11

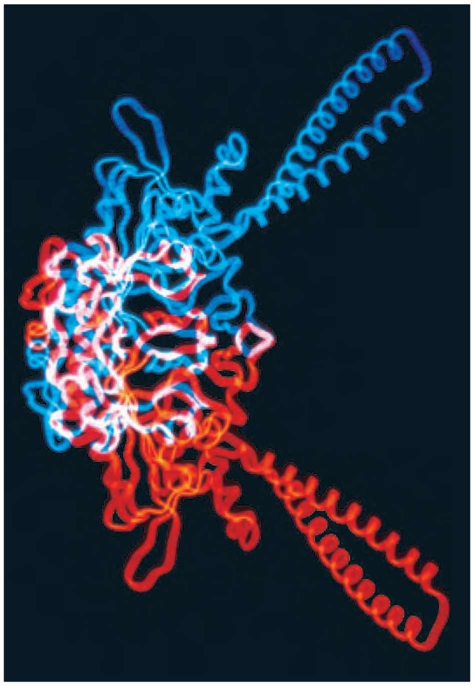
GENE EXPRESSION

Translation

STUDY OBJECTIVES

1. To study the mechanism of protein biosynthesis, in which organisms, using the information in DNA, string together amino acids to form proteins 281
2. To examine the genetic code 304

STUDY OUTLINE



Computer-generated model of the enzyme seryl tRNA synthetase, the enzyme that charges tRNA's with the amino acid serine. (© Dr. Stephen Cusack/EMBL/SPL/Photo Researchers, Inc.)

In this chapter, we continue our discussion of gene expression, concentrating on protein biosynthesis. This process, which translates the nucleotide information in messenger RNA into amino acid sequences in proteins, is the final step of the central dogma. Nucleotide sequences in DNA are transcribed into nucleotide sequences in RNA, which are then translated into amino acid sequences in proteins.

All proteins are synthesized from only twenty naturally occurring amino acids (fig. 11.1). (There is one exception, selenocysteine, which we discuss at the end of the chapter.) These are called α-amino acids because one carbon, the α carbon, has four specific groups attached to it: an amino group, a carboxyl (acidic) group, a hydrogen, and one of the twenty different R groups (side chains), imparting the specific properties of that amino acid. (Technically, proline is termed an *imino* acid because of its structure.) Having these four groups attached imparts a property known as *chirality* on the amino acid: like left- and right-handed gloves, the mirror images cannot be superimposed. Because of optical properties, the two forms of each amino acid are referred to as D and L, in which D comes from dextrorotatory (right turning) and L comes from levorotatory (left turning). All biologically active amino acids are of the L form, and hence we need not refer to this designation. Proteins (polypeptides) are synthesized when peptide bonds form between any two amino acids (fig. 11.2). In this manner, long chains of amino acids—called *residues* when incorporated into a protein—can join, and all chains will have an amino (N-terminal) end and a carboxyl (C-terminal) end.

The sequence of polymerized amino acids determines the **primary structure** of a protein. Included in the primary structure is the formation of disulfide bridges between cysteine residues (fig. 11.3). Polypeptides can fold into several structures, the most common of which are α helices and β sheets. These folding configurations constitute the **secondary structure** of the protein. In some proteins, the folding is spontaneous; in some, it is guided by other proteins. Further folding, bringing α helices and β sheets into three-dimensional configurations, creates the **tertiary structure** of the protein (fig. 11.4). Many proteins in the active state are composed of several subunits that together make up the **quaternary structure** of the protein. Translation is the process in which the primary structure of a protein is determined from the nucleotide sequence in a messenger RNA (box 11.1).

INFORMATION TRANSFER

Before proceeding to the details of translation, a sketch of the beginning of the process may be helpful (fig. 11.5). The ribosome with its ribosomal RNA and proteins is the site of protein synthesis. The information from the gene is in the form of messenger RNA, in which each group of three nucleotides—a codon—specifies an amino acid. The amino acids are carried to the ribosome attached to transfer RNAs, and these transfer RNAs have anticodons, three nucleotides complementary to a codon, located at the end opposite the amino acid attachment site. A peptide bond will form between the two amino acids present at the ribosome, freeing one transfer RNA (at codon 1 in fig. 11.5) and lengthening the amino acid chain attached to the second transfer RNA (at codon 2 in fig. 11.5). The messenger RNA will then move one codon with respect to the ribosome, and a new transfer RNA will attach at codon 3. This cycle is then repeated, with the polypeptide lengthening by one amino acid each time. We can begin looking at the details of translation by looking at the transfer RNAs. As before, we concentrate on the prokaryotic system, noting details about eukaryotes as appropriate.

Transfer RNA

Attachment of Amino Acid to Transfer RNA

The function of transfer RNA is to ensure that each amino acid incorporated into a protein corresponds to a particular codon (a group of three consecutive nucleotides) in the messenger RNA. The transfer RNA serves this function through its structure: It has an anticodon at one end and an amino acid attachment site at the other end. The "correct" amino acid, the amino acid corresponding to the anticodon, is attached to the transfer RNA by enzymes known as aminoacyl-tRNA synthetases (e.g., arginyl-tRNA synthetase, leucyl-tRNA synthetase). A transfer RNA with an amino acid attached is said to be "charged."

An aminoacyl-tRNA synthetase joins a specific amino acid to its transfer RNA in a two-stage reaction that takes place on the surface of the enzyme. In the first stage, the amino acid is activated with ATP. In the second stage of the reaction, the amino acid is attached with a high-energy bond to the 2′ or 3′ carbon of the ribose sugar at the 3′ end of the transfer RNA (fig. 11.6). In the figure, we denote high-energy bonds, bonds that liberate a lot of free energy when hydrolyzed, as "~." Thus, during the process of protein synthesis, the energy for the formation of the peptide bond will be present where it is needed, at the point of peptide bond formation.

Component Numbers

In bacteria, there are twenty aminoacyl-tRNA synthetases, one for each amino acid. A particular enzyme recognizes a particular amino acid, as well as all the transfer RNAs that

