tendency of nonpolar molecules to gather together to exclude water molecules). The most common secondary structures are the α -helix and the β -sheet ([Figure 9-4b](#)). Proteins can contain neither, one, or both of these structures. There are 3.6 amino acids per turn in an α -helix, which means that each amino acid occupies 100 degrees of rotation (360 degrees/3.6). The α -helix structure is stabilized by hydrogen bonds between carbonyl oxygen atoms (C=O) and amide groups (NH) four amino acids away. β -sheets consist of pairs of β -strands (stretches of 3–10 amino acids in an extended conformation) lying side by side that are kept together by inter-strand hydrogen bonds, again between carbonyl oxygen atoms and amide groups. Antiparallel β -strands are oriented (N-terminus to C-terminus) in opposite directions, as in

Figure 9-4b, and parallel β -strands are oriented in the same direction. Lastly, turns composed of a few amino acids and loops of longer stretches of amino acids often connect α -helices to α -helices, β -strands to β -strands, and α -helices to β -strands.



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 9-4 Proteins have four levels of structure. (a) Primary structure: the sequence of amino acids. (b) Secondary structure: the three-dimensional form of regions of a protein. Examples are α -helices and β -sheets. Antiparallel β -sheets have two polypeptide segments arranged in opposite polarity, as indicated by the arrows. (c) Tertiary structure: the overall three-dimensional shape of a polypeptide. In hemoglobin, heme is a non-protein molecule. (d) Quaternary structure: the arrangement of polypeptides in a protein complex. As an example, hemoglobin is composed of four subunits, two α subunits and two β subunits.

Tertiary structures are the overall three-dimensional shape of an entire polypeptide. In addition to noncovalent interactions, tertiary structures can be stabilized by covalent disulfide bridges between cysteine side chains. Cysteine is the only amino acid whose side chain can form a covalent bond. Each enzyme has a pocket called the **active site** into which its substrate or substrates fit. The active sites of enzymes are good illustrations of the precise interactions of side chains. Within the active site, side chains of certain amino acids are strategically positioned to interact with a substrate and catalyze a specific chemical reaction.

Most proteins have a hydrophobic core containing nonpolar amino acid side chains. In contrast, the surface of proteins, which is exposed to the aqueous environment, is made up of polar amino acids, including those that are positively or negatively charged. The surface location of polar amino acids such as serine, threonine, tyrosine, lysine, and arginine makes them accessible to post-translation modification by enzymes, a topic discussed later in this chapter. The nonpolar amino acid proline is unique among all amino acids because it incorporates the amino group into the side chain (see [Figure 9-2](#)). Prolines are infrequently found in the middle of α -helices and β -sheets because they are unable to contribute to the hydrogen bonding pattern of the helices. Instead, prolines are often found in turns and loops, at the ends of α -helices, and in the edge strands of β -sheets.

The folding of polypeptides into their correct conformation will be discussed at the end of this chapter. At present, the rules by which primary structure is converted into secondary and tertiary structures are imperfectly understood. However, from knowledge of the primary amino acid sequence of a polypeptide, the functions of specific regions can be predicted. For example, some characteristic polypeptide sequences are the contact points with membrane phospholipids that position a protein in a membrane. Other characteristic sequences act to bind DNA or RNA. Amino acid sequences or protein folds that are associated with particular functions are called **domains**. A polypeptide may contain one or more separate domains.

Lastly, **quaternary structure** refers to how polypeptides interact with one another to form a multi-polypeptide protein complex. Individual polypeptides in complexes are called **subunits** and are joined together by weak bonds. Quaternary associations can be between different types of polypeptides (resulting in a heterodimer, if there are two subunits) or between identical polypeptides (making a homodimer). Hemoglobin is an example of a heterotetramer (tetramer meaning four subunits), composed of two copies each of two different polypeptides ([Figure 9-4d](#)).

There are two general types of proteins, globular and fibrous. **Globular proteins** have a compact, round shape and play functional roles. Enzymes, hemoglobin, and antibodies are examples of globular proteins. In contrast, **fibrous proteins** have a long, narrow shape and play structural roles. Collagen and keratin are examples of fibrous proteins. Collagen is the main structural protein found in connective tissues such as skin, and keratin is involved in the structure of hair and fingernails.

KEY CONCEPT Proteins have four levels of structure. Primary structure is the sequence of amino acids. Secondary structure is the shape of a region of amino acids, such as α -helices or β -sheets. Tertiary structure is the three-dimensional shape of a whole polypeptide, and quaternary structure is the assembly of multiple polypeptides into a protein complex.

9.2 THE GENETIC CODE

LO 9.2 Outline the experimental evidence supporting the rules of the genetic code.

LO 9.3 Describe features of the genetic code that minimize effects of point mutations on protein function.

The one-gene—one-polypeptide hypothesis of Beadle and Tatum ([Chapter 5](#)) was the source of the first exciting insight into the functions of genes: genes were somehow responsible for the function of enzymes, and each gene apparently controlled one enzyme. This hypothesis became one of the great unifying principles in biology because it provided a bridge between the concepts and research techniques of genetics and biochemistry. When the structure of DNA was deduced in 1953, it seemed likely that there was a linear correspondence between the nucleotide sequence in DNA and the amino acid sequence in a protein. It was soon deduced that the nucleotide sequence in mRNA going from 5' to 3' corresponds to the amino acid sequence in protein going from N-terminus to C-terminus.

If genes are segments of DNA, and if a strand of DNA is just a string of nucleotides, the sequence of nucleotides must somehow dictate the sequence of amino acids in proteins. How does DNA sequence dictate protein sequence? Simple logic tells us that, if nucleotides are the “letters” in a code, a combination of letters can form “words” representing different amino acids. However, in the 1960s, researchers were faced with many questions about how the code is read. How many letters make up a word, or [codon](#), in the code? Are codons overlapping or nonoverlapping? Is the code continuous or discontinuous? Which codon or codons represent each amino acid? The cracking of the [genetic code](#) is the story told in this section.

A degenerate three-letter genetic code specifies the 20 amino acids

If an mRNA is read from one end to the other, only one of four different bases, A, C, G, or U, is found at each position. Thus, if the words encoding amino acids were one letter long, only four words would be possible. This vocabulary cannot be the genetic code because there must be a word for each of the 20 amino acids commonly found in proteins. If the words were two letters long, $4 \times 4 = 16$ words would be possible; for example, AU, CU, or CC. This vocabulary is still not large enough. But if the words were three letters long, $4 \times 4 \times 4 = 64$ words would be

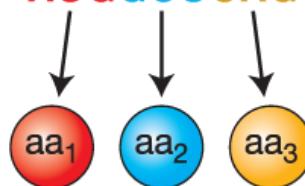
possible; for example, AUU, GCG, or UGC. This vocabulary provides more than enough words to describe the 20 amino acids. Therefore, codons must consist of at least three nucleotides. However, if all three-nucleotide combinations specify an amino acid, the genetic code must be **degenerate**, meaning that some amino acids are specified by two or more different triplets.

The genetic code is nonoverlapping and continuous

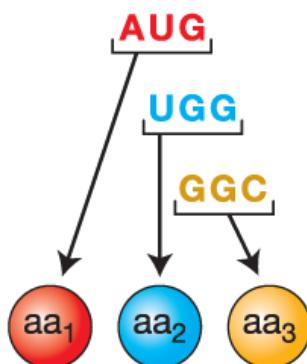
The genetic code could either be overlapping or nonoverlapping. **Figure 9-5** illustrates these possibilities for a three-nucleotide, or **triplet**, code. For a nonoverlapping code, consecutive amino acids are specified by consecutive code words (codons), and a single nucleotide mutation would only alter one codon and one amino acid. For an overlapping code, consecutive amino acids are specified by codons that have nucleotides in common; for example, the third nucleotide in one codon could be the second or first nucleotide in adjacent codons. In this case, a single nucleotide mutation would alter three codons and three amino acids. By 1961, it was already clear that the genetic code was nonoverlapping. Analyses of mutant proteins showed that almost all of the time, only one amino acid changed, which is predicted by a nonoverlapping code.

Nonoverlapping versus overlapping genetic codes

Nonoverlapping code **AUGGCUUCAGCUUGAC**



Overlapping code **AUGGCUUCAGCUUGAC**



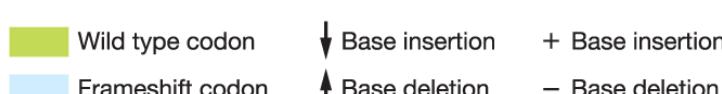
Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 9-5 Nonoverlapping and overlapping genetic codes would translate into different amino acid sequences. The example uses codons with three nucleotides (a triplet code). (Top) In a nonoverlapping code, a protein is translated by reading nucleotides sequentially in sets of three. Each nucleotide is found in only one codon. In this example, the third nucleotide (G) in the RNA is only in the first codon. (Bottom) In contrast, in an overlapping code, each nucleotide occupies positions in multiple codons. In this case, the third nucleotide (G) is found in all three codons.

The genetic code could be continuous or discontinuous. In a continuous code, codons are arranged side by side with no gaps, whereas in a discontinuous code, codons are separated by one or more nucleotides that act to pause translation and restart anew at the next codon. If the code was continuous, which turns out to be the case, insertion or deletion of a single nucleotide would cause a shift in the reading frame starting at the site of the mutation and continuing to the end of the open reading frame (**Figures 9-6b** and **c**). In contrast, if the code was discontinuous, insertion or deletion of a single nucleotide would only affect one codon, and this error would not be propagated through the rest of the open reading frame.

Evidence that the genetic code is continuous and read in triplets

| | | Mutation | Phenotype |
|-----|----------------------|----------|------------------|
| (a) | Wild type | None | rII ⁺ |
| (b) | FCO mutant | + | rII ⁻ |
| (c) | FCO mutant | - | rII ⁻ |
| (d) | Suppression of FCO | + - | rII ⁺ |
| (e) | Two base insertion | + + | rII ⁻ |
| (f) | Two base deletion | -- | rII ⁻ |
| (g) | Three base insertion | + + + | rII ⁺ |
| (h) | Three base deletion | -- - | rII ⁺ |


 Wild type codon ↓ Base insertion + Base insertion
 Frameshift codon ↑ Base deletion - Base deletion

Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 9-6 Examination of single, double, and triple mutations by Crick and colleagues demonstrated that the genetic code is continuous and read in triplets. A wild-type sequence (a) produces the wild-type rII^+ phenotype. A single base insertion (b) or deletion (c) causes the rII^- phenotype. The phenotype is suppressed by combining these mutations (d), but not by a second insertion (e) or deletion (f). Suppression restores the normal reading frame and indicates that the genetic code is continuous. Furthermore, a double insertion or deletion is suppressed by a third insertion (g) or deletion (h), indicating that the genetic code is read in triplets.

Pal Convincing proof that the genetic code is continuous came from genetic experiments first reported in 1961 by Francis Crick, Sidney Brenner, and their co-workers. These experiments used mutants in the *rII* locus of T4 phage. The use of *rII* mutations in recombination analyses was discussed in [Chapter 6](#). Phage T4 is usually able to grow on two different *E. coli* strains, called B and K. However, mutations in the *rII* gene change the host range of the phage: mutant phages can still grow on an *E. coli* B host, but they cannot grow on an *E. coli* K host. Mutations causing the rII^- phenotype were induced using a chemical called proflavin, which causes insertion or deletion of single base pairs in DNA. Starting with one particular proflavin-induced mutation called FCO, Crick and his colleagues found “reversions” (reversals of the mutation) that were able to grow on *E. coli* strain K. Genetic analysis of these phages revealed that the “revertants” were not identical to true wild types. In fact, revertions were found to be due to the presence of a second mutation at a different site from that of FCO, although in the same gene. This second mutation “suppressed” mutant expression of the original FCO. Recall from [Chapter 5](#) that a suppressor mutation counteracts the effects of another mutation so that the bacterium is more like wild type.

How can these results be explained? If we assume that the gene is read from one end only, the original insertion or deletion induced by proflavin could interrupt the normal reading mechanism that establishes the grouping of bases to be read as words. For example, if each group of three bases in an mRNA makes a word, the [reading frame](#) might be established by taking the first three bases from the end as the first word, the next three as the second word, and so forth ([Figure 9-6a](#)). In that case, a proflavin-induced insertion of a single base pair in the DNA would shift the reading frame on the mRNA from that point on, causing all following words to be misread ([Figure 9-6b](#)). Such [frameshift mutations](#) could reduce most of the genetic message to garbage. However, the proper reading frame could be restored by a compensatory deletion somewhere else, limiting the garbage to the segment between the two mutations ([Figure 9-6d](#)). We have assumed here that the original frameshift mutation was an insertion, but the explanation works just as well if the original FCO mutation was a deletion and the suppressor was an insertion ([Figures 9-6c](#) and [d](#)). The few wrong words in the suppressed genotype could account for the fact that the revertants (suppressed

phenotypes) did not look exactly like true wild types. These data demonstrated that the genetic code is continuous.

Crick and his colleagues also found that insertions or deletions of two bases produced the mutant phenotype ([Figures 9-6e](#) and [f](#)); however, a third mutation of the same type restored a wild-type phenotype because it corrected the reading frame ([Figures 9-6g](#) and [h](#)). This observation provided the first experimental evidence that a word in the genetic code consists of three successive nucleotides, or a triplet.

KEY CONCEPT The genetic code has the following features:

1. The linear sequence of nucleotides in a gene determines the linear sequence of amino acids in the encoded protein.
 2. A codon of three nucleotides specifies an amino acid.
 3. The genetic code is degenerate; more than one codon can specify the same amino acid.
 4. The genetic code is nonoverlapping; each nucleotide is part of only one codon.
 5. The genetic code is continuous; it is read from a fixed starting point and continues uninterrupted to the end of the open reading frame.
-

Cracking the code

Deciphering the genetic code—determining the amino acid specified by each triplet—is one of the most exciting genetic breakthroughs to occur since elucidation of the structure of DNA. After the necessary experimental techniques became available, the genetic code was cracked quickly.

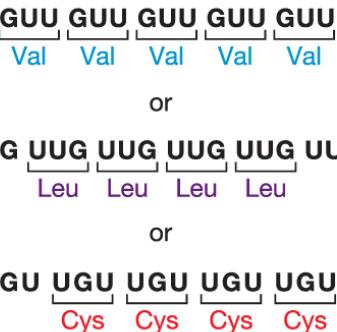
One technical breakthrough was the discovery that single-stranded RNA can be synthesized *in vitro* by the enzyme polynucleotide phosphorylase. Unlike transcription, no DNA template is needed for this synthesis, so the nucleotides are incorporated at random. The ability to enzymatically synthesize RNA offered the exciting prospect of creating specific RNA sequences and then seeing which amino acids they would specify. The first synthetic RNA was made using only uracil nucleotides, producing ... UUUU ... [poly(U)]. In 1961, Marshall Nirenberg and Heinrich Matthaei mixed poly(U) with the protein-synthesizing machinery of *E. coli* *in vitro* and observed the formation of a protein. The main excitement centered on the question of the amino acid sequence of this protein. It proved to be polyphenylalanine—a string of phenylalanine (Phe) amino acids ([Figure 9-7a](#)). Thus, the triplet UUU codes for phenylalanine. Expanding this approach to other single nucleotides and combinations of nucleotides led to assignment of about

40 codons to particular amino acids ([Figures 9-7b](#) and [c](#)). The technical breakthrough that led to assignment of 61 of the 64 codons was the development by H. Gobind Khorana of methods to chemically synthesize RNAs of defined sequences. In 1968, Nirenberg and Khorana were awarded the Nobel Prize for deciphering the genetic code.

Cracking the genetic code using synthetic RNAs

(a)  Reading frames 1, 2, and 3
Phe Phe Phe Phe Phe

(b)  Reading frames 1, 2, and 3
Pro Pro Pro Pro Pro

(c)  Reading frame 1
Val Val Val Val Val
or
Reading frame 2
Leu Leu Leu Leu
or
Reading frame 3
Cys Cys Cys Cys

Griffiths et al., *Introduction to Genetic Analysis*, 12e, ©
2020 W. H. Freeman and Company

FIGURE 9-7 The genetic code (shown in [Figure 9-8](#)) was established by discovering the amino acids coded by synthetic RNAs containing repeats of a single nucleotide (a and b) or combinations of nucleotides (c). Regardless of the reading frame, repeats of single nucleotides code for the same amino acid, whereas repeats of different nucleotides can code for three different amino acids in the three different reading frames.

Virtually all organisms use the same genetic code ([Figure 9-8](#)). There are just a few exceptions in which a small number of codons have different meanings—for example, in mitochondrial and nuclear genomes of ciliates and some other protozoans. How is it that genomes can have very different base compositions but use the same code? Part of the answer is that degeneracy of the code permits the DNA base composition of genomes to vary over a wide range and still encode all 20 amino acids. For example, the G + C content of the coding regions of bacterial genomes ranges from about 20 percent to 70 percent but encodes proteins with very similar sequences. Furthermore, not all codons for an amino acid, termed **synonymous codons**, are used with equal frequency in an organism, and this frequency can change dramatically among organisms. The topic of synonymous codon usage bias is discussed in greater detail in [Chapter 14](#). From a practical standpoint, the almost universality of the genetic code means that mRNAs from one

organism can be correctly translated in another organism. For example, bacteria can be used to make large amounts of human proteins by translating human mRNA sequences.