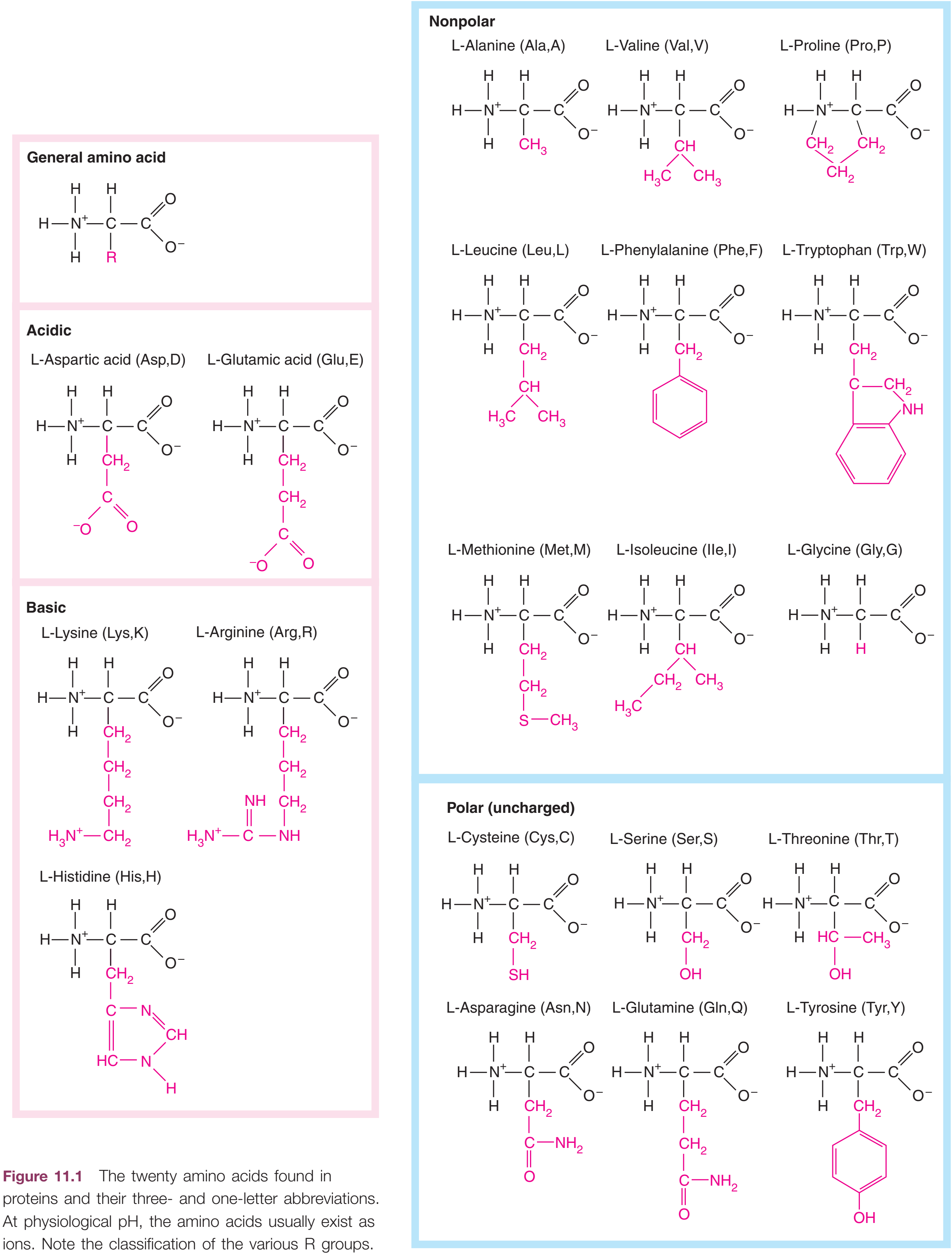


Figure 11.1 The twenty amino acids found in proteins and their three- and one-letter abbreviations. At physiological pH, the amino acids usually exist as ions. Note the classification of the various R groups.

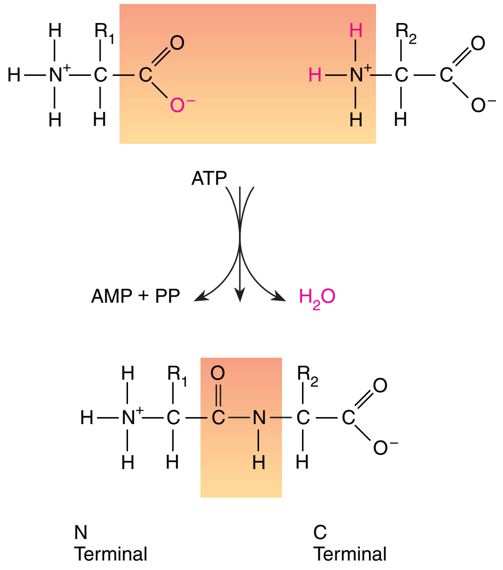


Figure 11.2 Protein synthesis: formation of a peptide bond between two amino acids. The bond is between the carboxyl group of one amino acid and the amino group of the other.

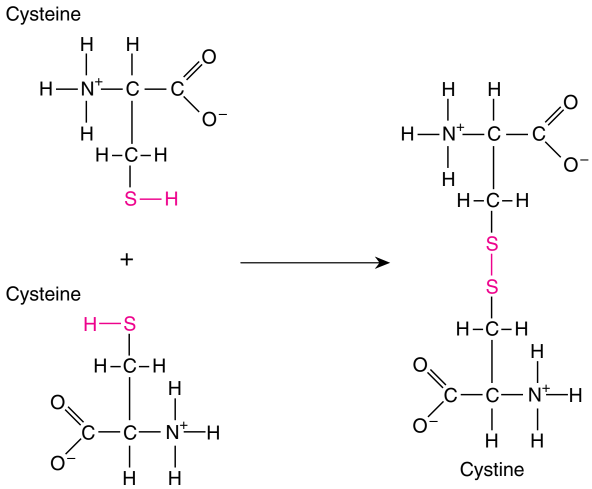


Figure 11.3 A disulfide bridge can form when two cysteines are brought into apposition. If the two amino acids are in the free form, the new structure is called *cystine.* When the two cysteines are in the same or different polypeptides, the disulfide bridge creates stability.

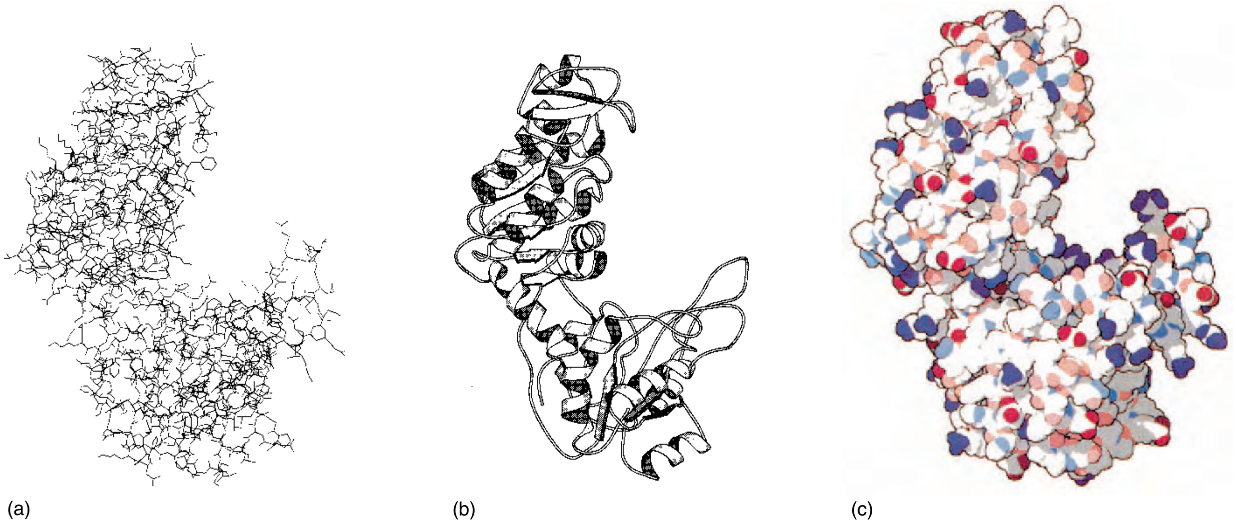


Figure 11.4 Three different ways of depicting a protein, the enzyme phosphoglycerate kinase. At *left* is a bond diagram; all the lines shown represent bonds between the various atoms of the molecule. In the *middle* is a ribbon diagram that emphasizes the secondary structure of the protein. Shown are alpha helices (spiral ribbons) and beta pleated sheets (flat arrows). Finally, on the *right* is a space-filling diagram that emphasizes the volume the molecule fills. The space-filling diagram is what the molecule would generally look like if it were magnified eight million times. (Images by David S. Goodsell, the Scripps Research Institute.)

BOX 11.1

Experimental Methods

Amino Acid Sequencing

Protein-sequencing techniques have been known since 1953, when F. Sanger worked out the complete sequence of the protein hormone insulin. The basic strategy is to purify the protein and then sequence it, beginning at one end. However, since most proteins contain too many amino acids to do this successfully, proteins are first broken into small peptides in several different ways. These peptides are sequenced, and the whole protein sequence can be determined by the overlap pattern of the sequenced subunits.

A protein can be broken into peptide fragments by many different methods, including acid and alkaline hydrolysis. For the most part, proteolytic enzymes (proteases) that hydrolyze the peptides at specific points are used. *Pepsin,* for example, preferentially hydrolyzes peptide bonds involving aromatic amino acids, methionine, and leucine; *chymotrypsin* hydrolyzes peptide bonds involving carboxyl groups of aromatic amino acids; and *trypsin* hydrolyzes bonds involving the carboxyl groups of arginine and lysine.

The proteolytic digest is usually separated into a *peptide map,* or *peptide fingerprint,* by using a two-dimensional combination of paper chromatography, electrophoresis, or column chromatography. In two-dimensional chromatography, a sample is put onto a piece of paper that is then placed in a solvent system. After an allotted time, the paper is dried, turned 90 degrees, and placed in a second solvent system for another allotted time (fig. 1). In each solvent, different peptides travel through the paper at different rates. The spots are then developed using ninhydrin, which reacts with the N-terminal amino acid and produces a colored product when heated.

The spots, which represent small peptides, can be cut out of a second, identical chromatogram that has not been sprayed with ninhydrin. These spots can then be sequenced by, for example, the Edman method, whereby the peptide is sequentially degraded

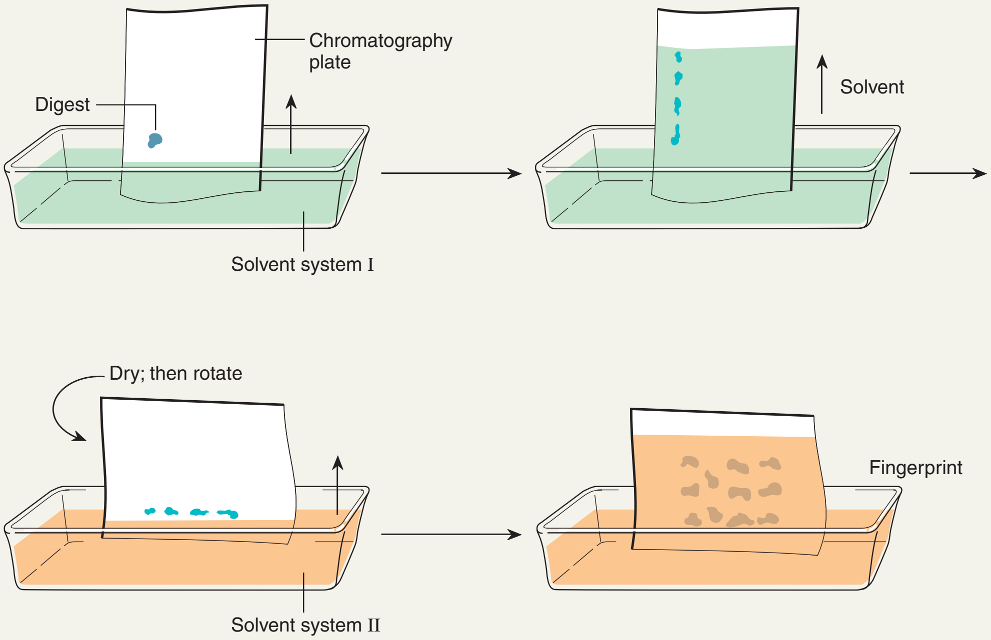


Figure 1 Two-dimensional paper chromatography of a protease digest. Chromatography is done first in one solvent system. The paper is then dried, rotated, and placed into a second solvent system. The pattern on the resulting plate is called a peptide fingerprint.

from the N-terminal end. Phenylisothiocyanate (PITC) reacts with the amino end of the peptide. When acid is added, the N-terminal amino acid is removed as a PITC derivative and can be identified. The process is then repeated until the whole peptide has been sequenced (fig. 2).

If the fingerprint pattern is worked out for two different digests of the same polypeptide, the unique sequence of the original polypeptide can be determined by overlap. In figure 3, the letters A–J represent the ten amino acids in a polypeptide. A is known to be the first (N-terminal) amino acid since the Edman method sequences peptides from this end. We can thus summarize the methodology as follows:

1. A protein is purified. If it is made up of several subunits, these subunits are separated and purified. (If disulfide bridges exist within a peptide, they must be reduced. The bridges are later determined by digestion, keeping the bridges intact, and then resequencing.)
2. Different proteolytic enzymes are used on separate subsamples so that the protein is broken into different sets of peptide fragments.
3. Two-dimensional chromatography, electrophoresis, or column chromatography can be used to isolate the peptides.
4. The Edman method of sequentially removing amino acids from the N-terminal end is used to sequence each peptide.
5. The amino acid sequence from the N- to C-terminal ends of the protein is deduced from the overlap of sequences in peptide digests generated with different proteolytic enzymes.

Today, a machine known as an amino acid sequencer (*sequenator*) can automatically sequence protein. Taking about two hours per amino acid residue, sequenators can carry out Edman degradation on polypeptides up to about fifty amino acids long.

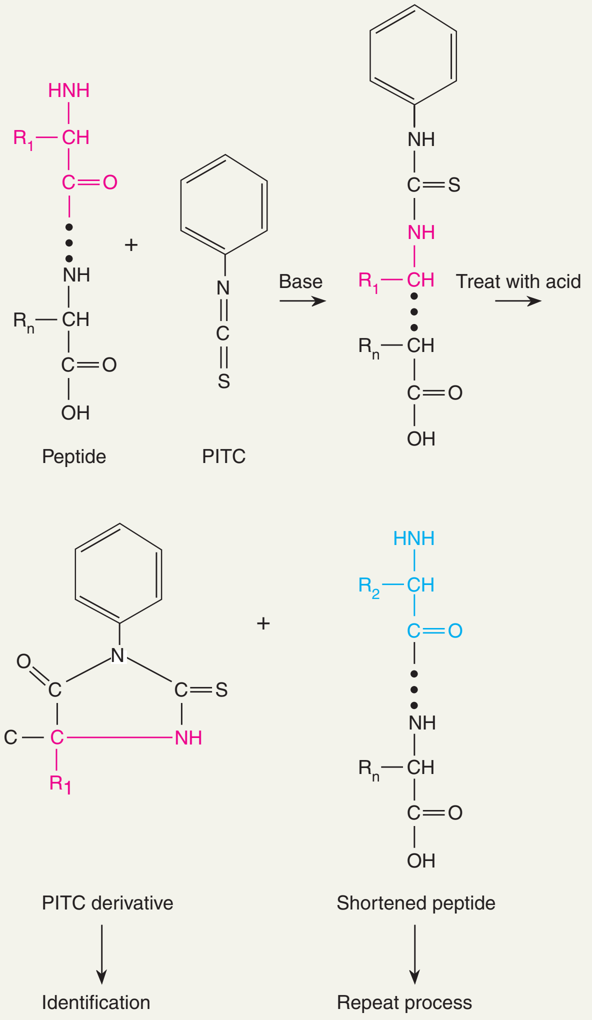


Figure 2 Isolation of amino acids from a peptide for sequencing purposes. First, the peptide reacts with PITC (phenylisothiocyanate) at the amino end. Acid treatment produces a PITC derivative of the amino-terminal amino acid and a peptide one amino acid shorter than the original. The PITC derivative can be identified. These steps are then repeated, isolating one amino acid at a time.

continued

BOX 11.1 CONTINUED

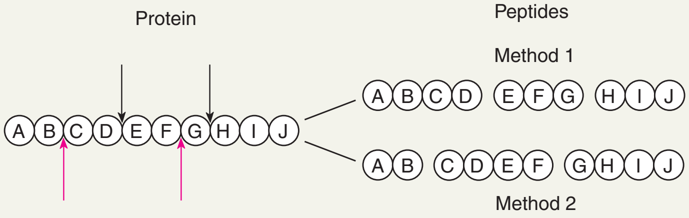


Figure 3 The overlap of peptides digested two different ways provides the sequence of the original peptide.