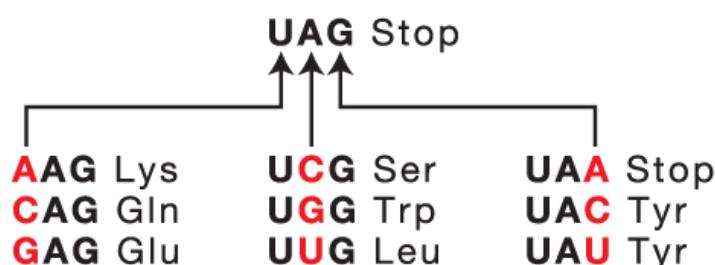
| The genetic code | | | | | | |
|---|--|-----------------------|--|---|-------------|-------------|
| | Second letter | | | | | |
| | U | C | A | G | | |
| U | UUU } Phe UUC UUA } Leu UUG } | UCU } UCC UCA } | UAU } Tyr UAC UAA Stop UAG Stop | UGU } Cys UGC UGA Stop UGG Trp | U C A | G |
| | CUU } CUC CUA } Leu CUG } | CCU } CCC CCA } | CAU } His CAC CAA } Gln CAG } | CGU } CGC CGA } | Arg | U C A |
| | AUU } AUC AUA } Ile AUG Met | ACU } ACC ACA } | AAU } Asn AAC AAA } Lys AAG } | AGU } Ser AGC AGA } | Ser | U C A |
| | GUU } Val GUC GUA } | GCU } GCC GCA } | GAU } Asp GAC GAA } | GGU } GGC GGA } | Gly | G |
| First letter | | | | | | |
| Third letter | | | | | | |
| FIGURE 9-8 The genetic code designates the amino acids specified by each codon. Sixty-one codons specify amino acids, and three codons specify translation stop. This code is almost universal among all organisms. | | | | | | |

Stop codons

Three of the 64 codons are **stop codons**, or **termination codons**, that, instead of encoding an amino acid, stop translation. They can be regarded as being similar to periods punctuating the message encoded in the DNA. One of the first indications of the existence of stop codons came in 1965 from Sydney Brenner's work with the T4 phage. Brenner analyzed certain mutations ($m_1 - m_6$) in a single gene that controls the head protein of the phage. He found that the head protein of

each mutant was shorter than that of the wild type. Brenner examined the ends of the shortened proteins and compared them with the wild-type protein. For each mutant, he recorded the next amino acid that would have been inserted to continue the wild-type chain. The amino acids for the six mutations were glutamine, lysine, glutamic acid, tyrosine, tryptophan, and serine. These results presented no immediately obvious pattern, but Brenner deduced that each of these codons can mutate to the codon UAG by a single nucleotide change ([Figure 9-9](#)). He therefore postulated that UAG is a stop (termination) codon—a signal to the translation mechanism that the protein is now complete. The two other stop codons are UGA and UAA. Stop codons are often called **nonsense codons** because they do not designate an amino acid.

Single nucleotide mutations that produce a UAG stop codon



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

FIGURE 9-9 The UAG stop codon is produced by a single nucleotide mutation in any of nine codons, eight that code for amino acids and one that codes for another stop codon. Note that a single nucleotide mutation is not sufficient to convert some codons to any of the three stop codons.



Degeneracy of the genetic code limits the effects of point mutations

The genetic code is not random, and degeneracy of the code minimizes the potentially detrimental effects of point mutations (single nucleotide mutations). The genetic code can be thought of as 16 blocks, each containing four entries defined by the first two nucleotides in a codon ([Figure 9-10](#)). Eight of the 16 blocks specify only one amino acid, meaning that mutations in the third position of a codon do not change the encoded amino acid. As an example, GUA, GUC, GUG, and GUU all code for valine. Furthermore, mutations in the first position of a codon often lead to a conservative change in amino acid—that is, an amino acid with similar chemical properties and size. For example, mutation of the first nucleotide of the leucine codon CUU to AUU, GUU, or UUU changes the amino acid to isoleucine, valine, or phenylalanine, respectively, all of which are

nonpolar and have moderately sized side chains (see [Figure 9-2](#)). In these cases, the structure and function of a mutant protein is likely to remain largely unaffected.

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

Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 9-10 Amino acids that are similar in polarity and charge are represented by codons that are similar in sequence and, thus, located near one another in the table. Amino acids are color-coded based on their polarity and charge, as in [Figure 9-2](#) (yellow: nonpolar, blue: polar/uncharged, green: polar/positively charged, and purple: polar/negatively charged). Stop codons are white.

FO Given that there are more than 10^{84} possible ways to create a three-letter code with each of the 20 amino acids and three stop codons assigned to at least one codon, it remains to be determined how a single code that is almost perfectly conserved among organisms originated and evolved. In the face of these questions, it is interesting to note that the odds of this specific code being randomly selected from among 10^{84} possibilities is astronomically less likely than finding a particular grain of sand from among the roughly 10^{19} grains on all of the earth's beaches and deserts.

KEY CONCEPT The genetic code contains 64 codons (61 codons specify amino acids and 3 codons specify translation stop). The genetic code is nearly universal, and the arrangement of codons in the codon table is highly non-random.

9.3 tRNAs AND RIBOSOMES

LO 9.4 Explain how the structures of tRNAs and ribosomes determine how they function in protein synthesis.

Once the genetic code was deciphered, scientists began to wonder how the sequence of amino acids of a protein was determined by the triplet codons of an mRNA. An early model, quickly dismissed as naive and unlikely, proposed that mRNA codons could fold up and form 20 distinct cavities that directly bind specific amino acids in the correct order. Instead, in 1958, Crick recognized the following:

It is therefore a natural hypothesis that the amino acid is carried to the template by an adaptor molecule, and that the adaptor is the part which actually fits on to the RNA. In its simplest form [this hypothesis] would require twenty adaptors, one for each amino acid.¹

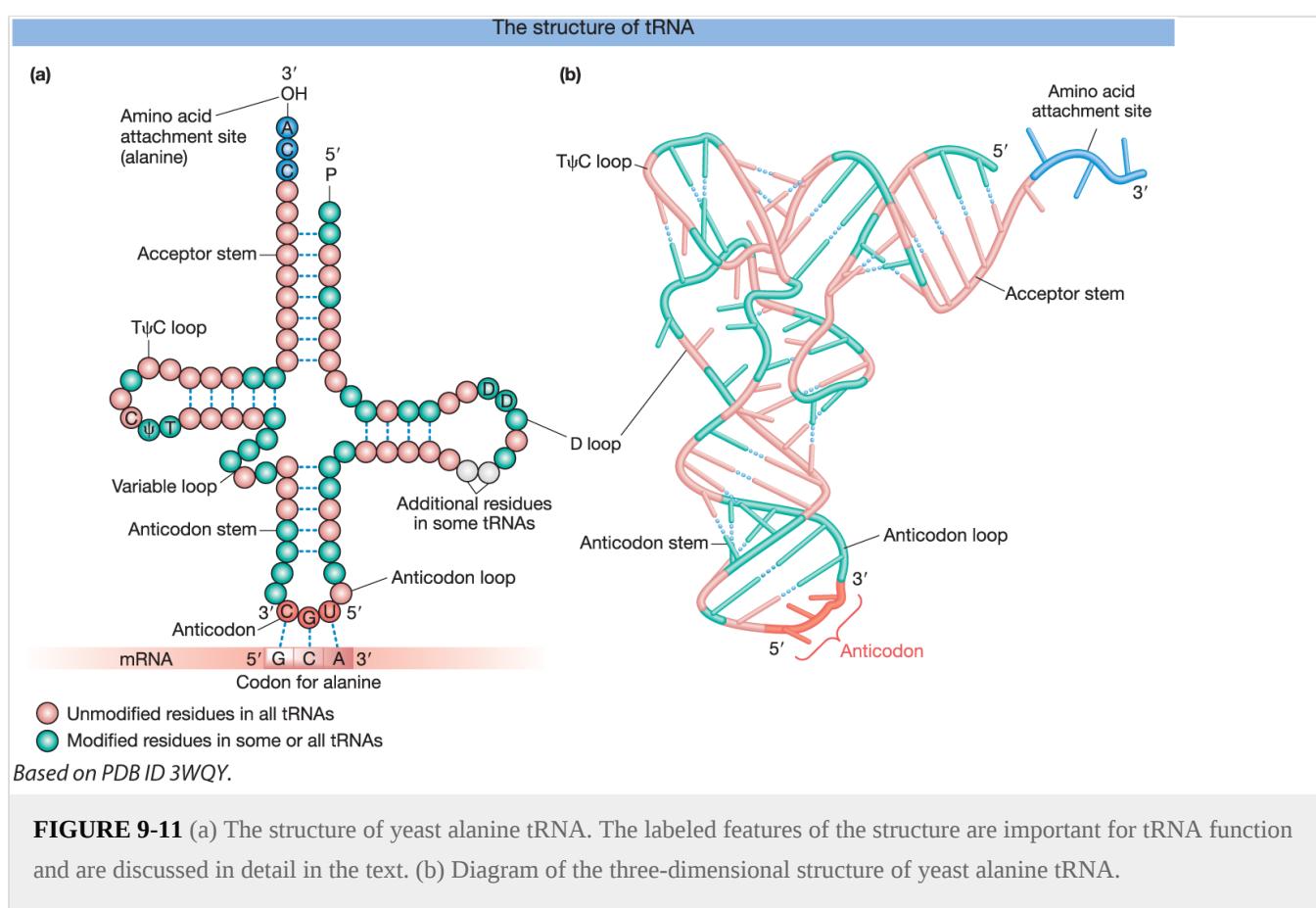
He speculated that the adaptor “might contain nucleotides. This would enable them to join on the RNA template by the same ‘pairing’ of bases as is found in DNA.” Furthermore, “a separate enzyme would be required to join each adaptor to its own amino acid.” We now know that Crick’s “adaptor hypothesis” is largely correct. Transfer RNAs (tRNA) are the adaptors that base pair to codons in mRNA.

Enzymes join amino acids to specific tRNAs, which then bring the amino acid to the ribosome, the molecular complex that will attach the amino acid to a growing polypeptide. In all organisms, ribosomes consist of one **small ribosomal subunit** and one **large ribosomal subunit**, each made up of RNA (called ribosomal RNA, rRNA) and protein. Each subunit is composed of one to three rRNA types and as many as 50 proteins.

tRNAs are adaptors

Structural features of tRNA serve as the components of bridges that link mRNA codons to specific amino acids. Single-stranded tRNAs, which are about 75 nucleotides in length, have a cloverleaf shape consisting of four double-helical stems and three single-stranded loops (**Figure 9-11a**). The middle loop of each tRNA is called the anticodon loop because it carries a nucleotide triplet called an **anticodon**. This sequence is complementary to the mRNA codon for the amino acid carried by the tRNA. The anticodon in tRNA and the codon in the mRNA interact by specific

RNA:RNA base pairing. Again, we see the principle of nucleic acid complementarity at work, this time in the binding of two separate RNAs. Because codons in mRNA are read in the 5'-to-3' direction, anticodons are oriented and written in the 3'-to-5' direction ([Figure 9-11a](#)).



MIn addition to the anticodon, tRNAs have other structural features that are important for their function.

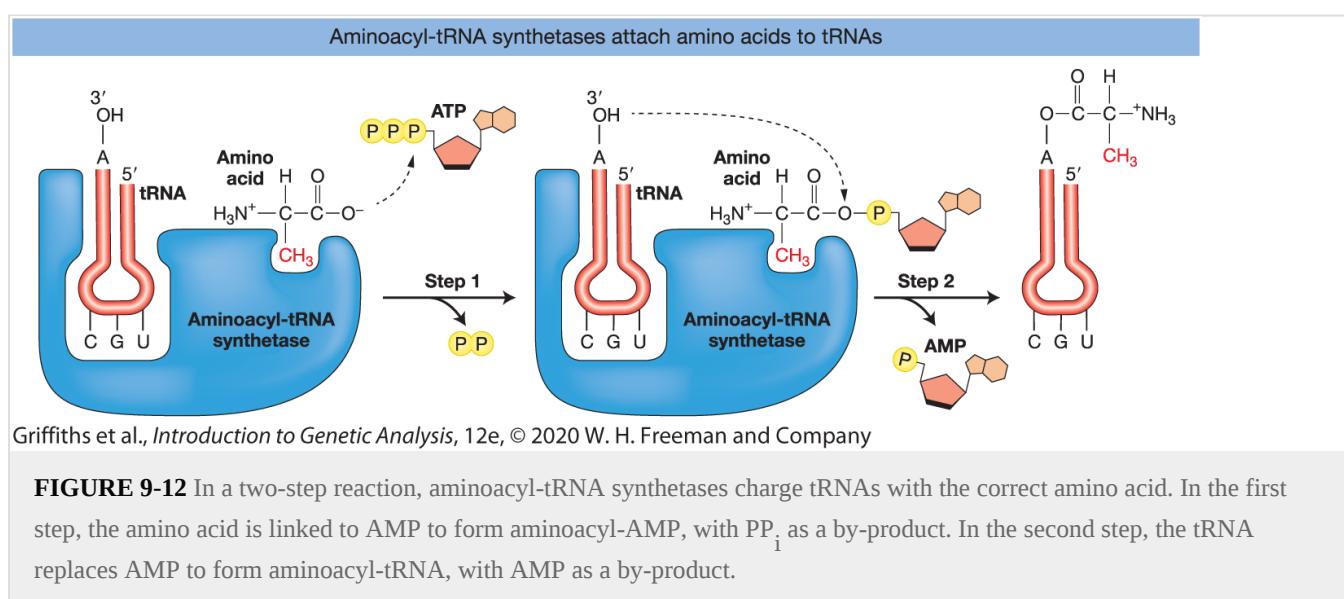
1. The sequence 5'-CCA-3' is found at the very 3' end of all tRNAs. In bacteria, the CCA is generally encoded in tRNA genes, but in eukaryotes, the CCA is added post-transcriptionally by an enzyme called nucleotidyltransferase, or CCA-adding enzyme. The CCA extends beyond the base paired acceptor stem, and the 3'-OH of the A is the site of attachment of the amino acid.
2. Numerous nucleotides in all tRNAs undergo post-transcriptional modification; that is, tRNAs are transcribed with the standard four nucleotides, which are then altered by enzymes. In yeast, each tRNA contains between 7 and 17 modified nucleotides. For example, the D loop contains the nucleotide dihydrouridine (D), which has two extra hydrogen atoms on the uracil base; the T ψ C loop contains pseudouridine (ψ), which has uracil attached to the ribose sugar at a carbon rather than a nitrogen; and the anticodon can contain the base inosine (I), which is structurally similar to guanosine (G). Nucleotide

modifications in the anticodon affect base pairing with the codon (see the discussion about wobble in the next section), and nucleotide modifications at other sites affect tRNA recognition, folding, and stability. However, the precise function of many modifications is yet to be determined.

3. A tRNA normally exists as an inverted L-shaped three-dimensional structure, as shown in [Figure 9-11b](#), rather than the flattened cloverleaf shown in [Figure 9-11a](#). Although tRNAs differ in their primary nucleotide sequence, all tRNAs fold into virtually the same L-shaped conformation, indicating that the shape of a tRNA is important for its function.

KEY CONCEPT tRNAs have four important structural features: (1) the sequence CCA at the 3' end; (2) modified nucleotides such as dihydrouridine, pseudouridine, and inosine; (3) an overall inverted L shape; and (4) an anticodon.

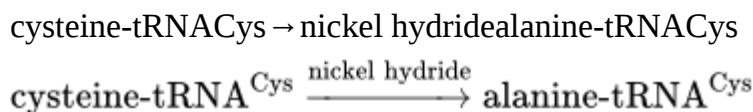
Amino acids are attached to tRNAs by enzymes called [aminoacyl-tRNA synthetases](#). The tRNA with an attached amino acid is said to be [charged tRNA](#). There are 20 synthetases, one for each of the 20 amino acids. Because the code is degenerate, some synthetases act on multiple tRNAs. Charging by aminoacyl-tRNA synthetases occurs in two steps ([Figure 9-12](#)). In the first step, the carboxyl group of the amino acid reacts with the α -phosphate of ATP to form 5'-aminoacyl-AMP and release pyrophosphate (PP_i). 5'-aminoacyl-AMP is referred to as an activated amino acid. In the second step of charging, the amino acid is transferred to the adenine (A) of the invariant CCA sequence at the 3' end of the tRNA, and AMP is released as a by-product.



ANIMATED ART  Sapling Plus

tRNA charging

What would happen if the wrong amino acid was covalently attached to a tRNA? A convincing experiment answered this question. The experiment used tRNA^{Cys}, the tRNA specific for cysteine. This tRNA was charged with cysteine, meaning that cysteine was attached to the tRNA. The charged tRNA was treated with nickel hydride, which converted the cysteine (while still bound to tRNA^{Cys}) into another amino acid, alanine, without affecting the tRNA:



Proteins synthesized with alanine-tRNA^{Cys} had alanine wherever cysteine was coded. The experiment demonstrated that amino acids are “illiterate”; they are inserted at the proper position because the tRNA adaptors recognize the mRNA codons. Thus, attachment of the correct amino acid to its corresponding tRNA (also called its cognate tRNA) by an aminoacyl-tRNA synthetase is the critical step in ensuring that the mRNA code is translated correctly. If the wrong amino acid is attached, there is no way to prevent it from being incorporated into a growing polypeptide chain.