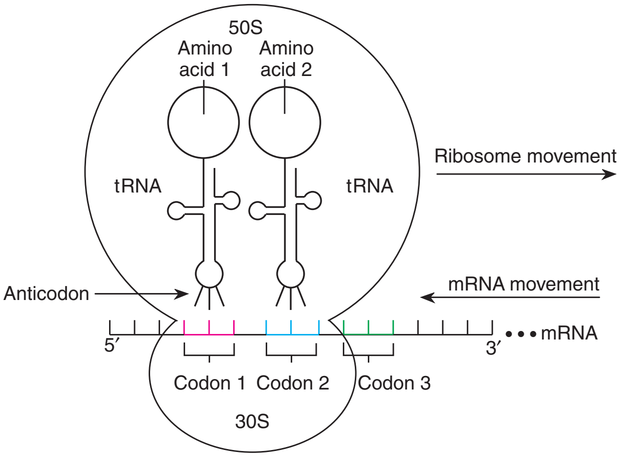


Figure 11.5 The initiation of the translation process at the ribosome. Note the two charged transfer RNAs and the messenger RNA. They are in position to form the first peptide bond between the two amino acids attached to the transfer RNAs.

code for that amino acid. In eukaryotes, there are separate sets of twenty cytoplasmic and twenty mitochondrial synthetases, all coded in the nucleus.

Aminoacyl-tRNA synthetases are a heterogeneous group of enzymes. In *E. coli,* they vary from monomeric proteins (one subunit) to tetrameric proteins, made up of two copies each of two subunits. The enzymes fall into two categories based on sequence similarity, structural features, and whether the amino acid is attached at the 2′-OH (in class I enzymes) or 3′-OH (in class II enzymes) of the 3′-terminal adenosine of the transfer RNA.

To add its appropriate amino acid to the appropriate transfer RNA, a synthetase recognizes many parts of the transfer RNA. This can be shown by experiments that introduce specific changes in transfer RNAs by site-directed mutagenesis (see chapter 13). In seventeen of the twenty *E. coli* synthetases, recognition involves part of the anticodon itself. This makes sense since the anticodon is the defining element of a transfer RNA in protein synthesis.

A synthetase can initially make errors and attach the "wrong" amino acid to a tRNA. For example, isoleucyl-tRNA synthetase will attach valine about once in 225 times. This type of error occurs because a similar, but smaller, amino acid can sometimes occupy the active site of the enzyme (compare isoleucine and valine in fig. 11.1). However, because of a proofreading step, only 1 in 270 to 1 in 800 of the errors are released intact from the enzyme. The amino acids on the rest of the incorrectly charged transfer RNAs are hydrolyzed before the transfer RNAs are released. The overall error rate is the product of the two steps; this means only about one incorrectly charged transfer RNA occurs per 60,000 to 80,000 formed.

In several cases, the number of amino acyl-tRNA synthetases in a particular organism is below twenty. For example, in some archaea, there is no cysteinyl-tRNA synthetase. However, the prolyl-tRNA synthetase activates the tRNAs for both cysteine and proline with their appropriate amino acids. Similarly, in some eubacteria, there is no glutaminyl-tRNA synthetase; the glutaminyl tRNA is charged with glutamic acid, rather than glutamine. An amido transferase enzyme then converts the glutamic acid to glutamine (see fig. 11.1).

There are sixty-four possible codons in the genetic code (four nucleotide bases in groups of three $= 4 \times 4 \times 4 = 64$). Three of these codons are used to terminate translation. Thus, sixty-one transfer RNAs are needed because there are sixty-one different nonterminator codons. About fifty transfer RNAs are known in *E. coli.* The number fifty can be explained by the wobble phenomenon, which occurs in the third position of the codon. We examine this phenomenon in the section on the genetic code. The transfer RNAs for each amino acid are designated by the convention tRNA^{Leu} (for leucine), tRNA^{His} (for histidine), and so on.

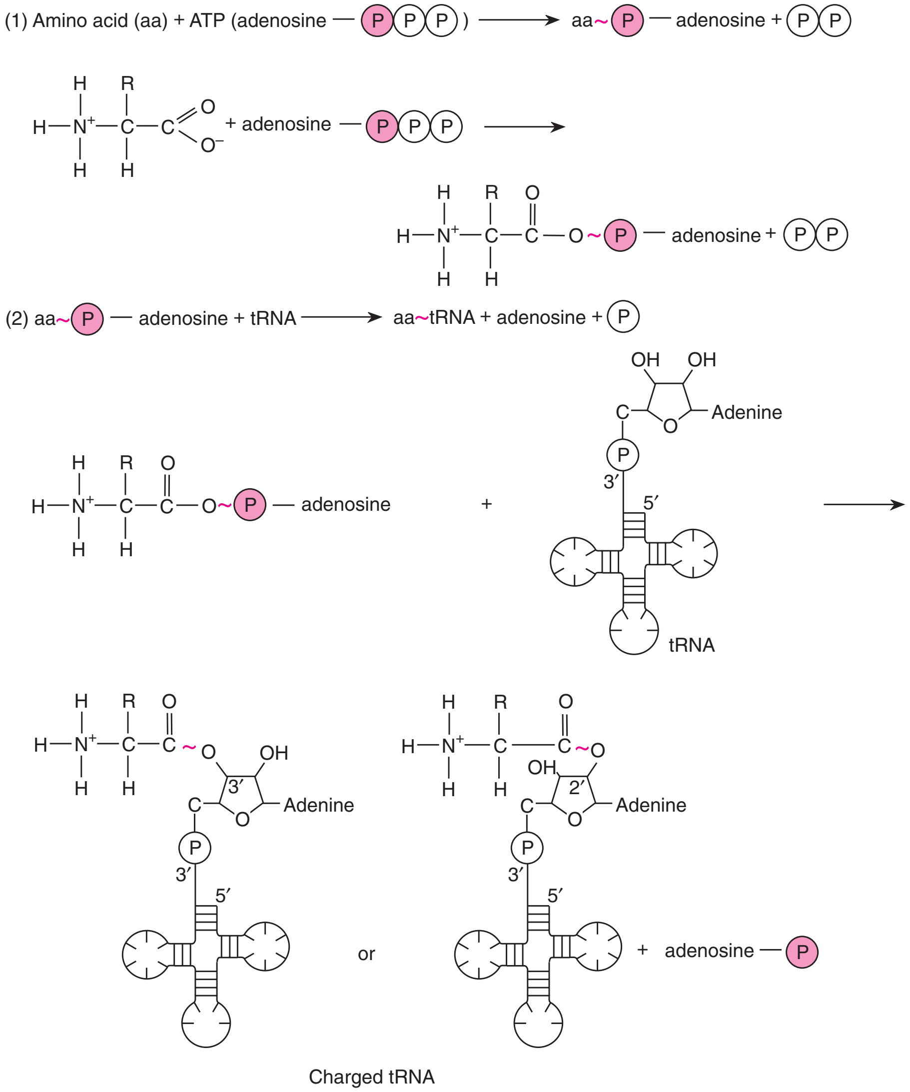


Figure 11.6 It takes a two-step process to attach a specific amino acid to its transfer RNA by an aminoacyl synthetase. High-energy bonds are indicated by ~. In the first step, an amino acid is attached to AMP with a high-energy bond. In the second step, the high-energy bond is transferred to the tRNA, which is then referred to as "charged." Depending on which class of aminoacyl-tRNA synthetase is involved, the amino acid will be attached to either the 2′ or 3′ carbon of the sugar of the 3′ terminal adenosine.

Recognition of the Aminoacyl-tRNA During Protein Synthesis

Although amino acids enter the protein-synthesizing process attached to transfer RNAs, it was theoretically possible that the ribosome recognized the amino acid itself during translation. A simple experiment was done to determine whether the amino acid or the transfer RNA was recognized.

In 1962, F. Chapeville and colleagues isolated transfer RNA with cysteine attached. They chemically converted the cysteine to alanine by using Raney nickel, a catalytic form of nickel that removes the SH group of cysteine (fig. 11.7). When these transfer RNAs were used in protein synthesis, alanine was incorporated where cysteine should have been, demonstrating that the transfer RNA, not the amino acid, was recognized during protein synthesis. The synthetase puts a specific amino acid on a specific transfer RNA; then, during protein synthesis, the anticodon on the transfer RNA—not the amino acid itself—determines which amino acid is incorporated.

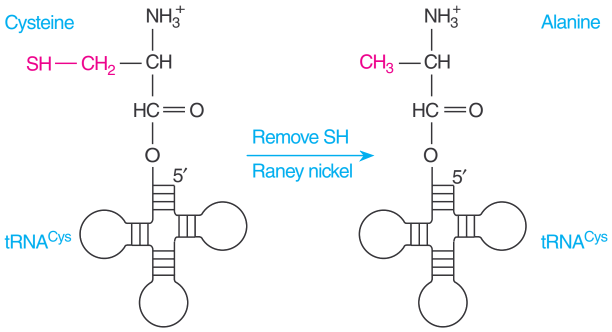


Figure 11.7 Cysteine-tRNA^{Cys} treated with Raney nickel becomes alanine-tRNA^{Cys} by the removal of the SH group of cysteine. During protein synthesis, alanine is incorporated in place of cysteine in proteins, indicating that the specificity of amino acid incorporation into proteins resides with the tRNA.

Initiation Complex

Translation can be divided into three stages: initiation, elongation, and termination. Elongation is the repetitive process of adding amino acids to a growing peptide chain. However, added complexity enters the picture in the initiation and termination of protein synthesis.

It is especially important that the translation process start precisely. Remember that the genetic code is translated in groups of three nucleotides (codons). If the reading of the messenger RNA begins one base too early or too late, the reading frame is shifted so that an entirely different set of codons is read (fig. 11.8). The protein produced, if any, will probably bear no structural or functional resemblance to the protein the gene is coded for.

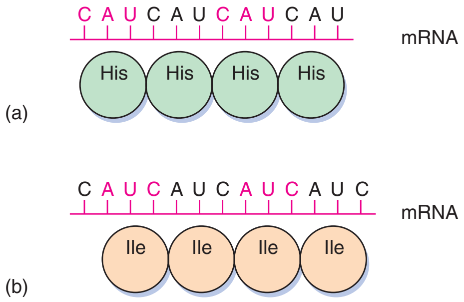


Figure 11.8 (*a*) In the normal reading of the messenger RNA, these codons are read as repeats of CAU, coding for histidine. (*b*) A shift in the reading frame of the messenger RNA causes the codons to be read as AUC repeats coding for isoleucine.

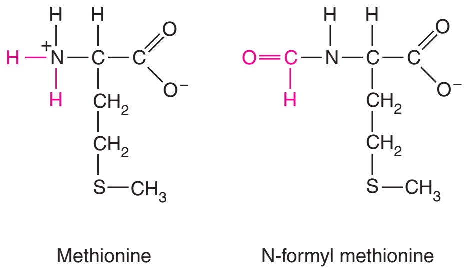


Figure 11.9 The structures of the amino acids methionine and N-formyl methionine.

Role of N-Formyl Methionine

The synthesis of every protein in *Escherichia coli* begins with the modified amino acid N-formyl methionine (fig. 11.9). However, none of the completed proteins in *E. coli* contains N-formyl methionine. Many of these proteins do not even have methionine as their first amino acid. Obviously, before a protein becomes functional, the initial amino acid is modified or removed. In eukaryotes the initial amino acid, also methionine, does not have an N-formyl group.

Methionine, with a codon of 5′-AUG-3′, known as the **initiation codon,** has two transfer RNAs with the same complementary anticodon (3′-UAC-5′) but with different structures (fig. 11.10). One of these transfer RNAs (tRNA$_f^{Met}$) serves as a part of the initiation complex. Before the initiation of translation, this transfer RNA will have its methionine chemically modified to N-formyl methionine (fMet). The other transfer RNA will not have its methionine modified (tRNA$_m^{Met}$). The translation machinery will use it to insert methionine into proteins, where called for, in all but the first position. The cell thus has a mechanism to make use of methionine in the normal way as well as to use a modified form of it to initiate protein synthesis. Because of the structure of the prokaryotic initiation transfer RNA, it can recognize AUG, GUG, and, rarely, UUG as initiation codons. In eukaryotes, CUG as well as AUG can serve as an initiation codon. Since the initiation methionine is not formylated in eukaryotes, the eukaryotic transfer RNA is designated tRNA$_i^{Met}$; there is a separate internal methionine transfer RNA, termed tRNAm$_m^{Met}$, in eukaryotes, as in prokaryotes.

Translation Initiation

The subunits of the ribosome (30S and 50S) usually dissociate from each other when not involved in translation. To begin translation, an **initiation complex** forms, consisting of the following components in prokaryotes: the 30S subunit of the ribosome, a messenger RNA, the charged N-formyl methionine tRNA (fMET-tRNA$_f^{Met}$), and