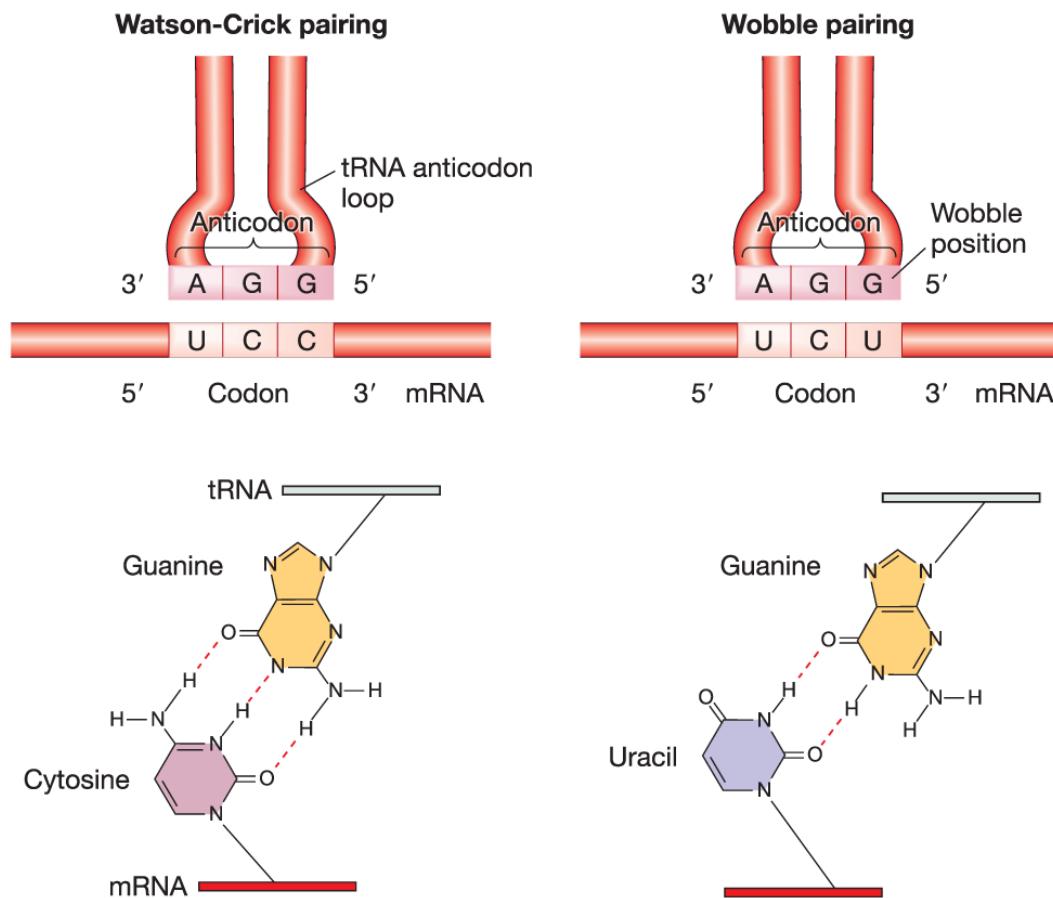
Correct charging of tRNAs depends on selection of appropriately paired tRNAs and amino acids by aminoacyl-tRNA synthetases. These enzymes are good at recognizing the correct tRNA because tRNAs have numerous distinguishing structural features, including nucleotide sequence and nucleotide modifications. However, the only distinguishing feature of amino acids is their side chain, which can be very similar. Therefore, to prevent mistakes, aminoacyl-tRNA synthetases have a two-step mechanism that discriminates between chemically similar amino acids such as valine (Val) and isoleucine (Ile), which differ by only a single CH₂ group (see [Figure 9-2](#)). The first discrimination step occurs in the activation site of the enzyme where the amino acid is bound and activated to form aminoacyl-AMP. This step rejects amino acids that do not fit into the activation site because they are too large. So, Val-tRNA synthetase will reject Ile because it is too large. In contrast, Ile-tRNA synthetase will sometimes charge tRNA^{Ile} with valine to produce Val-tRNA^{Ile}. But, in the second discrimination step, Val-tRNA^{Ile}, but not Ile-tRNA^{Ile}, fits into a separate active site of the synthetase and is hydrolyzed to valine and tRNA^{Ile}. Because of this proofreading mechanism, the error rate of protein synthesis is very low, in the range of 1 in 10⁴–10⁵ amino acids incorporated.

KEY CONCEPT tRNAs are charged by tRNA synthetases in a two-step reaction requiring ATP. There is a different synthetase for each amino acid. Two proofreading steps ensure that synthetases charge tRNAs with the correct amino acid.

Wobble base pairing allows tRNAs to recognize more than one codon

If perfect Watson-Crick base pairing between tRNA anticodons and mRNA codons was required to recognize all of the codons, there would need to be 61 different tRNAs. However, this is not the case; some tRNAs can recognize multiple codons through a different kind of base pairing at the third position of a codon, termed the **wobble** position. For example, the charged tRNA^{Ser} can form either a normal Watson-Crick G–C base pair or an unusual G–U wobble base pair with serine codons (**Figure 9-13**), so one tRNA can be used for both serine codons. In addition, inosine (I), a rare modified base in tRNA, can base pair to C, U, and A (**Table 9-1**). Therefore, because of wobble base pairing, cells require fewer than 61 tRNAs to read all of the codons.

Wobble base pairing allows a tRNA to recognize two or more codons



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 9-13 The third position in a codon is often called the wobble position because multiple nucleotides in this position can base pair with same 5' nucleotide in a tRNA anticodon. In this case, both C and U in the third position base pair with a G in the anticodon. [Table 9-1](#) lists all of the wobble base pairs. (Top) Normal (Watson-Crick) and wobble base pairing interactions of tRNA anticodons and codons. (Bottom) Hydrogen bonds formed in Watson-Crick and wobble base pairs.

TABLE 9-1 Wobble Base-pairing Rules

In

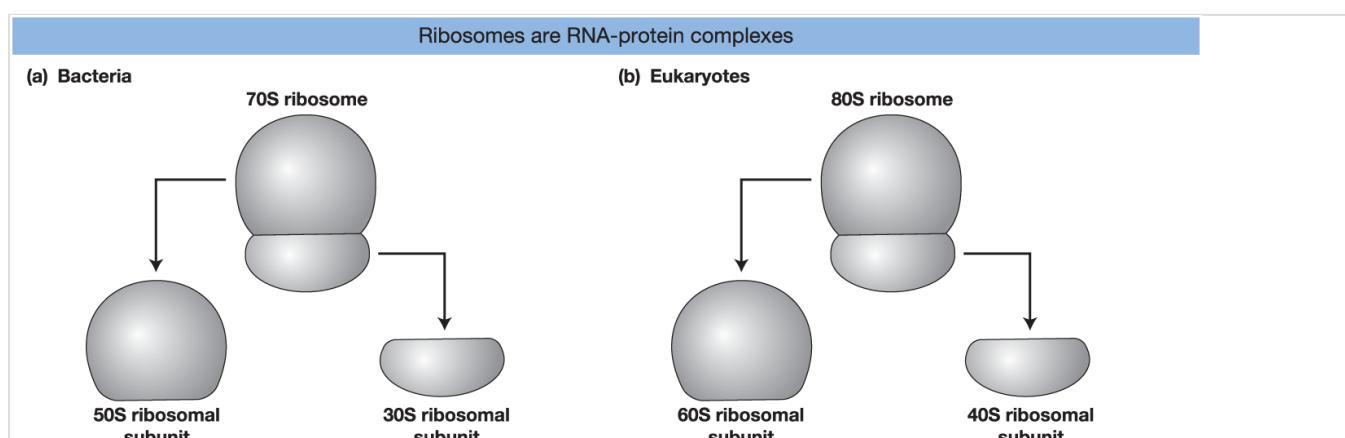
| 5' end of anticodon | 3' end of codon |
|---------------------|-----------------|
| A | U |
| C | G |
| G | C or U |
| U | A or G |
| I (inosine) | A, C, or U |

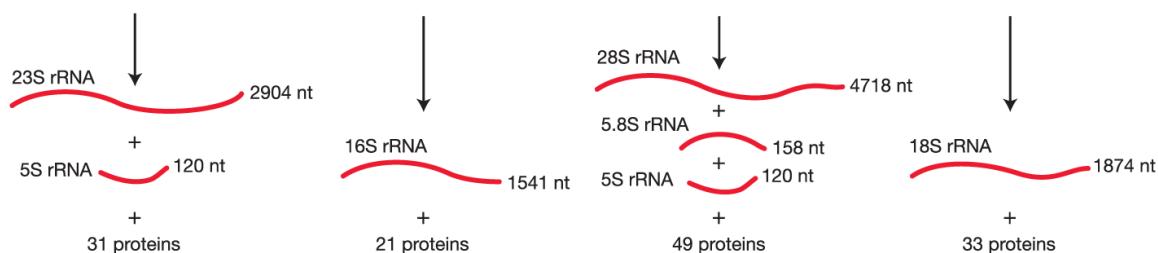
Wobble base pairs are shown in red.

KEY CONCEPT The genetic code is called degenerate because, in many cases, more than one codon is assigned to a single amino acid; in addition, wobble base pairing allows the anticodon of some tRNAs to pair with more than one codon.

Ribosome structure and function

Ribosomes are made up of two subunits that were originally characterized by their rate of sedimentation when spun in an ultracentrifuge. Therefore, their names are derived from their sedimentation coefficients in Svedberg (S) units, which is an indication of molecular size. In bacteria, the small and large subunits are called 30S and 50S, respectively, and they associate to form a 70S particle ([Figure 9-14a](#)). The eukaryotic subunits are called 40S and 60S, and the complete eukaryotic ribosome is called 80S ([Figure 9-14b](#)).





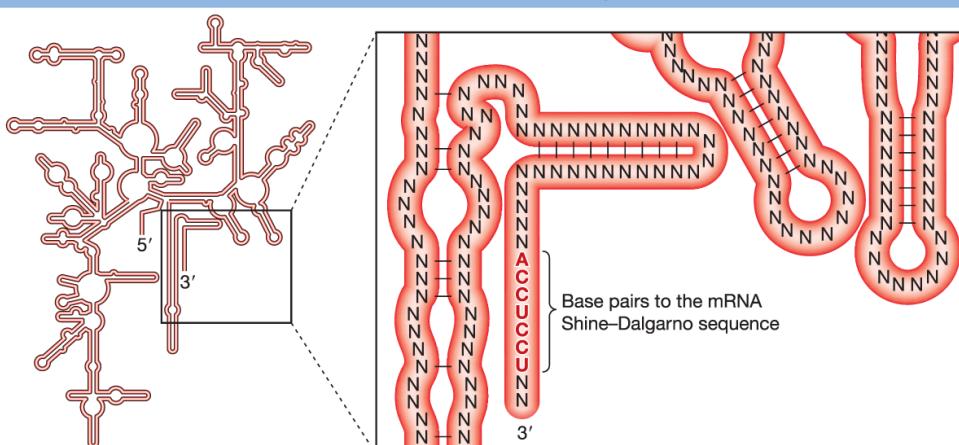
Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 9-14 Ribosomes contain a large and a small subunit. Each subunit contains one large RNA and a set of proteins. In addition, the large subunit of bacterial ribosomes contains one small RNA, 5S rRNA, whereas (b) the large subunit of eukaryotic ribosomes contains two small rRNAs, 5S rRNA and 5.8S rRNA.

Although bacterial and eukaryotic ribosomes differ in size and composition, the steps in protein synthesis are similar overall. The similarities clearly indicate that translation is an ancient process that originated in the common ancestor of bacteria and eukaryotes. On the other hand, because of differences between bacterial and eukaryotic ribosomes, antibiotics are able to inactivate bacterial ribosomes but leave eukaryotic ribosomes untouched. More than half of all antibiotics currently in use target the bacterial ribosome, including penicillin, tetracycline, ampicillin, and chloramphenicol.

When ribosomes were first studied, the fact that almost two-thirds of their mass is RNA and only one-third is protein was surprising. For decades, rRNAs had been assumed to function as a scaffold for assembly of ribosomal proteins. That role seemed logical because rRNAs fold up by intramolecular base pairing into stable secondary structures ([Figure 9-15](#)). According to this model, the ribosomal proteins catalyzed protein synthesis. This view changed with the discovery in the 1980s of catalytic RNAs (see [Chapter 8](#)). As you will see, there is now considerable evidence that rRNAs, assisted by the ribosomal proteins, catalyze protein synthesis.

rRNAs fold into particular secondary structures



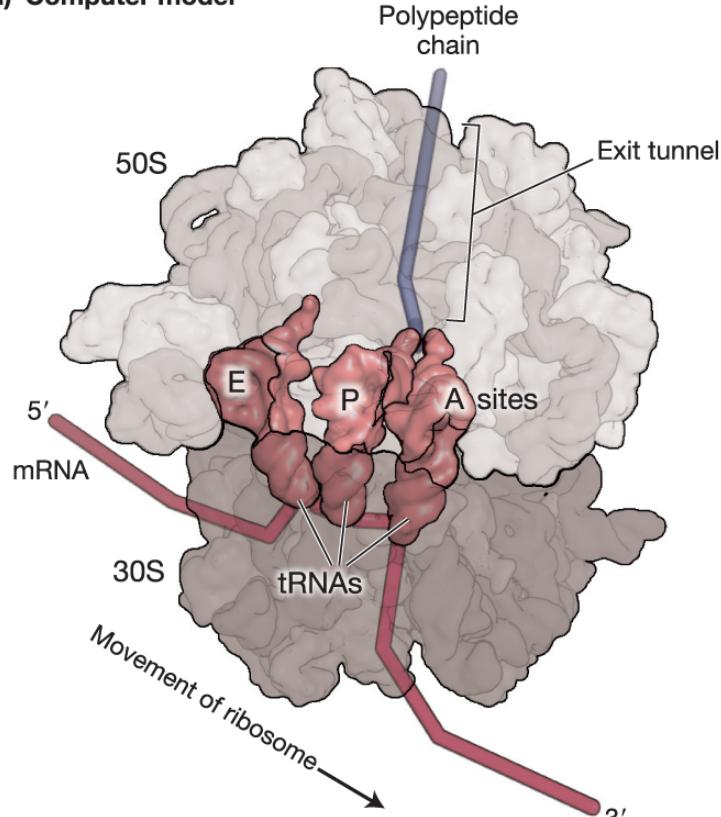
Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

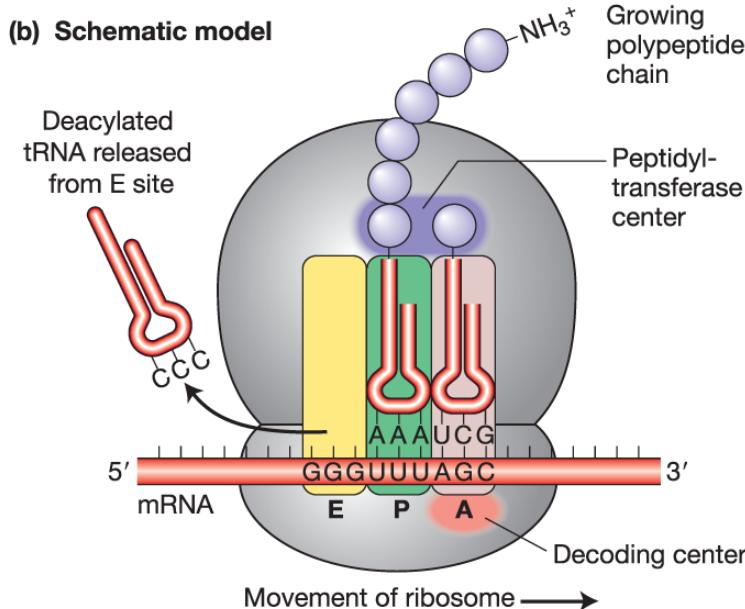
FIGURE 9-15 The folded structure of bacterial 16S rRNA. The magnified area shows the details of the complex secondary structure of 16S rRNA and the sequence at the 3' end of 16S that binds an mRNA Shine–Dalgarno sequence.

Ribosomes bring together the important players in protein synthesis—charged tRNA and mRNA—to translate the nucleotide sequence of an mRNA into the amino acid sequence of a protein. tRNAs and mRNAs are positioned in the ribosome so that codons of the mRNA can interact with anticodons of tRNAs. Key sites of interaction are illustrated in [Figure 9-16](#). The binding site for mRNA is completely within the small subunit. There are three binding sites for tRNA molecules. Each bound tRNA bridges the 30S and 50S subunits, positioned with its anticodon end in the 30S subunit and its aminoacyl end (carrying the amino acid) in the 50S subunit. The [A site](#) (for [aminoacyl-tRNA binding site](#)) binds an incoming aminoacyl-tRNA whose anticodon is complementary to the mRNA codon in the A site of the 30S subunit. Proceeding in the 5' direction on the mRNA, the next codon interacts with the anticodon of the tRNA in the [P site](#) (for [peptidyl site](#)) of the 30S subunit. The P and A sites are situated to facilitate formation of a peptide bond between their amino acids, disconnecting the P site amino acid from its tRNA. The growing peptide chain fits into a tunnel-like structure in the 50S subunit. The [E site](#) (for [exit site](#)) contains a deacylated tRNA (it no longer carries an amino acid) that is ready to be released from the ribosome. Whether codon–anticodon interactions also take place between the tRNA and the mRNA in the E site is not yet clear.