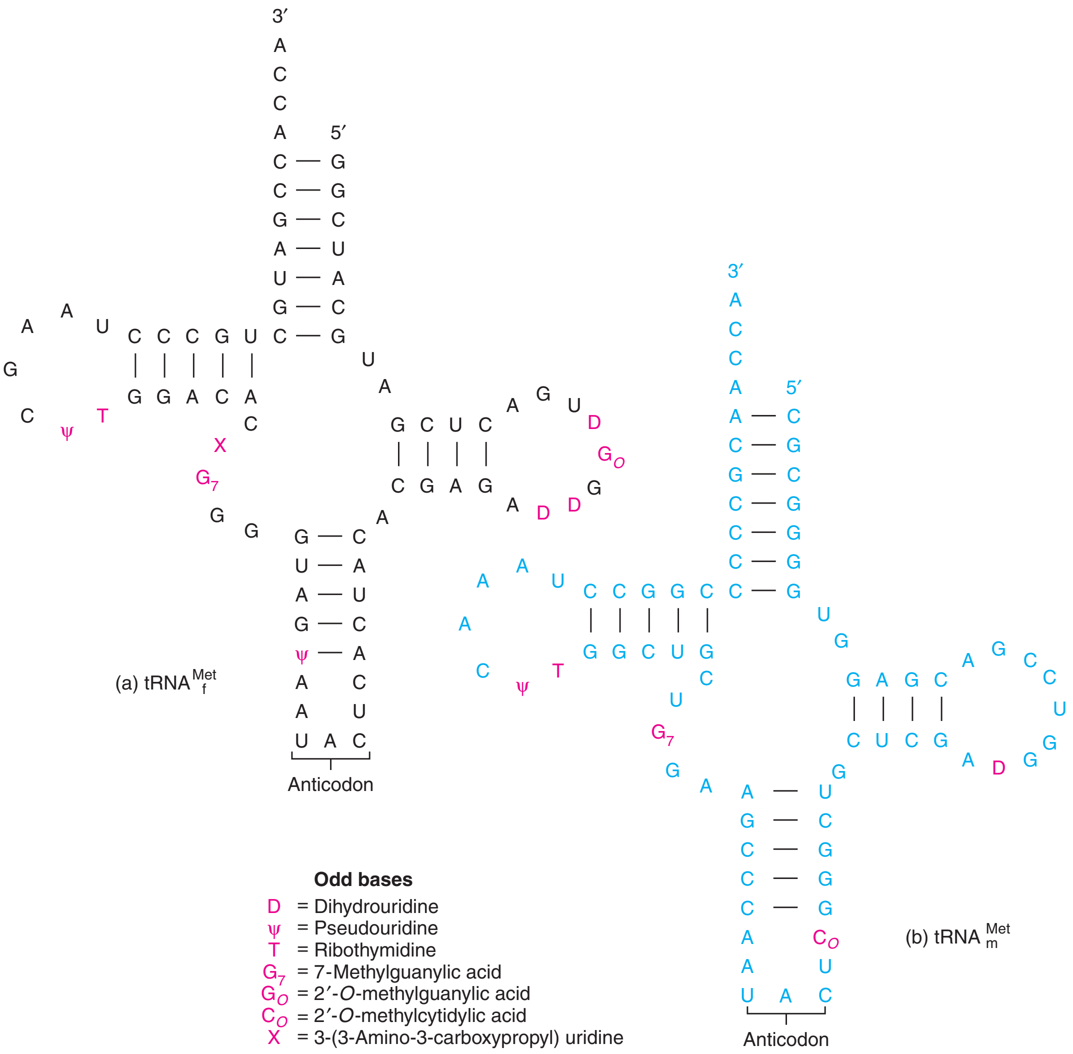


Figure 11.10 The two tRNAs for methionine in *E. coli.* (*a*) The initiator tRNA. (*b*) The interior tRNA.

three **initiation factors (IF1, IF2, IF3).** Initiation factors (as well as elongation and termination factors) are proteins loosely associated with the ribosome. They were discovered when ribosomes were isolated and then washed, losing the ability to perform protein synthesis.

The components that form the initiation complex interact in a series of steps. It is known that IF3 binds to the 30S ribosomal subunit, allowing the 30S subunit to bind to messenger RNA (fig. 11.11, *step 1*). Meanwhile, a complex forms with IF2, the charged N-formyl methionine tRNA (fMET-tRNA$_f^{Met}$) and GTP (guanosine triphosphate; fig. 11.11, *step 2*). It is IF2 that brings the initiator transfer RNA to the ribosome. IF2 binds only to the charged initiator transfer RNA, and, without IF2, the initiator transfer RNA cannot bind to the ribosome. The final step in initiation-complex formation is bringing together the first two components (fig. 11.11, *step 3*).

The hydrolysis of GTP to GDP + P_i (inorganic phosphate, PO_4^{-3}—see fig. 9.8) produces conformational changes; these changes allow the initiation complex to join the 50S ribosomal subunit to form the complete ribosome and then allow the initiation factors and GDP to be released. Frequently, the hydrolysis of a nucleoside triphosphate (e.g., ATP, GTP) in a cell occurs to release the energy in the phosphate bonds for use in a metabolic process. However, in the process of translation, the hydrolysis apparently changes the shape of the GTP so that it and the initiation factors can be released from the ribosome after the 70S particle has been formed. Thus, hydrolysis of GTP in translation is for conformational

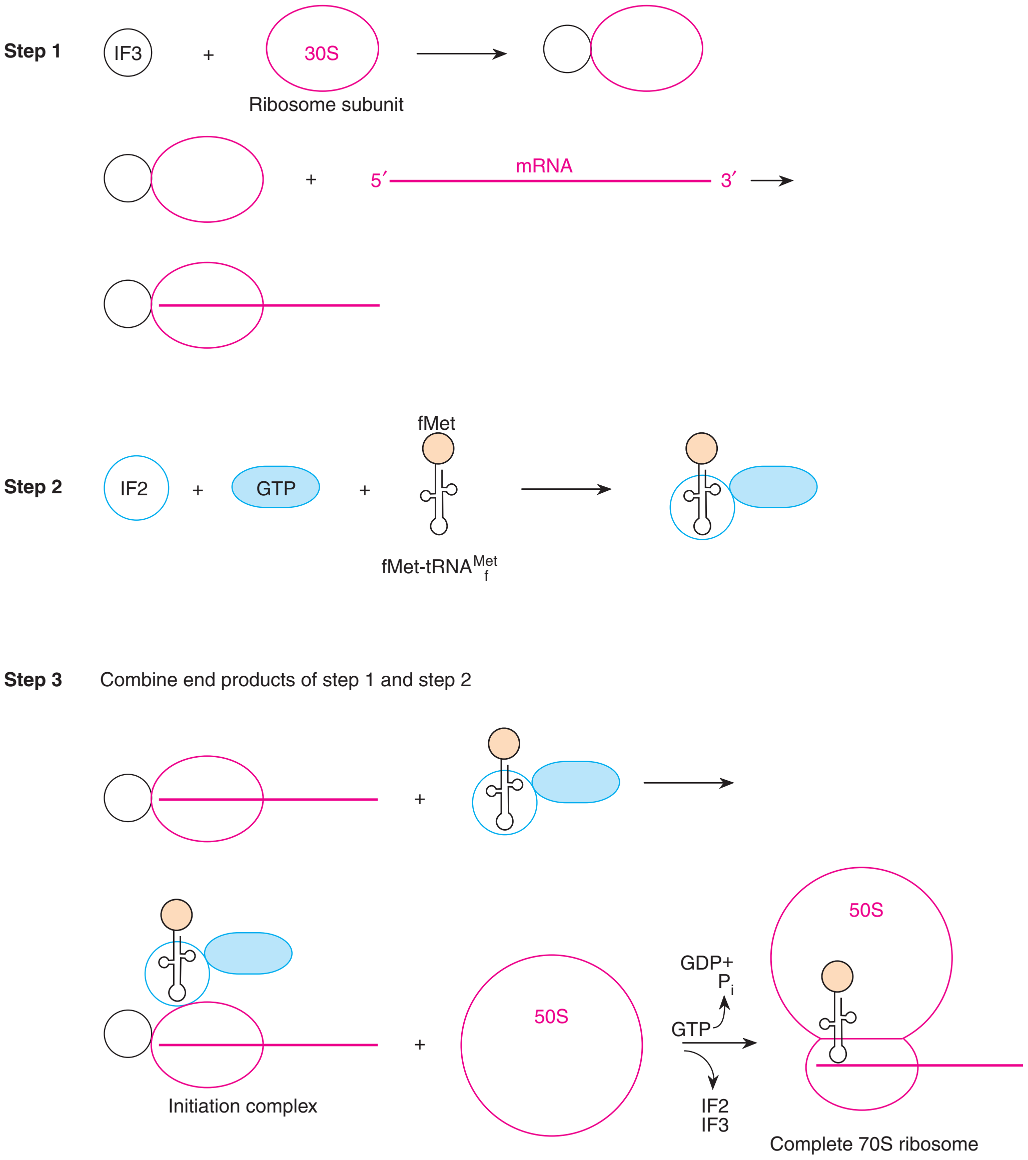


Figure 11.11 The prokaryotic 70S ribosome forms in a three-step process. In the first step, the 30S ribosomal subunit and the mRNA combine. In the second step, the initiator tRNA combines with IF2. In the final step, the components from steps 1 and 2 combine to form the initiation complex, followed by the formation of the 70S ribosome.

change rather than covalent bond formation. IF1 helps the other two initiation factors bind to the 30S ribosomal subunit or stabilizes the 30S initiation complex.

The process in eukaryotes is generally similar, but more complex. The eukaryotic initiation factor abbreviations are preceded by an "e" to denote that they are eukaryotic (eIF1, eIF2, etc.). At least eleven initiation factors are involved, including a specific cap-binding protein, eIF4E.

The ribosome apparently recognizes the prokaryotic messenger RNA through complementarity of a region at the 3′ end of the 16S ribosomal RNA and a region slightly upstream from the initiation sequence (AUG) on the messenger RNA. This idea, the **Shine-Dalgarno hypothesis** (fig. 11.12), is named after the people who first suggested it. The sequence (AGGAGGU) of complementarity between the messenger RNA and the 16S ribosomal RNA is referred to as the Shine-Dalgarno sequence. Although

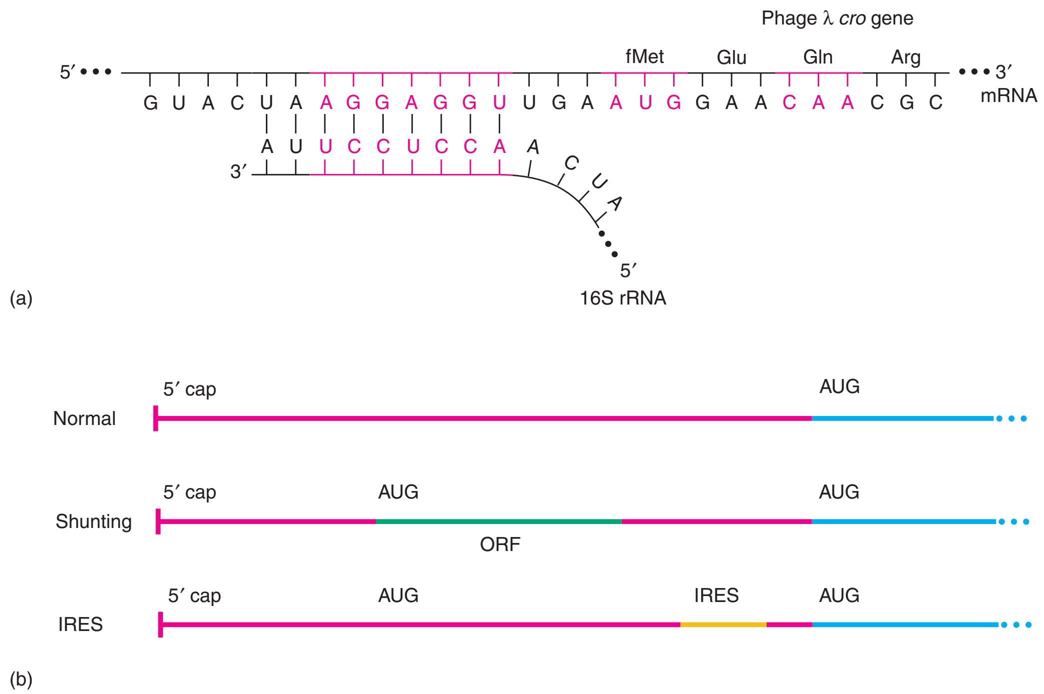


Figure 11.12 Translation initiation signals. (*a*) The Shine-Dalgarno hypothesis for prokaryotic translation. The Shine-Dalgarno sequence (AGGAGGU) is on the prokaryotic messenger RNA just upstream from the initiation codon AUG. Complementarity exists between this sequence and a complementary sequence (UCCUCCA) on the 3' end of the 16S ribosomal RNA. (*b*) Scanning, shunting, and internal ribosome entry in eukaryotic messenger RNA. The 5' untranslated region of a eukaryotic gene is shown in *red;* the beginning of the gene in *blue.* Normally, in the scanning model, the initiation codon of the gene is the first AUG encountered. In shunting, an open reading frame (ORF, *green*) may or may not be present to provide secondary structure in the messenger RNA to shunt scanning to the main gene. If the open reading frame is translated, reinitiation of translation at the same ribosome may occur at the main gene. Finally, an internal ribosome entry site (IRES, *yellow*) allows translation to begin within the messenger RNA without scanning.

there is a good deal of homology between prokaryotic and eukaryotic small ribosomal RNAs, the Shine-Dalgarno region is absent in eukaryotes. The actual mechanism for recognizing the 5′ end of eukaryotic messenger RNA appears to be based on recognition of the 5′ cap of the messenger RNA by the cap-binding protein with recruitment of other initiation factors and the small subunit of the ribosome. This is followed by the small subunit's movement down the messenger RNA. The ribosome scans the messenger RNA until it recognizes the initiation codon. This model is referred to as the **scanning hypothesis.**

In several known cases in eukaryotes, a process called **shunting** occurs, in which the first AUG does not serve as the initiation codon; rather, scanning begins, but it bypasses a region of the messenger RNA upstream of the initiation codon, called the *leader* or **5′ untranslated region (5′ UTR),** in favor of an AUG further down the messenger RNA. The cause of shunting seems to be secondary structure in the messenger RNA, upstream from the AUG codon that actually serves as the initiation codon. In some cases, very small genes, called **open reading frames (ORFs),** are present in this region of the messenger RNA and play some role in shunting. It seems also that some ORFs are translated, and then the main gene is translated by the same ribosome in a process called **reinitiation** (fig. 11.12*b*). We have seen this in the genes of some plant and animal viruses; it is a topic under current study.

Under some circumstances, eukaryotic ribosomes can initiate protein synthesis within the messenger RNA if that messenger RNA contains a sequence called an **internal ribosome entry site.** These sequences were discovered in the poliovirus RNA and in several cellular messenger RNAs. They are at least four hundred nucleotides long. Thus, although scanning accounts for the initiation of most eukaryotic messenger RNAs at their 5′ ends, some initiation can take place internally in messenger RNAs that have internal ribosome entry sites.

Aminoacyl and Peptidyl Sites in the Ribosome

When the initiator transfer RNA joins the 30S subunit of the prokaryotic ribosome with its messenger RNA attached, it fits into one of three sites in the ribosome. These sites, or cavities in the ribosome, are referred to as the aminoacyl site **(A site),** the peptidyl site **(P site),** and the exit site (**E site,** fig. 11.13) Here, we concentrate on the A and P sites, each of which contains a transfer RNA just before forming a peptide bond: the P site contains the transfer RNA with the growing peptide chain (peptidyl-tRNA); the A site contains a new transfer RNA with its single amino acid (aminoacyl-tRNA). The exit site helps eject depleted transfer RNAs after a peptide bond forms. When the complete 70S ribosome of figure 11.11 has formed, the initiation fMET-$\text{tRNA}_\text{f}^\text{Met}$ is placed directly into the P site (fig. 11.13), the only charged transfer RNA that can be placed directly there. The association of transfer RNA and ribosome is aided by a G-C base pairing between the 3′-CCA terminus of all transfer RNAs and a guanine in the 23S ribosomal RNA.

Elongation

Positioning a Second Transfer RNA

The next step in prokaryotic translation is to position the second transfer RNA, which is specified by the codon at the A site. The second transfer RNA is positioned in the A site of the ribosome so it is able to form hydrogen bonds between its anticodon and the second codon on the messenger RNA. This step requires the correct transfer RNA, another GTP, and two proteins called **elongation factors** (**EF-Ts** and **EF-Tu**). EF-Tu, bound to GTP, is required to position a transfer RNA into the A site of the ribosome (fig. 11.14). After the transfer RNA is positioned, the GTP is hydrolyzed to GDP + P_i. Upon hydrolysis of the GTP, the EF-Tu/GDP complex is released from the ribosome. EF-Ts is required to regenerate an EF-Tu/GTP complex. EF-Ts displaces the GDP on EF-Tu. Then a new GTP displaces EF-Ts, and now the EF-Tu/GTP complex can bind another transfer RNA. Here again, the hydrolysis of GTP changes the shape of the GTP so that the EF-Tu/GDP complex can depart from the ribosome after the transfer RNA is in place in the A site (fig. 11.15). Figure 11.16 shows the ribosome at the end of this step. EF-Tu does not bind fMet-$\text{tRNA}_\text{f}^\text{Met}$, so this blocked (formylated) methionine cannot be inserted into a growing peptide chain.