Key sites of interaction in the ribosome

(a) Computer model





Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020

W. H. Freeman and Company

Part (a) drawn using PDB IDs 1VSA, 2OW8, and 1GIX.

FIGURE 9-16 Ribosome interactions during the elongation phase of translation. (a) A computer model of the three-dimensional structure of the ribosome with mRNA, tRNAs, and the nascent polypeptide chain as it emerges from the large ribosomal subunit. (b) A schematic model of the ribosome during translation elongation.

KEY CONCEPT mRNA base pairs with tRNA in the small subunit of the ribosome, while tRNAs fit into sites that span both subunits. tRNAs begin in the A site, peptide bond formation occurs in the P site, and tRNAs exit from the E site.

Two additional regions in the ribosome are critical for protein synthesis. The **decoding center** in the 30S subunit ensures that only tRNAs carrying anticodons that match the codon will be accepted into the A site. The **peptidyltransferase center** in the 50S subunit is the site where peptide bond formation is catalyzed. The Nobel Prize in Chemistry was awarded in 2009 to Thomas Steitz, Venkatraman Ramakrishnan, and Ada Yonath for their laboratories' work using X-ray crystallography to determine the structure of the ribosome at the atomic level. The results of their elegant studies clearly show that both the decoding and peptidyltransferase centers are composed entirely of rRNA. Therefore, peptide bond formation is thought to be catalyzed by an active site in the rRNA and assisted only by ribosomal proteins. In other words, the large ribosomal subunit functions as a ribozyme to catalyze peptide bond formation.

KEY CONCEPT In all organisms, ribosomes have large and small subunits, each containing rRNA and proteins. Two key sites in ribosomes, the peptidyltransferase center in the large subunit that carries out peptide bond formation and the decoding center in the small subunit that accepts the correct tRNA, are built of rRNA.

9.4 TRANSLATION

LO 9.5 Outline the molecular events that take place during translation initiation, elongation, and termination.

Translation is carried out by ribosomes moving along mRNA in the 5'-to-3' direction. tRNAs bring amino acids to the ribosome, and their anticodons base pair to mRNA codons. An incoming amino acid becomes bonded to the amino end of the growing polypeptide chain in the ribosome. The process of translation can be divided into three phases: initiation, elongation, and termination. Aside from the ribosome, mRNA, and tRNAs, other proteins (factors) are required for each phase (**Table 9-2**). Because certain steps in initiation differ significantly in bacteria and eukaryotes, initiation is described separately for the two groups. The elongation and termination phases are described largely as they take place in bacteria.

TABLE 9-2 Translation Factors

| | Bacteria | Eukaryotes | Function |
|--------------------|-----------------------------|--------------------------------------|-------------------------------------|
| Initiation | fMET - tRNA ^{fMET} | Met-tRNA _i ^{Met} | Initiator tRNA |
| | IF1 | eIF1A | Blocks A site |
| | IF2 | eIF2, eIF5B | Entry of initiator tRNA |
| | IF3 | eIF3, eIF1 | Blocks association of large subunit |
| | | eIF4 complex | |
| | | eIF4A | Unwinds mRNA |
| | | eIF4E | Binds m ⁷ G cap |
| | | eIF4G | Binds PABP to circularize mRNA |
| Elongation | EF-Tu | eEF1 α | Delivers aminoacyl tRNA |
| | EF-G | eEF2 | Translocates ribosome |
| Termination | RF1 | eRF1 | Recognizes UAA and UAG stop codons |
| | RF2 | eRF1 | Recognizes UAA and UGA stop codons |
| | RF3 | eRF3 | Stimulates peptide release |

Translation initiation

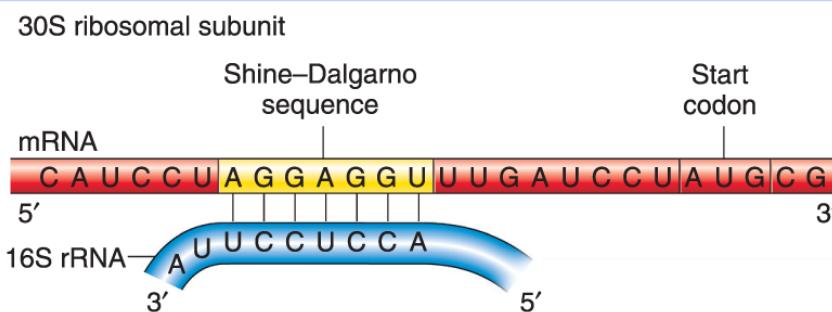
The main task of initiation is to place the first aminoacyl-tRNA in the P site of the ribosome and, in this way, establish the correct reading frame of the mRNA. In most bacteria and all eukaryotes,

the first amino acid in any newly synthesized polypeptide is methionine (Met), specified by the **initiation codon** AUG. In bacteria, there are two tRNAs for methionine. tRNA^{Met} is used at AUGs in internal positions in mRNA and an **initiator tRNA**, tRNA^{fMet}, is used at AUG initiation codons. tRNA^{fMet} is charged with methionine to form Met-tRNA^{fMet} and then a formyl (f) group is added to methionine to generate *N*-formylmethionyl-tRNA^{fMet} (fMet-tRNA^{fMet}). The formyl group on fMet-tRNA^{fMet} is removed during or shortly after synthesis of the polypeptide. Eukaryotes also use distinct methionine tRNAs for internal and initiation AUG codons, called tRNA^{Met} and tRNA_i^{Met}, respectively. The use of Met-tRNA_i^{Met} rather than Met-tRNA^{Met} for initiation is specified by interactions with translation initiation factors.

How does the translation machinery know where to begin? In other words, how is the initiation AUG codon selected from among the many AUG codons in an mRNA? Recall that, in both bacteria and eukaryotes, mRNA has a 5' untranslated region (UTR) consisting of the sequence between the transcription start site and the translational start site (see [Figure 8-7](#)). As you will see below, the nucleotide sequence of the 5' UTR adjacent to the AUG initiation codon is critical for ribosome binding in bacteria, and the 5' cap is critical for ribosome binding and scanning for the AUG initiation codon in eukaryotes.

In bacteria, AUG initiation codons in mRNA are preceded by a special sequence called the **Shine-Dalgarno sequence**, also known as the **ribosome-binding site (RBS)**, that base pairs with the 3' end of 16S rRNA in the 30S ribosomal subunit ([Figure 9-17](#)). This base pairing correctly positions the AUG in the P site where the initiator tRNA will bind. The mRNA can interact only with a 30S subunit that is dissociated from the rest of the ribosome. Note again that rRNA performs the key function in ensuring that the ribosome is at the right place to start translation. Once the initiator tRNA is bound, the 50S ribosomal subunit binds to form the 70S initiation complex.

The Shine–Dalgarno sequence positions the mRNA in the ribosome



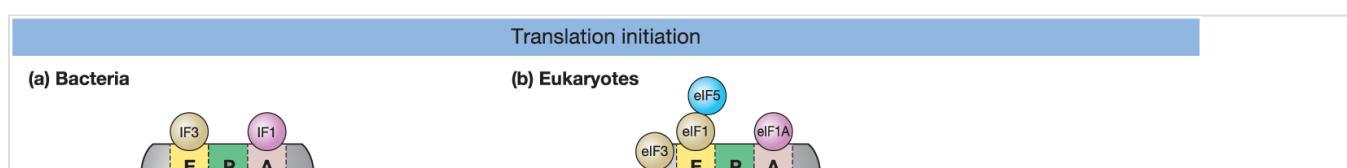
Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

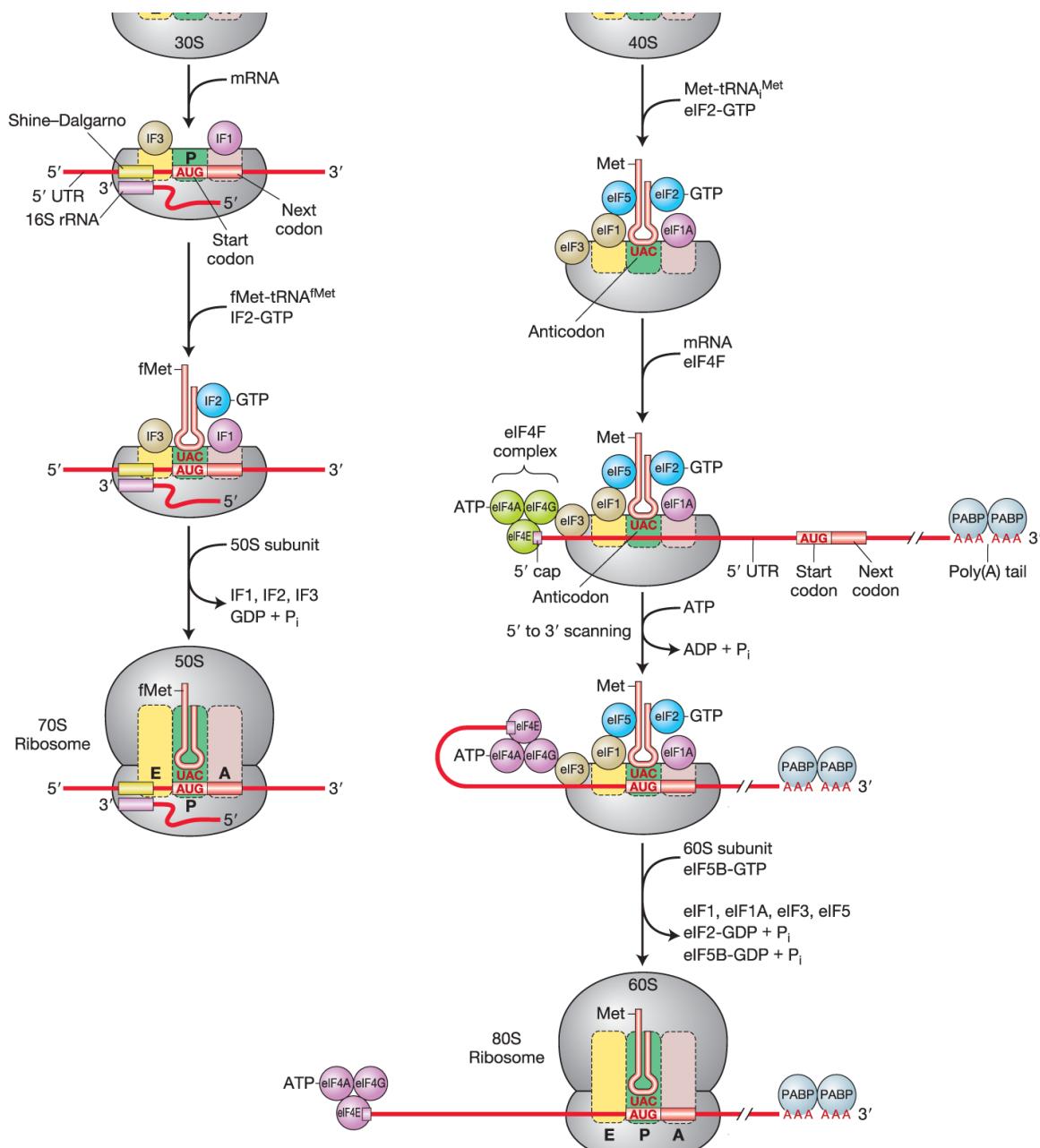
FIGURE 9-17 In bacteria, base pairing between the 3' end of the 16S rRNA of the small ribosomal subunit (30S) and the Shine–Dalgarno sequence of the mRNA positions the ribosome to correctly initiate translation at the downstream AUG initiation codon.

In eukaryotes, translation initiation involves binding of the 40S ribosomal subunit to the capped 5' end of an mRNA, followed by scanning of the 5' UTR for an AUG initiation codon. The 5' cap (m^7G) that is added to mRNAs during transcription (described in [Chapter 8](#)) is directly bound by a translation initiation factor, which in turn binds other initiation factors to recruit the small ribosomal subunit to the mRNA. The ribosome subsequently scans the mRNA in the 5'-to-3' direction until it encounters the first AUG codon. In 5 to 10 percent of cases, the ribosome will bypass the first AUG and initiate at the second, third, or subsequent AUG. Marilyn Kozak found that bypass occurs because the sequence surrounding the AUG codon affects the efficiency of its use in initiation. Kozak found that the sequence $\text{CC(A}^{-3}/\text{G})\text{CCA}^{+1}\text{UGG}^{+4}$, called the [Kozak sequence](#), commonly surrounds initiating AUGs, and mutagenesis studies showed that the A at -3 and the G at +4 are particularly important for specifying the AUG initiation codon.

KEY CONCEPT Translation initiation begins when a charged initiator tRNA anticodon and an mRNA AUG initiation codon assemble in the P site of a ribosomal small subunit. In bacteria, the Shine–Dalgarno sequence in the 5' UTR of the mRNA base pairs with the 16S rRNA to position the start codon in the P site. In eukaryotes, the 5' cap of the mRNA is bound by an initiation factor that recruits the ribosomal small subunit. Once bound, the ribosome scans the mRNA for the initiation codon within the Kozak sequence.

In bacteria, three proteins—IF1, IF2, and IF3 (for initiation factor)—are required to assemble an active 70S ribosome ([Figure 9-18a](#)). Assembly begins by positioning the mRNA AUG initiation codon in the 30S ribosomal subunit P site through base pairing of the Shine–Dalgarno sequence with 16S rRNA and by base pairing of fMet-tRNA^{fMet} with the AUG initiation codon. During these early assembly steps, IF2 bound to GTP promotes binding of fMet-tRNA^{fMet} to the P site. In addition, binding of IF1 in the A site blocks tRNA binding to the second codon, and binding of IF3, with help from IF1, blocks association of the 50S subunit. Once formation of the 30S initiation complex is complete, IF3 and IF1 are released, which enables association of the 50S subunit, and hydrolysis of GTP to GDP leads to release of IF2 to generate a functional 70S ribosome. Hydrolysis of GTP and sometimes ATP provides energy for conformation changes that are required to progress through several steps of translation.





Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 9-18 Initiation factors assist assembly of the ribosome at the translation start site and then dissociate before translation elongation. (a) In bacteria, three initiation factors (IF1, IF2, and IF3) position fMet-tRNA^{fMet} in the P site of the 70S ribosome. (b) In eukaryotes, many eIFs are required to position Met-tRNA_i^{Met} in the P site of the 80S ribosome. Factors that serve similar functions in bacteria and eukaryotes are colored the same.

The
ANIMATED ART Sapling Plus

Translation

Like bacteria, eukaryotes use a set of initiation factors (IFs) to assemble the 80S translation initiation complex containing the small and large ribosomal subunits, an mRNA bound to the small subunit, and a charged initiator tRNA (Met-tRNA_i^{Met}) in the P site that is base paired to the

AUG initiation codon ([Figure 9-18b](#)). The names of eukaryotic factors all begin with an “e” to distinguish them from bacterial factors. Four eIFs have functions similar to IFs in bacteria. Like bacterial IF1, eIF1A blocks tRNA binding in the A site of the small subunit (40S) as well as premature association of the large subunit (60S) with the small subunit. Eukaryotes have two factors, eIF2 and eIF5B, that have functions similar to bacterial IF2. A ternary complex composed of eIF2, GTP, and Met-tRNA_i^{Met} associates with the small ribosomal subunit, while eIF5B promotes association of the small and large subunits. In addition, both eIF2 and eIF5B use hydrolysis of GTP to GDP to carry out their jobs. Lastly, like bacterial IF3, eIF3 binds the small ribosomal subunit and blocks association of the large subunit.

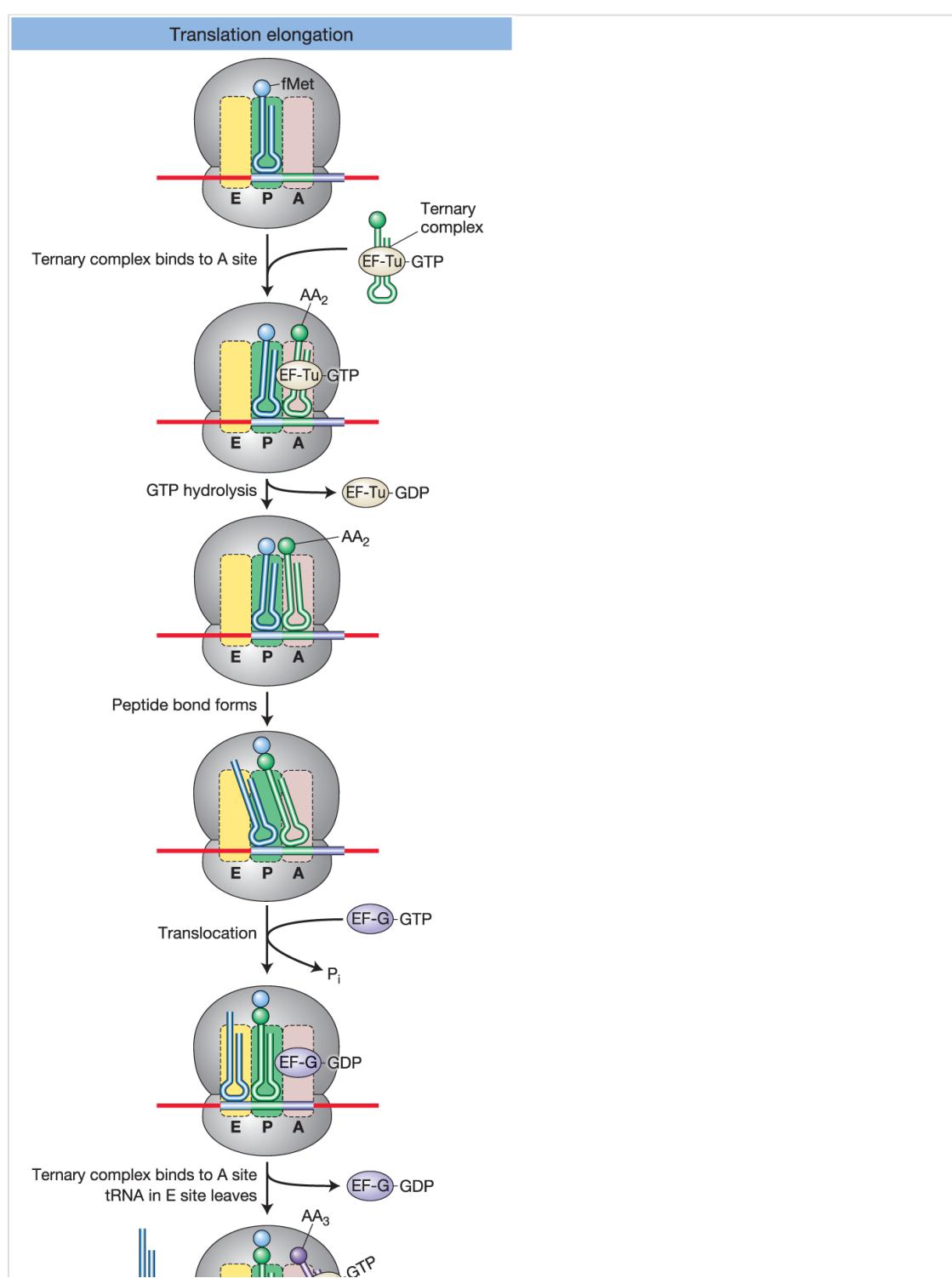
Because of differences between eukaryotic and bacterial mRNAs, there are other eIFs that do not have counterparts in bacteria. Three eIFs—eIF4A, eIF4E, and eIF4G—interact with one another in a complex called eIF4F that carries out activities unique to eukaryotes. eIF4E binds the cap structure at the 5' end of an mRNA. m⁷G caps are present only on mRNAs, so the requirement for cap binding by eIF4E ensures that only mRNAs are translated. eIF4A has RNA helicase activity that unwinds regions in the 5' UTR that are double-stranded RNA due to intramolecular base pairing. This allows ribosomes to scan along single-stranded RNA in search of the AUG initiation codon. Hydrolysis of ATP to ADP by eIF4A is required for its activity. Lastly, eIF4G binds poly(A) binding proteins (PABPs) that associate with the mRNA poly(A) tail, thereby bringing together the 5' and 3' ends of the mRNA. Circularization of the mRNA is thought to enhance the rate of translation of capped and polyadenylated mRNAs by coordinating the initiation of ribosomes that have recently terminated translation.

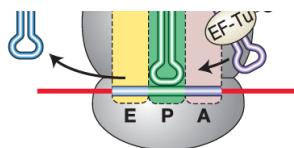
In the first round of translation, the function of eIF4F is carried out by the cap binding complex (CBC), which associates with newly synthesized mRNAs in the nucleus and is transported along with the mRNA to the cytoplasm ([Chapter 8](#)). Another eukaryotic-specific factor is eIF1, which associates with eIF4G and promotes scanning for the AUG initiation codon. Finally, as in bacteria, eukaryotic initiation factors dissociate from the ribosome before the elongation phase of translation begins. The exception is eIF4F, which remains associated with PABP, setting the stage for additional rounds of translation.

KEY CONCEPT In both bacteria and eukaryotes, initiation factors bring the initiator tRNA to the ribosomal small subunit and prevent premature binding of the large subunit. Some initiation factors hydrolyze GTP to proceed through the steps of initiation. In eukaryotes, additional initiation factors facilitate mRNA scanning and circularization.

Translation elongation

During translation elongation, the ribosome functions as a factory, repeating the same steps over and over again. The mRNA acts as a blueprint specifying the delivery of tRNAs, each carrying as cargo an amino acid. Each amino acid is added to the growing polypeptide chain, while the deacylated tRNA is recycled by being charged with another amino acid. [Figure 9-19](#) details the steps in elongation. In bacteria, two protein elongation factors (EFs) called elongation factor Tu (EF-Tu) and elongation factor G (EF-G) assist the elongation process.





Griffiths et al., *Introduction to Genetic Analysis*, 12e,
© 2020 W. H. Freeman and Company

FIGURE 9-19 In bacteria, two elongation factors, EF-Tu and EF-G, perform repetitive functions for each amino acid that is added to the growing polypeptide. EF-Tu escorts charged tRNAs to the A site and positions them for peptide bond formation with the peptide attached to the tRNA in the P site. The protein factor EF-G then drives the repositioning of tRNAs from the P and A sites into the E and P sites, respectively.

As described earlier in this chapter, an aminoacyl-tRNA is formed by the covalent attachment of an amino acid to the 3' end of a tRNA that contains the correct anticodon. Before aminoacyl-tRNAs can be used in protein synthesis, they associate with the protein factor EF-Tu to form a ternary complex composed of EF-Tu, GTP, and aminoacyl-tRNA. The elongation cycle commences with fMet-tRNA^{fMet} in the P site and with the A site ready to accept a ternary complex (Figure 9-19). Codon–anticodon recognition in the decoding center of the small subunit determines which of the different ternary complexes to accept (see Figure 9-16b). When the correct match has been made, the ribosome changes shape, EF-Tu hydrolyzes GTP to GDP and leaves the ternary complex, and the two amino acids are juxtaposed in the peptidyltransferase center of the large subunit (see Figure 9-16b). There, a peptide bond is formed with transfer of fMet in the P site to the amino acid in the A site.

At this point, the second elongation factor, EF-G, plays its part. EF-G is structurally similar to a ternary complex and fits into the A site, displacing the peptidyl-tRNA. Hydrolysis of GTP to GDP by EF-G changes its structure as well as that of the ribosome and shifts the tRNAs in the A and P sites to the P and E sites, respectively. When EF-G leaves the ribosome, the A site is open to accept the next ternary complex. As elongation progresses, the number of amino acids on the peptidyl-tRNA (at the P site) increases. Eventually, the amino-terminal end of the growing polypeptide emerges from the tunnel in the 50S subunit and protrudes from the ribosome.

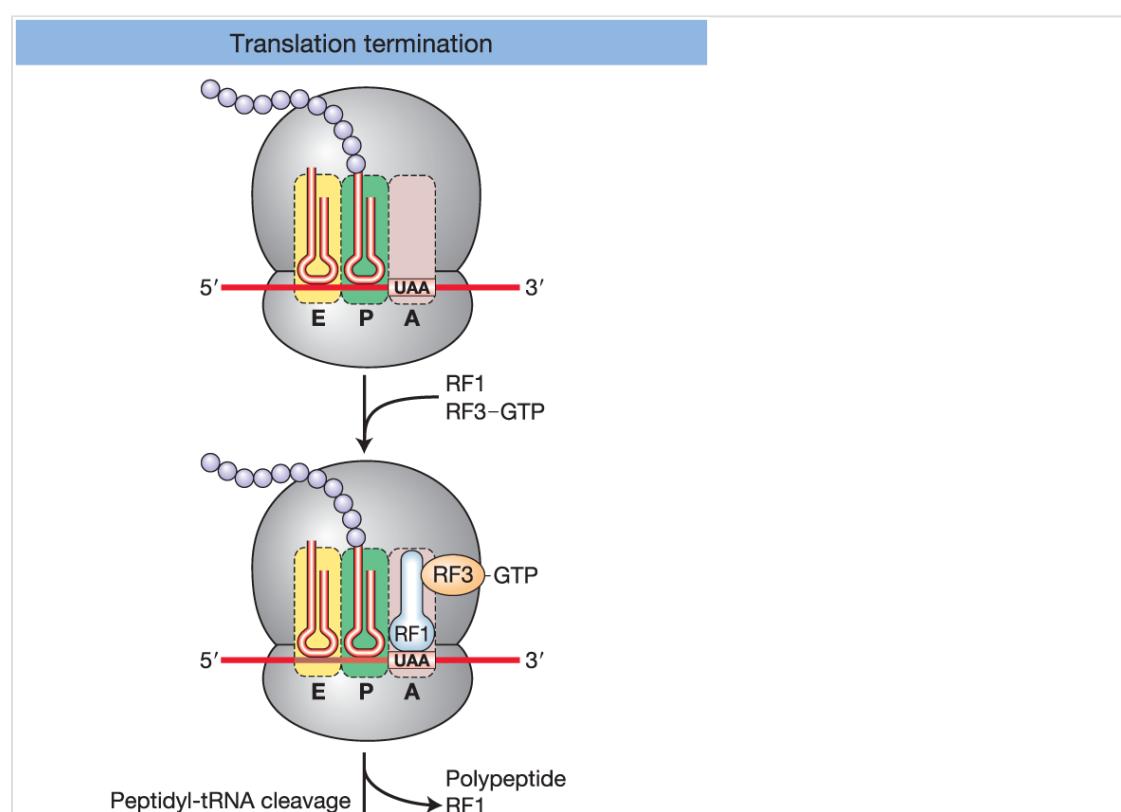
Translation elongation in eukaryotes is very similar. Eukaryotic elongation factor 1α (eEF1α) functions similarly to EF-Tu, and eEF2 functions similarly to EF-G.

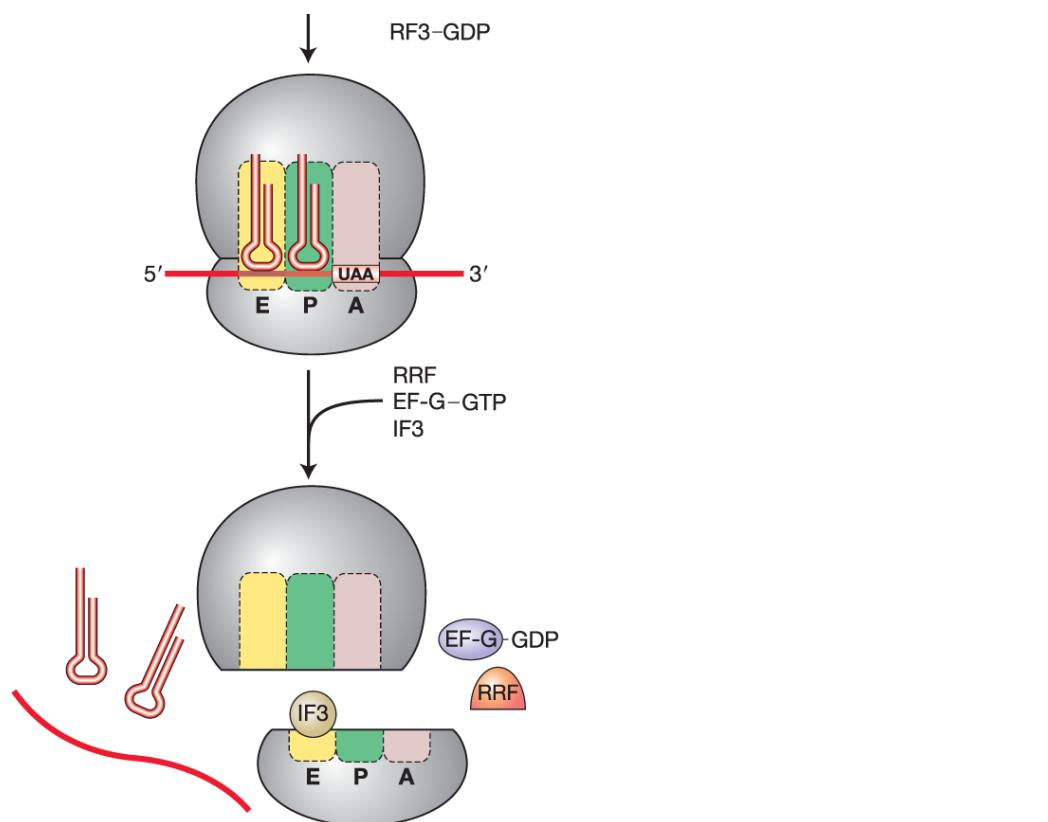
KEY CONCEPT During translation elongation, two protein elongation factors function repeatedly to grow the polypeptide chain. In bacteria, one elongation factor, EF-Tu, brings charged tRNAs to the ribosome A site to be

joined to the polypeptide chain in the P site. The other elongation factor, EF-G, binds in the A site and promotes the translocation of tRNAs from the P and A sites to the E and P sites, respectively. In eukaryotes, the analogous elongation factors are eEF1 α and eEF2.

Translation termination

The elongation cycle continues until the codon in the A site is one of the three stop codons: UGA, UAA, or UAG. tRNAs do not recognize these codons. Instead, proteins called release factors (RFs) recognize stop codons ([Figure 9-20](#)). Just like the structure of EF-G mimics the structure of a ternary complex, RF1 and RF2 mimic the structure of a tRNA. In bacteria, RF1 recognizes UAA or UAG, whereas RF2 recognizes UAA or UGA. The interaction between RF1 or RF2 and the A site differs from that of the ternary complex in two important ways. First, the stop codons are recognized by tripeptides in the RF proteins, not by an anticodon. Second, RFs fit into the A site of the 30S subunit but do not participate in peptide bond formation. Instead, a water molecule gets into the peptidyltransferase center, and its presence leads to release of the polypeptide from the tRNA in the P site. Following release of the peptide chain, RF3 promotes release of RF1 or RF2 from the ribosome. Hydrolysis of GTP to GDP is involved in releasing RF3 from the ribosome. In eukaryotes, translation termination is very similar; eRF1 recognizes stop codons, and eRF3 stimulates peptide chain release by eRF1. However, unlike bacteria, eRF1 recognizes all three stop codons (UAA, UAG, and UGA).





Griffiths et al., *Introduction to Genetic Analysis*, 12e,
© 2020 W. H. Freeman and Company

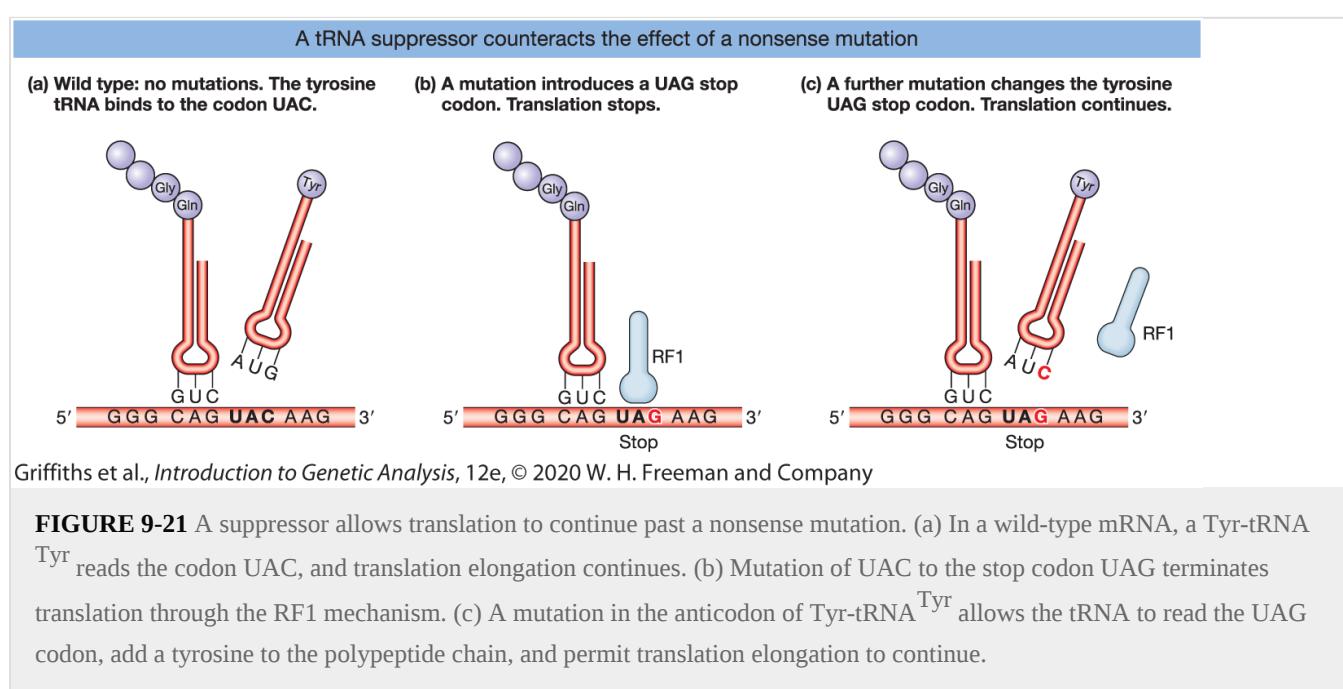
FIGURE 9-20 In bacteria, translation is terminated when a release factor (RF1 or RF2) recognizes a stop codon in the A site of the ribosome and liberates the polypeptide chain. RF3 then promotes the release of RF1 or RF2 (RF1, in this example). To get ready for another round of translation, other factors (RRF, EF-G, and IF3) displace the uncharged tRNAs and dissociate the ribosome subunits from one another and from the mRNA.

To prepare for a new round of translation, the ribosome recycling factor (RRF) disassembles the post-termination complex ([Figure 9-20](#)). With the help of EF-G and IF-3, RRF binds in the A site, translocates to the P site, releases deacylated tRNAs from the E and P sites, and dissociates the small and large ribosome subunits from each other and from the mRNA. IF-3 bound to the small subunit is now ready to initiate translation (see [Figure 9-18a](#)).

KEY CONCEPT During translation termination, release factors (proteins, not tRNAs) bind stop codons and release the polypeptide chain from the tRNA in the P site. Other factors recycle the ribosome to begin another round of translation.

Nonsense suppressor mutations

Experiments using nonsense suppressor mutations show that the ribosome relies on the anticodon-codon interaction in its decoding center to ensure fidelity in translation. In these experiments, wild-type codons in phages were mutated to stop codons, creating nonsense mutations that resulted in truncated phage proteins. However, suppressor mutations in the host chromosome counteracted the effects of these mutations. Many of these suppressors are mutations in genes encoding tRNAs and are known as tRNA suppressors. These mutations alter the anticodon loops of specific tRNAs in such a way that a tRNA becomes able to recognize a stop codon in mRNA. In [Figure 9-21](#), a mutation replaces a wild-type codon with the chain-terminating stop codon UAG. By itself, the UAG would cause the protein to be prematurely cut off at the corresponding position. The suppressor mutation in this case produces a tRNA^{Tyr} with an anticodon that recognizes the mutant UAG stop codon. Thus, in the suppressed mutant, tRNA^{Tyr} competes with the release factor for access to the UAG stop codon. As a result, if tyrosine is inserted, translation continues past that triplet.



ANIMATED ART Sapling Plus

Nonsense suppression at the molecular level

Presumably, tRNA suppressors also bind to normal termination signals and result in the synthesis of abnormally long proteins. Now that many genomes have been sequenced, it is known that the UAA stop codon is used much more often than UAG or UGA to terminate protein synthesis. As such, cells with UAA suppressors are usually sicker than cells with UAG or UGA suppressor mutations.

KEY CONCEPT Experiments with suppressor mutations show that the ribosome cannot proofread the match between the tRNA anticodon and the amino acid.

9.5 TRANSLATIONAL AND POST-TRANSLATIONAL REGULATION

LO 9.6 Describe how protein synthesis and function are regulated.

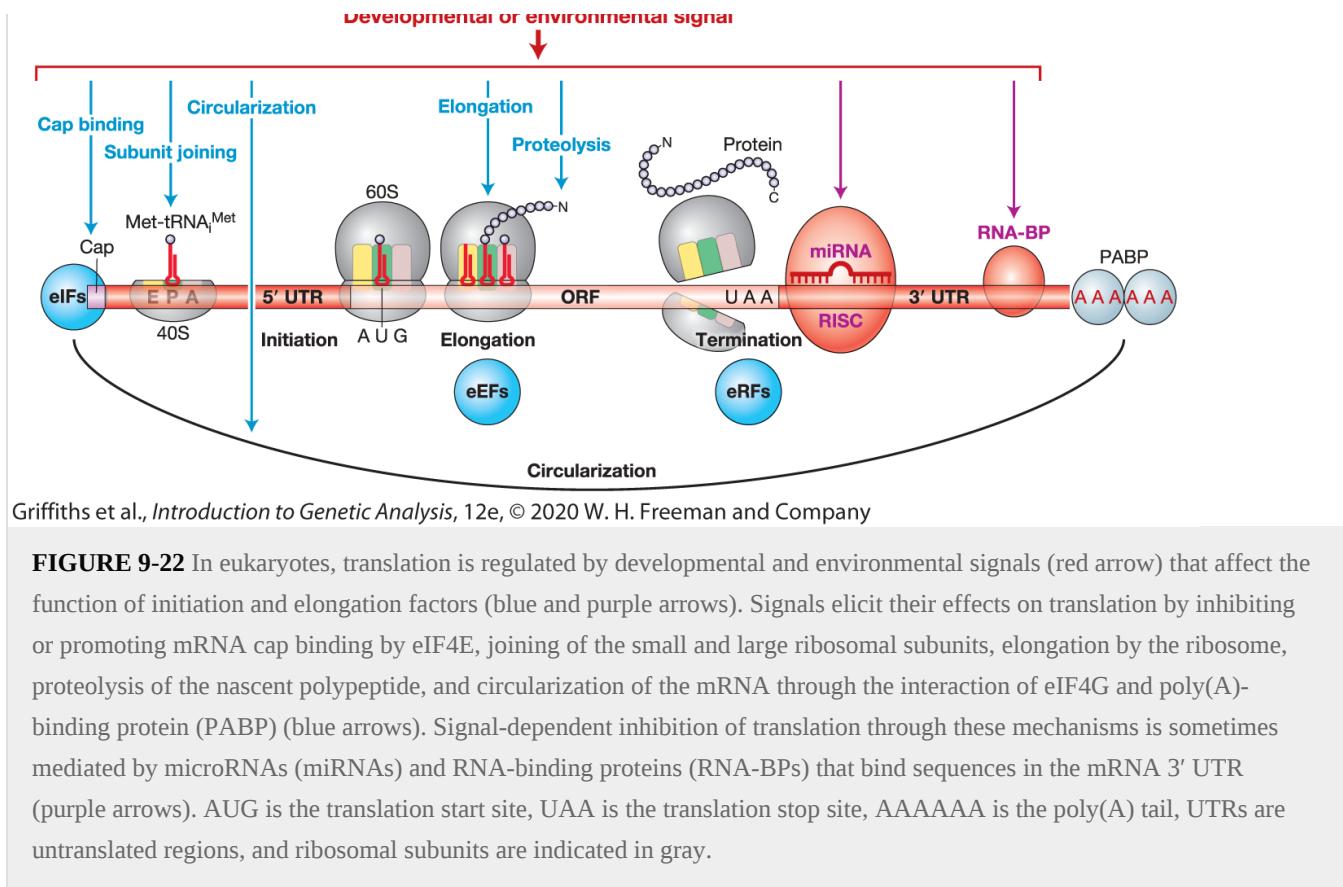
Translation is a regulated process in eukaryotic cells. As a result, the quantity of an mRNA is not always representative of the quantity of its encoded protein. For example, some mRNAs localize to the cell cytoplasm but are not assembled with ribosomes and translated until the cell receives a particular signal. In general, translation is controlled by signals from outside and inside a cell that alter the function of both general and specific translation initiation or elongation factors.

Stresses that are not favorable to growth, including nutrient deprivation, temperature shock, and DNA damage, produce signals that cause a general halt to translation and allow only selective translation of a few mRNAs encoding proteins that are required for responding to the stress. In contrast, favorable growth conditions that are rich in nutrients and growth stimuli lead to a global increase in translation and stimulated translation of specific mRNAs encoding proteins involved in cell growth (increase in cell size), proliferation (increase in cell number), and survival.

| | |
|------------------|---|
| Stress → | decrease in general transcription |
| | increase in transcription of stress-response proteins |
| Growth stimuli → | permissive to general transcription |
| | increase in transcription of cell growth proteins |

In eukaryotes, signals that inhibit the translation of many mRNAs often act by directly altering the ability of eIF4E to bind mRNA 5' caps and assemble eIF4F to initiate translation (see [Figure 9-18](#)). On the other hand, signals that inhibit the translation of particular mRNAs act through microRNAs (miRNAs) or RNA-binding proteins (RNA-BPs) that bind specific sequences or structural motifs in an mRNA's 5' or 3' UTR ([Figure 9-22](#)). Bound factors in turn block translation through multiple mechanisms, including inhibiting ribosome assembly and elongation, inhibiting mRNA circularization, and promoting cleavage of the polypeptide chain.

Signal-dependent regulation of translation



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 9-22 In eukaryotes, translation is regulated by developmental and environmental signals (red arrow) that affect the function of initiation and elongation factors (blue and purple arrows). Signals elicit their effects on translation by inhibiting or promoting mRNA cap binding by eIF4E, joining of the small and large ribosomal subunits, elongation by the ribosome, proteolysis of the nascent polypeptide, and circularization of the mRNA through the interaction of eIF4G and poly(A)-binding protein (PABP) (blue arrows). Signal-dependent inhibition of translation through these mechanisms is sometimes mediated by microRNAs (miRNAs) and RNA-binding proteins (RNA-BPs) that bind sequences in the mRNA 3' UTR (purple arrows). AUG is the translation start site, UAA is the translation stop site, AAAAAA is the poly(A) tail, UTRs are untranslated regions, and ribosomal subunits are indicated in gray.

miRNAs are small RNAs (~21 nucleotides) that bind with imperfect complementarity to their target mRNAs. Humans express over 2,500 miRNAs, and about 50% of human mRNAs are subject to regulation by miRNAs. So, most biological processes, including cell differentiation, growth, and proliferation, are regulated by miRNAs. Single-stranded miRNAs associate with Argonaute (Ago) proteins and other proteins to form an RNA-induced silencing complex (RISC). Guided to specific mRNAs by sequence complementary between the miRNA and the target mRNA, RISC interacts with other proteins to inhibit translation as well as promote mRNA decay.

Most newly synthesized proteins are unable to function until regulatory mechanisms alter their structure and cellular location. To become functional, proteins need to be folded correctly, the amino acids of some proteins need to be chemically modified, and proteins need to be transported to their sites of action within or outside the cell. Some protein folding, modification, and targeting takes place co-translationally (while the protein is being synthesized), and the rest takes place post-translationally (after synthesis is complete).

KEY CONCEPT Developmental and environmental signals regulate the translation of many mRNAs at the same time by altering the function of translation initiation factors and of selected mRNAs via RNA-binding factors (proteins and RNAs) that act on initiation and elongation factors.

Protein folding

Protein folding is the process by which proteins attain their functional tertiary structure. A protein that is folded correctly is said to be in its native conformation (in contrast with an unfolded or misfolded protein that is nonnative). Folding involves the stepwise formation of secondary structures such as α -helices and β -sheets that are stabilized by non-covalent hydrogen bonds. Folded regions then guide and stabilize subsequent folding to progressively build the tertiary structure. The distinct three-dimensional structures of proteins are essential for their enzymatic activity, for their ability to bind to other molecules, and for their structural roles in the cell. Although it has been known since the 1950s that the amino acid sequence of a protein determines its three-dimensional structure, it is also known that the aqueous environment inside the cell does not favor the correct folding of most proteins. Given that proteins do in fact fold correctly in the cell, a long-standing question has been, how is correct folding accomplished?

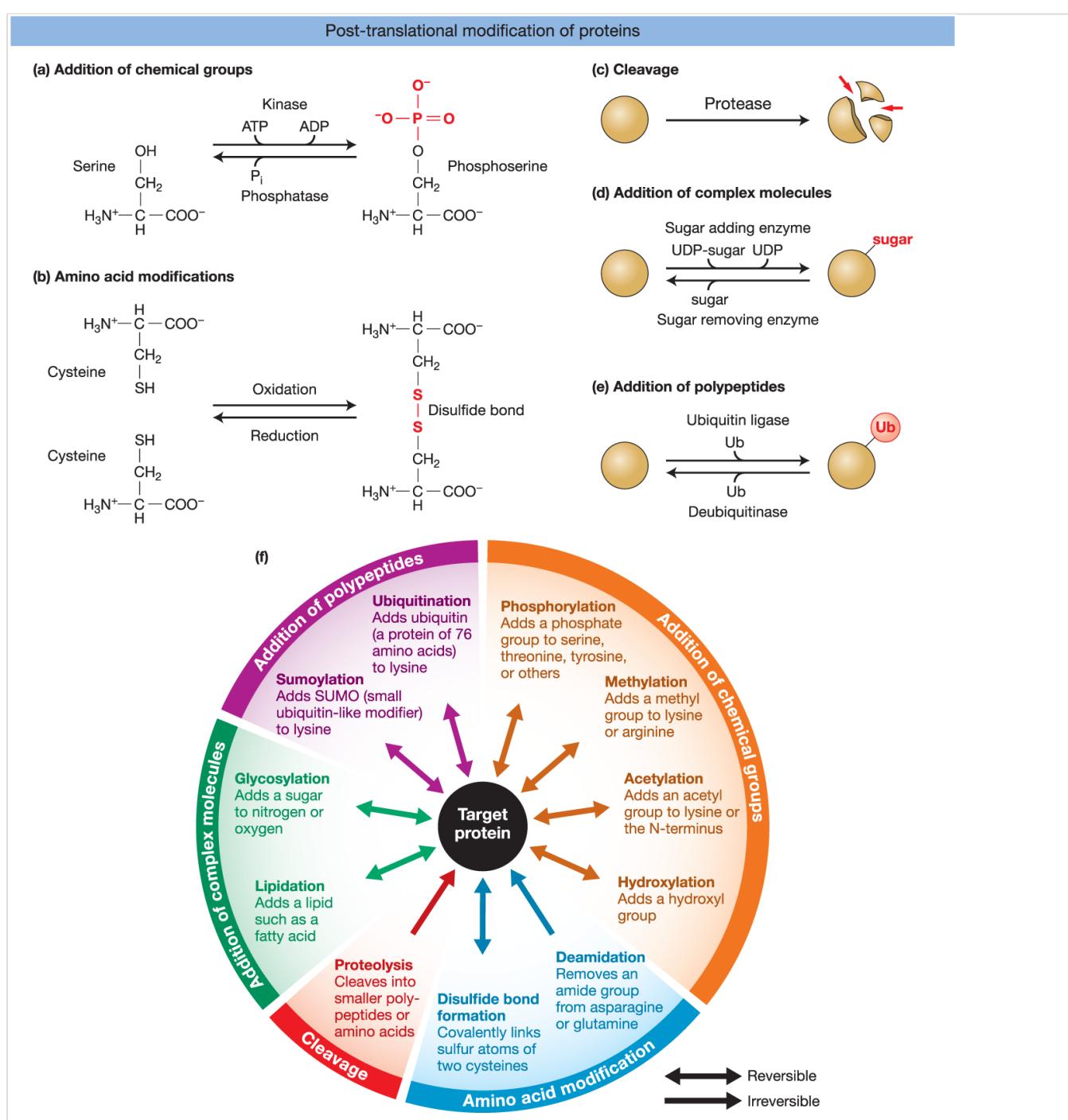
The answer seems to be that folding often begins co-translationally and is helped by **chaperones**—a class of proteins found in all organisms from bacteria to plants to humans. Chaperones typically bind hydrophobic regions of incorrectly or incompletely folded proteins to arrest folding or promote unfolding and then release them to undergo spontaneous refolding. Multiple rounds of binding and release that are driven by ATP hydrolysis occur until the protein is properly folded. Chaperones exist in all cell compartments. Some chaperones are expressed all the time in cells, while others are upregulated by heat shock and other stresses that increase protein misfolding. The latter class of chaperones is classified as stress or heat shock proteins.

KEY CONCEPT The folding of newly synthesized proteins into precise three-dimensional structures is determined by the primary amino acid sequence and assisted by a class of proteins called chaperones.

Post-translational modification of amino acid side chains

Chemical modifications of amino acids greatly increase the functionality of proteins. More than 200 different types of amino acid modifications have been identified, many of which occur post-translationally. Amino acids that undergo post-translational modification often have a functional group that acts as a nucleophile in the enzymatic modification reaction. Examples of nucleophiles include the hydroxyl group of serine, threonine, and tyrosine; the amine group of lysine, arginine,

and histidine, the thiol group of cysteine; and the carboxylate group of aspartic acid and glutamic acid (see [Figure 9-2](#)). The modifications themselves fall into five broad categories: the addition of chemical groups, complex molecules, or polypeptides to amino acids; the modification of amino acids; and the cleavage of peptide bonds between amino acids to convert inactive precursor proteins into smaller, active proteins ([Figure 9-23](#)).



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 9-23 Proteins can be reversibly and irreversibly modified in numerous ways. (a–e) Examples of the five types of post-translational modifications, shown in red. (f) A larger list of examples of the five types of post-translational modifications.

KEY CONCEPT Proteins can undergo five types of post-translational modification:

1. Addition of chemical groups to amino acids (e.g., phosphorylation, methylation, acetylation, hydroxylation)
 2. Addition of complex molecules to amino acids (e.g., glycosylation, lipidation)
 3. Addition of polypeptides to amino acids (e.g., ubiquitination, sumoylation)
 4. Modification of amino acids (e.g., deamidation, disulfide bond formation)
 5. Cleavage of peptide bonds to convert inactive precursor proteins into smaller, active proteins (proteolysis)
-

Three types of proteins are involved in chemical modifications: enzymes called *writers* that add chemical modifications, enzymes called *erasers* that remove chemical modifications, and structural proteins called *readers* that bind chemical modifications. Sometimes, a single protein or a protein complex can be both a writer or eraser and a reader.

Addition and removal of chemical modifications serves as a reversible switch to control the traits of proteins, including increasing or decreasing their biological activity, promoting or inhibiting their transport between cell compartments, increasing or decreasing their stability, and enhancing or suppressing their interactions with DNA, RNA, and other proteins. Two of the more commonly encountered post-translational modifications—phosphorylation and ubiquitination—are considered next.

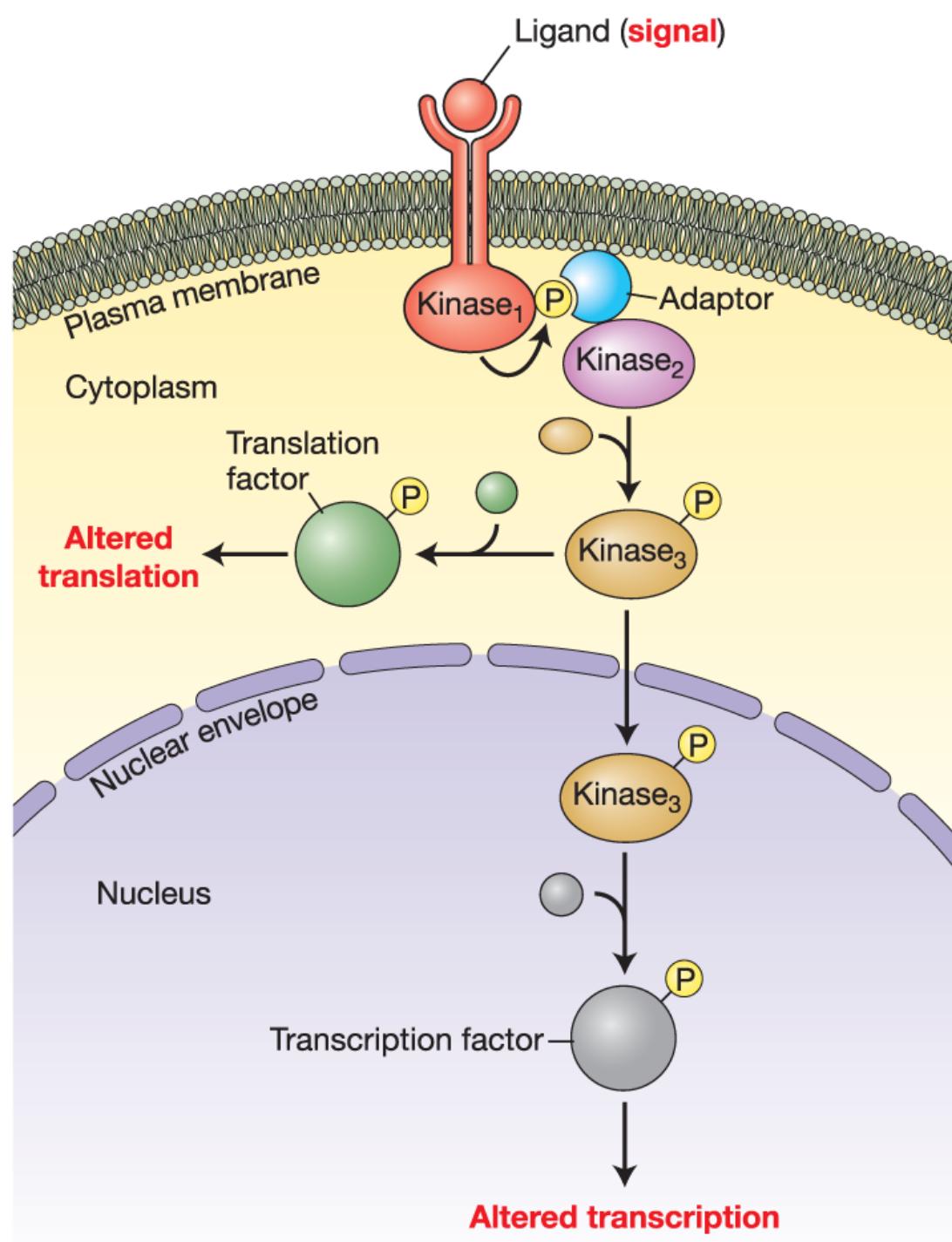
Phosphorylation

Enzymes called **kinases** (i.e., writers) catalyze **phosphorylation**, the addition of a phosphate group from ATP to the hydroxyl group of the amino acids serine, threonine, or tyrosine, whereas enzymes called **phosphatases** (i.e., erasers) catalyze dephosphorylation, resulting in the removal of phosphate groups. Protein phosphorylation is an important regulatory mechanism in eukaryotic cells. For example, phosphorylation and other modifications of the C-terminal domain (CTD) of RNA polymerase II regulate the processing of nascent mRNAs (see [Figure 8-13](#)). It is estimated that more than 30 percent of human proteins are regulated by phosphorylation, and abnormal phosphorylation is the cause or consequence of many human diseases.

Post-translational modifications are also used by eukaryotic cells to rapidly convert signals from the cellular environment into changes in gene expression; that is, changes in the transcription and translation of particular genes that allow the cell to respond to the signal ([Figure 9-24](#)). Commonly, signaling pathways begin when a plasma membrane-bound receptor such as a receptor tyrosine kinase is activated by binding a ligand (the signal from the environment). The activated receptor phosphorylates itself (i.e., autophosphorylation), which creates a binding site to

recruit and activate other kinases in the cell cytoplasm. Then, these kinases phosphorylate other kinases that in turn phosphorylate translation factors to alter the translation of specific mRNAs, or they translocate to the nucleus and phosphorylate transcription factors to alter the transcription of specific genes. The transfer of information from kinase to kinase serves to amplify the strength of the initial signal. Thus, in response to signals, sequential cascades of post-translational modifications transmit information from one place to another in a cell and often culminate in the regulation of gene expression.

Signal-dependent regulation of gene expression



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 9-24 Extracellular signals alter gene expression via signaling pathways that use post-translational modifications to alter the function of translation and transcription factors. Signaling begins with a receptor-ligand interaction at the cell surface, proceeds through a kinase cascade in the cell cytoplasm, and culminates in the phosphorylation of translation factors in the cytoplasm or transcription factors in the nucleus that alter gene expression. Colored circles indicate different proteins whose function is labeled, and circled P's indicate a phosphorylated amino acid.

Ubiquitination

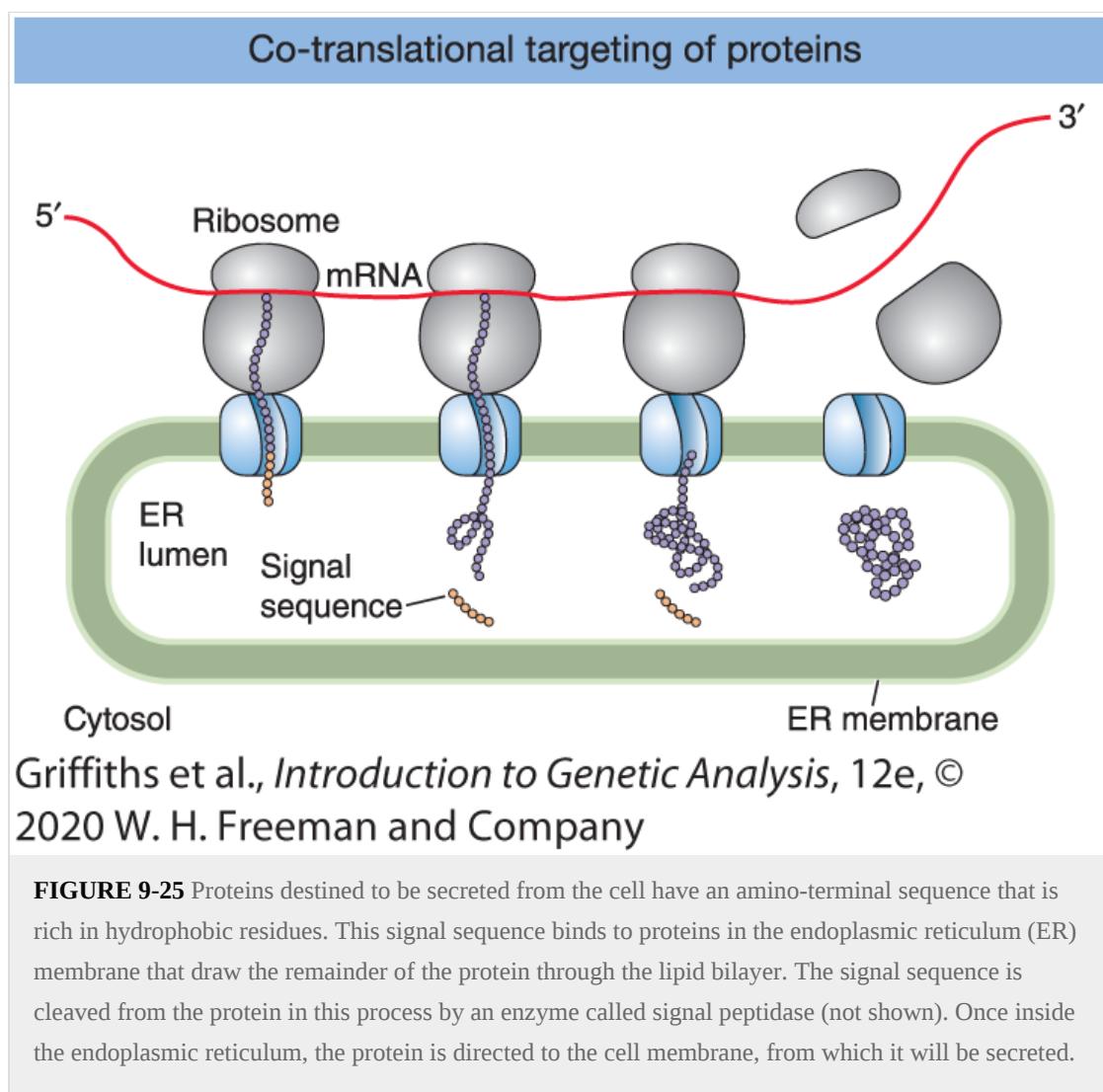
Amino acids can also be modified by small polypeptides (see [Figure 9-23](#)). For example, the ε-amine of lysines in proteins can be linked to a glycine in the 76-amino-acid polypeptide called **ubiquitin** in a process called **ubiquitination**. This addition of ubiquitin targets proteins for decay by a multiprotein protease called the **proteasome**. Two broad classes of proteins are targeted for destruction by ubiquitination: short-lived proteins such as cell cycle regulators, and proteins that have become damaged or mutated. Ubiquitin can also be covalently linked to itself in many different conformations to form polyubiquitin chains. Both monoubiquitination and polyubiquitination target proteins for decay by the proteasome. Monoubiquitination can also localize proteins to specific cellular compartments and regulate the formation of protein complexes.

KEY CONCEPT Post-translational modification of amino acids alters the structure of proteins with consequent effects on protein activity, interactions, localization, and stability. Phosphorylation is a common regulatory mechanism used in signaling cascades. Ubiquitination is often used to target proteins for degradation.

Protein targeting

In eukaryotes, all proteins are synthesized on ribosomes in the cytoplasm. However, some of these proteins end up in the nucleus, others in the mitochondria, and still others anchored in a membrane or secreted from the cell. How do these proteins “know” where they are supposed to go? The answer to this seemingly complex problem is actually quite simple: a newly synthesized protein contains a short sequence that targets the protein to the correct place or cellular compartment. For example, a newly synthesized membrane protein or a protein destined for an organelle has a short leader peptide, called a **signal sequence**, at its amino-terminal end. For

membrane proteins, this stretch of 15 to 25 amino acids directs the protein to channels in the endoplasmic reticulum (ER) membrane where the signal sequence is cleaved by a peptidase ([Figure 9-25](#)). From the ER, the protein is directed to its ultimate destination. A similar phenomenon exists for certain bacterial proteins that are secreted.



Proteins destined for the nucleus include RNA and DNA polymerases and transcription factors. Amino acid sequences embedded in the interior of such nucleus-bound proteins are necessary for transport from the cytoplasm into the nucleus. These [**nuclear localization sequences \(NLSs\)**](#) are recognized by cytoplasmic receptor proteins that transport newly synthesized proteins through nuclear pores—sites in the membrane through which large molecules are able to pass into and out of the nucleus. A protein not normally found in the nucleus will be directed to the nucleus if an NLS is artificially attached to it.

Why are signal sequences cleaved during targeting, whereas an NLS, located in a protein's interior, remains after the protein moves into the nucleus? One explanation might be that, in the

nuclear dissolution that accompanies mitosis (see [Chapter 2](#)), proteins localized to the nucleus may find themselves in the cytoplasm. Because such a protein contains an NLS, it can relocate to the nucleus of a daughter cell that results from mitosis.

KEY CONCEPT Signal sequences target proteins co-translationally or post-translationally to the inner space of organelles, to organelle or plasma membranes, or to the exterior of a cell.

SUMMARY

This chapter has described translation, the process by which the nucleotide sequence of an mRNA is converted into the amino acid sequence of a protein. Translation is the last step in the transfer of information from DNA to proteins. It occurs in three phases—initiation, elongation, and termination—that are regulated by distinct factors—IFs, EFs, and RFs, respectively. In addition, initiation, elongation, and termination involve conformational changes driven by the hydrolysis of GTP, and elongation and termination involve protein factors that mimic the function of tRNAs. Some parts of the initiation mechanism are similar between bacteria and eukaryotes, and others are different because bacterial translation occurs co-transcriptionally, whereas eukaryotic transcription and translation are physically separated into the nucleus and cytoplasm, respectively. In addition, eukaryotic mRNAs have 5' caps and 3' poly(A) tails that play critical roles in translation. On the other hand, translation elongation and termination mechanisms are largely conserved between bacteria and eukaryotes.

Ribosomes, tRNAs charged with amino acids, and mRNAs are central players in translation. Ribosomes are complex structures made up of small and large subunits, each containing rRNAs and proteins. rRNAs occupy important sites in ribosomes such as the decoding center, which has E, A, and P binding sites for tRNAs, and the peptidyltransferase center, where peptide bonds are formed between amino acids. The accuracy of translation depends on the enzymatic linkage of an amino acid with its cognate tRNA, generating a charged tRNA molecule. As adaptors, tRNAs decode the information in mRNA by base pairing between their anticodons and an mRNA codon, a triplet of nucleotides. Some tRNAs recognize more than one codon by forming both Watson-Crick and wobble base pairs. RNA base pairing also occurs between the mRNA Shine-Dalgarno sequence and the 3' end of 16S rRNA to position bacterial ribosomes at translation initiation codons. Translation begins at a fixed starting point (an AUG initiation codon that codes for methionine) and continues uninterrupted to the end of the open reading frame (an in-frame stop codon that does not code for an amino acid).

Just because an mRNA is present in a cell does not mean that it is translated. Translation is regulated, and this regulation largely occurs at the initiation and elongation phases. Some signals globally affect translation by altering the activity of initiation factors that act on most mRNAs, while other signals affect the translation of particular mRNAs via factors that bind specific sequences or structures in mRNA untranslated regions and influence multiple steps in translation initiation and elongation.

The genetic code that deciphers mRNA sequence into amino acids is almost universal among organisms. It contains 64 triplet codons, 61 that code for amino acids and 3 that code for stop codons. The code is degenerate, nonoverlapping, continuous, and organized in a non-random pattern. The organization minimizes effects of nonsynonymous mutations on the functionality of proteins by having codons that differ by one nucleotide code for amino acids with similar polarity, charge, and size.

Proteins are polymers of amino acids linked together by peptide bonds. The 20 common amino acids all have an amino group, a carboxyl group, and a unique side chain. Differences in the polarity, charge, and size of amino acid side chains affect the structure and function of proteins. With the assistance of chaperones, proteins fold into secondary structures such as α -helices and β -sheets as well as tertiary structures. Proteins also assemble into stable multiprotein complexes called quaternary structures. Proteins are the enzymes responsible for cell metabolism, including DNA and RNA synthesis; they are regulatory factors required for expression of the genetic program, and they are structural factors that confer stiffness to otherwise flexible biological machines. Lastly, proteins contain short sequence elements that act as zip codes to target them to distinct places in cells such as inside organelles, organelle or cell membranes, or the exterior of a cell.

An immense variety of amino acid modifications affect the structure, function, and localization of proteins. Many modifications are reversible, so they serve as on/off switches that control molecular processes such as transcription and translation in response to signals. Modifications fall into five broad categories: cleavage of the peptide backbone, modifications of amino acids, or the addition of complex molecules, polypeptides, or chemical groups to amino acids. The process of post-translational modifications involves enzymes that add modifications (writer), enzymes that remove modifications (erasers), and proteins that interpret modifications (readers). In summary, the versatility of proteins as biological molecules is manifested in the diversity of shapes that they can assume.

KEY TERMS

[active site](#)

[amino acid](#)

[aminoacyl tRNA-binding site \(A site\)](#)

[aminoacyl-tRNA synthetase](#)

[amino group](#)

anticodon
carboxyl group
chaperone
charged tRNA
codon
decoding center
degenerate code
domain
elongation factor (EF)
exit site (E site)
fibrous protein
frameshift mutation
genetic code
globular protein
initiation codon
initiation factor (IF)
initiator tRNA
kinase
Kozak sequence
large ribosomal subunit
nonsense codon
nuclear localization sequence (NLS)
peptide bond
peptidyl site (P site)
peptidyltransferase center
phosphatase
phosphorylation
polypeptide
primary structure (of a protein)
proteasome
quaternary structure (of a protein)
reactive group (R group), side chain
reading frame
release factor (RF)
ribosomal RNA (rRNA)
ribosome
ribosome-binding site (RBS)
secondary structure (of a protein)

Shine–Dalgarno sequence
signal sequence
small ribosomal subunit
stop codon (termination codon)
subunit
synonymous codon
termination factor
tertiary structure (of a protein)
transfer RNA (tRNA)
triplet
ubiquitin
ubiquitination
wobble

SOLVED PROBLEMS

SOLVED PROBLEM 1

Using [Figure 9-8](#), show the consequences on subsequent translation of the addition of an adenine nucleotide to the beginning of the following coding sequence:



-CGA-UCG-GAA-CCA-CGU-GAU-AAG-CAU-
- Arg - Ser - Glu - Pro - Arg - Asp - Lys - His -

SOLUTION

With the addition of A at the beginning of the coding sequence, the reading frame shifts, and a different set of amino acids is specified by the sequence, as shown here (note that a set of nonsense codons is encountered, which results in chain termination):

-ACG-AUC-GGA-ACC-ACG-UGA-UAA-GCA—Thr-Ile-Gly-Thr-Thr-stop
-ACG-AUC-GGA-ACC-ACG-UGA-UAA-GCA-
- Thr-Ile-Gly-Thr-Thr-stop

SOLVED PROBLEM 2

A single nucleotide addition followed by a single nucleotide deletion approximately 20 bp apart in DNA causes a change in the protein sequence from

—His—Thr—Glu—Asp—Trp—Leu—His—Gln—Asp—
—His—Thr—Glu—Asp—Trp—Leu—His—Gln—Asp—

to

—His—Asp—Arg—Gly—Leu—Ala—Thr—Ser—Asp—
—His—Asp—Arg—Gly—Leu—Ala—Thr—Ser—Asp—

Which nucleotide has been added, and which nucleotide has been deleted? What are the original and the new mRNA sequences? (*Hint:* Consult [Figure 9-8](#).)

SOLUTION

We can draw the mRNA sequence for the original protein sequence (with the inherent ambiguities at this stage):

— His — Thr — Glu — Asp — Trp — Leu — His — Gln — Asp
 —CA^U_C—ACC—GA^A_G—GA^U_C—UGG—CUC—CA^U_C—CA^A_G—GA^U_C
 A G A A
 G G
 G
 UUA
 G

C Because the protein-sequence change given to us at the beginning of the problem begins after the first amino acid (His) owing to a single nucleotide addition, we can deduce that a Thr codon must change to an Asp codon. This change must result from the addition of a G directly before the Thr codon (indicated by a box), which shifts the reading frame, as shown here:

—CA^U_C—GAC—UGA ^A_GGA—^U_CUG—G^C U—UCA—^U_CCA^A_G—GA^U_C—
 C C A
 A A
 G G
 — His — Asp — Arg — Gly — Leu — Ala — Thr — Ser — Asp —

C Additionally, because a deletion of a nucleotide must restore the final Asp codon to the correct reading frame, an A or G must have been deleted from the end of the original next-to-last codon, as shown by the arrow. The original protein sequence permits us to draw the mRNA with a number of ambiguities.

However, the protein sequence resulting from the frameshift allows us to determine which nucleotide was in the original mRNA at most of these points of ambiguity. Nucleotides that could have appeared in the original sequence are circled. In only a few cases does the ambiguity remain.

PROBLEMS

Visit SaplingPlus for supplemental content. Problems with the  icon are available for review/grading.

Problems with the  icon have an Unpacking the Problem exercise.

WORKING WITH THE FIGURES

(The first 25 questions require inspection of text figures.)

1. In [Figure 9-1](#), circle a protein α -helix, an RNA stem, and an RNA loop.
2. In [Figure 9-2](#), for which amino acids does the single-letter abbreviation not match the first letter of the amino acid name?
3. In [Figure 9-3a](#), draw the reaction that takes place to add a third amino acid to the chain.
4. In [Figure 9-4c](#), where are hydrophobic and hydrophilic amino acids most likely located?
5. In [Figure 9-5](#), what is the sequence of the fourth codon in the nonoverlapping mechanism? 
6. In [Figure 9-6](#), what phenotype (rII^+ or rII^-) would be expected for two insertions and one deletion? 
7. In [Figure 9-7](#), what amino acids are encoded by a repeat of the sequence CCA?
8. In [Figure 9-8](#), list the amino acids that are coded for by 1, 2, 3, 4, and 6 codons. 
9. In [Figure 9-9](#), create an analogous figure for the UAA stop codon.
10. In [Figure 9-10](#), in general, what effect does a mutation of a purine to a purine or a pyrimidine to a pyrimidine in the first nucleotide of a codon have on the polarity and charge of an encoded amino acid?
11. In [Figure 9-11](#), draw the secondary structure of tRNA^{Trp}, include the sequence of the anticodon and label the 5' and 3' ends as well as the location of amino acid attachment.
12. In [Figure 9-12](#), draw 5'-aminoacyl-AMP for proline.

13. In [Figure 9-13](#) and [Table 9-1](#), what codons are recognized by tRNA^{Val} with an anticodon sequence 5'-UAC-3'?
14. In [Figure 9-14](#), in eukaryotes, what RNA polymerases are required to transcribe rRNAs and ribosomal protein genes?
15. In [Figure 9-15](#), how many nucleotides are commonly found in the loop region of stem-loops?
16. In [Figure 9-16](#), which subunit of the ribosome binds the mRNA, and which subunit carries out peptide bond formation? 
17. Using [Figure 9-17](#), circle the Shine–Dalgarno sequence in [Figure 9-15](#).
18. In [Figure 9-18](#), describe three mechanistic differences between translation initiation in bacteria and eukaryotes.
19. In [Figure 9-19](#), draw the next step in elongation.
20. In [Figure 9-20](#), is RF1 a tRNA or a protein?
21. In [Figure 9-21](#), what mutation in tRNA^{Tyr} would suppress a UAA nonsense mutation?
22. In [Figure 9-22](#), why do miRNAs and RNA-BPs affect only the translation of specific mRNAs?
23. In [Figure 9-23](#), draw phosphotyrosine.
24. In [Figure 9-24](#), how might this pathway get turned off?
25. In [Figure 9-25](#), how does this diagram provide insight into the picture of ribosomes shown on the first page of the chapter?

BASIC PROBLEMS

26.

- a. Use the codon table in [Figure 9-8](#) to complete the following table. Assume that reading is from left to right and that the columns represent transcriptional and translational alignments.

| | | | | | | | | | |
|---|---|-----|---|---|---|---|---|---|---------------------------------------|
| | | | T | G | A | | | | DNA double helix |
| C | | | | | | | | | |
| | C | A | | U | | | | | mRNA transcribed |
| | | | | | | G | C | A | Appropriate tRNA anticodon |
| | | | | | | | | | |
| | | Trp | | | | | | | Amino acids incorporated into protein |

b. Label the 5' and 3' ends of DNA and RNA, as well as the amino and carboxyl ends of the protein.



27. Consider the following segment of DNA:

5' GCTTCCCAA 3' 3' CGAAGGGTT 5' 5' GCTTCCCAA 3'
3' CGAAGGGTT 5'

Assume that the bottom strand is the template strand used by RNA polymerase.

- a. Draw the RNA transcribed.
 - b. Label its 5' and 3' ends.
 - c. Draw the corresponding amino acid chain, assuming that the reading frame starts at the first nucleotide.
 - d. Label its amino and carboxyl ends.

Repeat parts *a* through *d*, assuming that the top strand is the template strand.

28. A mutational event inserts an extra base pair into DNA. Which of the following outcomes do you expect? (1) No protein at all; (2) a protein in which one amino acid is changed; (3) a protein in which three amino acids are changed; (4) a protein in which two amino acids are changed; (5) a protein in which most amino acids after the site of the insertion are changed.

29.

- a. In how many cases in the genetic code would you fail to know the amino acid specified by a codon if you knew only the first two nucleotides of the codon?
 - b. In how many cases would you fail to know the first two nucleotides of the codon if you knew which amino acid is specified by it?
30. If a polyribonucleotide contains equal amounts of randomly positioned adenine and uracil bases, what proportion of its triplets will encode (a) phenylalanine, (b) isoleucine, (c) leucine, (d) tyrosine? 
31. In the fungus *Neurospora*, some mutants were obtained that lacked activity for a certain enzyme. The mutations were found, by mapping, to be in either of two unlinked genes. Provide a possible explanation in reference to quaternary protein structure.
 32. What is meant by the statement “The genetic code is universal”? What is the significance of this finding?
 33. A mutant has no activity for the enzyme isocitrate lyase. Does this result prove that the mutation is in the gene encoding isocitrate lyase? Why? 
 34. A certain nonsense suppressor corrects a non-growing mutant to a state that is near, but not exactly, wild type (it has abnormal growth). Suggest a possible reason why the reversion is not a full correction.
 35. In bacterial genes, as soon as a partial mRNA transcript is produced by RNA polymerase, the ribosome assembles on it and starts translating. Draw a diagram of this process, identifying 5' and 3' ends of mRNA, the amino and carboxyl ends of the protein, the RNA polymerase, and at least one ribosome. Why couldn't this system work in eukaryotes?
 36. Researchers have found that aspartic acid and glutamic acid can sometimes mimic the function of phosphoserine and phosphothreonine. Why might this be?
 37. Why might a mutation in the untranslated region of a bacterial mRNA affect translation? How about for a eukaryotic mRNA?
 38. In vitro translation systems have been developed in which a specific mRNA can be added to a test tube containing a bacterial cell extract that includes all the components needed for translation (ribosomes, tRNAs, and amino acids). If a radioactively labeled amino acid is included, any protein translated from that mRNA can be detected on a gel. If a eukaryotic mRNA is added to the in vitro system, would radioactive protein be produced? Explain why or why not.
 39. An in vitro translation system contains a eukaryotic cell extract that includes all the components needed for translation (ribosomes, tRNAs, and amino acids). If bacterial RNA is added to the test tube, would a protein be produced? Explain why or why not.

40. Would a chimeric translation system containing the large ribosomal subunit from *E. coli* and the small ribosomal subunit from yeast (a unicellular eukaryote) be able to function in protein synthesis? Explain why or why not.
41. Mutations that change a single amino acid in the active site of an enzyme can result in the synthesis of wild-type amounts of an inactive enzyme. In what other regions of a protein might a single amino acid change have the same result?
42. What evidence supports the view that ribosomes are ribozymes? 
43. Explain why antibiotics, such as erythromycin and Zithromax, that bind the large ribosomal subunit do not harm us.
44. Our immune system makes many different proteins that protect us from viral and bacterial infection. Biotechnology companies must produce large quantities of these immune proteins for human testing and eventual sale to the public. To this end, their scientists engineer bacterial or human cell cultures to express these immune proteins. Explain why proteins isolated from bacterial cultures are often inactive, whereas the same proteins isolated from human cell cultures are active (functional).
45. Would you expect to find nuclear localization sequences (NLSSs) in the proteins that make up bacterial and eukaryotic DNA and RNA polymerases? Explain why or why not.

CHALLENGING PROBLEMS

46. Draw the structure and hydrogen bonding of a parallel β -sheet.
47. How were synthetic RNAs such as poly(U) that lacked a Shine–Dalgarno sequence translated in an *E. coli* extract?
48. A single nucleotide addition and a single nucleotide deletion approximately 15 bases apart in the DNA cause a protein change in sequence from
Phe–Ser–Pro–Arg–Leu–Asn–Ala–Val–Lys
Phe–Ser–Pro–Arg–Leu–Asn–Ala–Val–Lys
- to
- Phe–Val–His–Ala–Leu–Met–Ala–Val–Lys**
Phe–Val–His–Ala–Leu–Met–Ala–Val–Lys
- a. What are the old and new mRNA nucleotide sequences? Use the codon table in [Figure 9-8](#).
- b. Which nucleotide has been added? Which has been deleted? 
49. You are studying an *E. coli* gene that specifies a protein. A part of its sequence is
–Ala–Pro–Trp–Ser–Glu–Lys–Cys–His–**–Ala–Pro–Trp–Ser–Glu–Lys–Cys–His–**

You recover a series of mutants for this gene that show no enzymatic activity. By isolating the mutant enzyme products, you find the following sequences:

Mutant 1:

-Ala-Pro-Trp-Arg-Glu-Lys-Cys-His-**-Ala-Pro-Trp-Arg-Glu-Lys-Cys-His-**

Mutant 2:

-Ala-Pro-**-Ala-Pro-**

Mutant 3:

-Ala-Pro-Gly-Val-Lys-Asn-Cys-His-**-Ala-Pro-Gly-Val-Lys-Asn-Cys-His-**

Mutant 4:

-Ala-Pro-Trp-Phe-Phe-Thr-Cys-His-**-Ala-Pro-Trp-Phe-Phe-Thr-Cys-His-**

What is the molecular basis for each mutation? What is the DNA sequence that specifies this part of the protein? 

50. What structural features are shared by spliceosomes (see [Figure 8-19](#)) and ribosomes? Why are both structures used to support the RNA World theory?
51. A double-stranded DNA molecule with the sequence shown here produces, *in vivo*, a polypeptide that is five amino acids long.

TACATGATCATTCA CGGA ATTCTAGCATGTA
ATGTACTAGTAAAGTGCCTAAAGATCGTACAT
TACATGATCATTCA CGGA ATTCTAGCATGTA
ATGTACTAGTAAAGTGCCTAAAGATCGTACAT

- a. Which strand of DNA is the template strand, and in which direction is it transcribed?
 - b. Label the 5' and 3' end of each strand.
 - c. If an inversion occurs between the second and the third triplets from the left and right ends, respectively, and the same strand of DNA is transcribed, how long will the resultant polypeptide be?
 - d. Assume that the original molecule is intact and that the bottom strand is transcribed from left to right. Give the RNA base sequence and label the 5' and 3' ends of the anticodon that inserts the *fourth* amino acid into the nascent polypeptide. What is this amino acid? 
52. One of the techniques Khorana used to decipher the genetic code was to synthesize polypeptides *in vitro*, using synthetic mRNA with various repeating base sequences. For example, $(AGA)_n$, which can be written out as AGAAGAAGAAGAAGA.... Sometimes the resulting polypeptide contained

just one amino acid (a homopolymer), and sometimes it contained more than one amino acid (a heteropolymer), depending on the repeating sequence used. Khorana found that sometimes different polypeptides were made from the same synthetic mRNA, suggesting that the initiation of protein synthesis in the system in vitro does not always start at the first nucleotide of the messenger. For example, from $(CAA)_n$, three polypeptides may have been made: aa₁ homopolymer (abbreviated aa₁-aa₁), aa₂ homopolymer (aa₂-aa₂), and aa₃ homopolymer (aa₃-aa₃). These polypeptides probably correspond to the following readings derived by starting at different places in the sequence:

CAA CAA CAA CAA...ACA ACA ACA ACA...AAC AAC AAC AAC...

CAA CAA CAA CAA ...

ACA ACA ACA ACA ...

AAC AAC AAC AAC ...

The following table shows the results of Khorana's experiment.

| Synthetic mRNA | Polyptide(s) synthesized |
|----------------|---------------------------------------|
| $(UC)_n$ | (Ser-Leu) |
| $(UG)_n$ | (Val-Cys) |
| $(AC)_n$ | (Thr-His) |
| $(AG)_n$ | (Arg-Glu) |
| $(UUC)_n$ | (Ser-Ser) and (Leu-Leu) and (Phe-Phe) |
| $(UUG)_n$ | (Leu-Leu) and (Val-Val) and (Cys-Cys) |
| $(AAG)_n$ | (Arg-Arg) and (Lys-Lys) and (Glu-Glu) |
| $(CAA)_n$ | (Thr-Thr) and (Asn-Asn) and (Gln-Gln) |
| $(UAC)_n$ | (Thr-Thr) and (Leu-Leu) and (Tyr-Tyr) |
| $(AUC)_n$ | (Ile-Ile) and (Ser-Ser) and (His-His) |
| $(GUA)_n$ | (Ser-Ser) and (Val-Val) |
| $(GAU)_n$ | (Asp-Asp) and (Met-Met) |
| $(UAUC)_n$ | (Tyr-Leu-Ser-Ile) |
| $(UUAC)_n$ | (Leu-Leu-Thr-Tyr) |
| $(GAUA)_n$ | None |
| $(GUAA)_n$ | None |

Note: The order in which the polypeptides or amino acids are listed in the table is not significant except for $(UAUC)_n$ and $(UACU)_n$

- Why do $(GUA)_n$ and $(GAU)_n$ each encode only two homopolypeptides?
- Why do $(GAUA)_n$ and $(GUAA)_n$ fail to stimulate synthesis?

- c. Using Khorana's results, assign an amino acid to each triplet in the following list. Remember that there are often several codons for a single amino acid and that the first two letters in a codon are usually the important ones (but that the third letter is occasionally significant). Also keep in mind that some very different-looking codons sometimes encode the same amino acid. Try to solve this problem without consulting [Figure 9-8](#).

| | | | |
|-----|-----|-----|-----|
| GUA | GAU | UUG | AAC |
| GUG | UUC | UUA | GAA |
| GUU | AGU | UAU | AGA |
| AUG | CUU | AUC | GAG |
| UGU | CUA | UAC | CAA |
| ACA | UCU | AAG | UAG |
| CAC | CUC | ACU | UGA |

Solving this problem requires both logic and trial and error. Don't be disheartened: Khorana received a Nobel Prize for doing it. Good luck!

(Data from J. Kuspira and G. W. Walker, *Genetics: Questions and Problems*. McGraw-Hill, 1973.)

GENETICS AND SOCIETY

If life were found on another planet, do you think that it would have the same genetic code? Justify your answer.