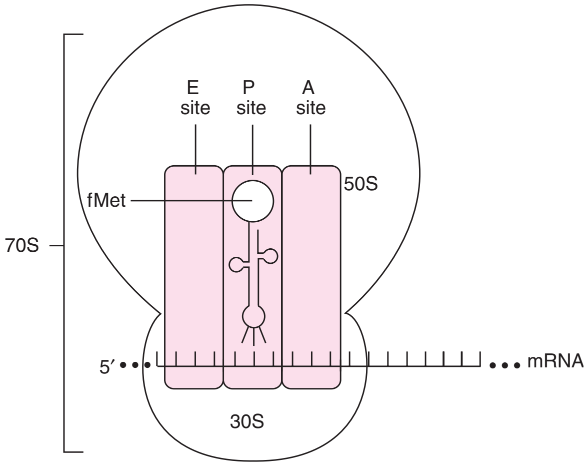


Figure 11.13 The 70S ribosome contains an A site, a P site, and an E site that can receive tRNAs. The messenger RNA runs through the bottom of the sites.

It takes several milliseconds for the GTP to be hydrolyzed, and another few milliseconds for the EF-Tu/GDP to actually leave the ribosome. During those two intervals of time, the codon-anticodon fit of the transfer RNA is scrutinized. If the correct transfer RNA is in place, a peptide bond forms. If not, the charged transfer RNA is released and a new cycle of EF-Tu/GTP-mediated testing of transfer RNAs begins. The error rate is only about one mistake in ten thousand amino acids incorporated into protein. The speed of amino acid incorporation is about fifteen amino acids per second in prokaryotes and about two to five per second in eukaryotes.

Peptide Bond Formation

The two amino acids on the two transfer RNAs are now in position to form a peptide bond between them; both amino acids are juxtaposed to an enzymatic center, **peptidyl transferase,** in the 50S subunit. This enzymatic center, an integral part of the 50S subunit, was originally believed to be composed of parts of several of the 50S proteins. Now, however, it is believed to have ribozymic activity, enzymatic activity of the ribosomal RNA of the ribosome. The enzymatic activity involves a bond transfer from the carboxyl end of N-formyl methionine to the amino end of the second amino acid (phenylalanine in fig. 11.16). Every subsequent peptide bond is identical, regardless of the amino acids involved. The energy used is contained in the high-energy ester bond between the transfer RNA in the P site and its amino acid (fig. 11.17). Immediately after the formation of the peptide bond, the transfer RNA with the dipeptide is in the A site, and a depleted transfer RNA is in the P site (box 11.2).

Translocation

The next stage in elongation is translocation of the ribosome in relation to the transfer RNAs and the messenger

RNA. Elongation factor EF-G, earlier called translocase, catalyzes the translocation process. The ribosome must be converted from the *pretranslocational state* to the *posttranslocational state* by the action of EF-G, which physically moves the messenger RNA and its associated transfer RNAs (fig. 11.18). This movement is accomplished by the hydrolysis of a GTP to GDP after EF-G enters the ribosome at the A site. After the first posttranslocational state is reached, the depleted transfer RNA in the E site is ejected, leaving the ribosome ready to accept a new charged transfer RNA in the A site. A computer-generated diagram of a ribosome with all three transfer RNA sites occupied is shown in figure 11.18*b*. In eukaryotes, three elongation factors perform the same tasks that EF-Tu, EF-Ts, and EF-G perform in prokaryotes. The factor eEF1α replaces EF-Tu, eEF1βγ replaces EF-Ts, and eEF2 replaces EF-G.

When translocation is complete, the situation is again as diagrammed in figure 11.13, except that instead of fMet-tRNA$_f^{Met}$, the P site contains the second transfer RNA (tRNA^{Phe}) with a dipeptide attached to it. The process of elongation is then repeated, with a third transfer RNA coming into the A site. The process repeats from here to the end (fig. 11.19), synthesizing a peptide starting from the amino (N-terminal) end and proceeding to the carboxyl (C-terminal) end. During the repetitive aspect of

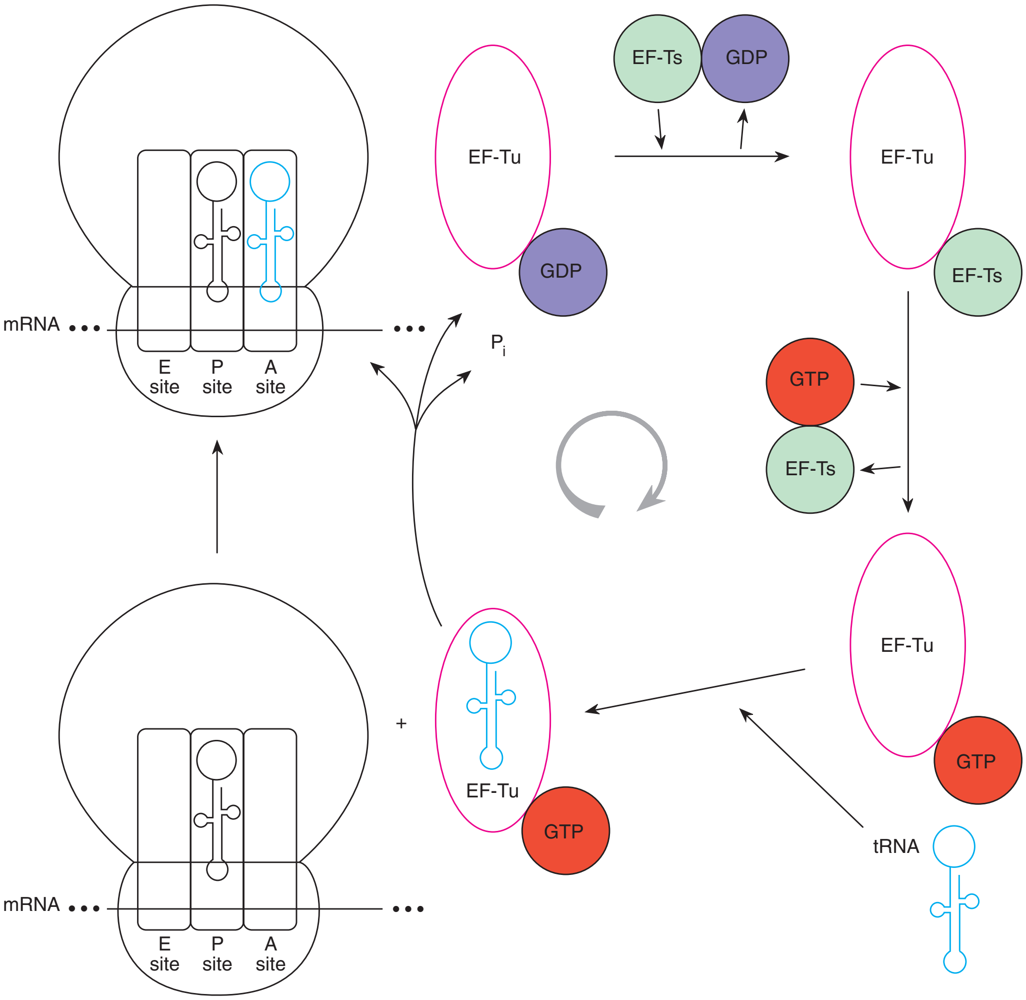


Figure 11.14 The EF-Ts/EF-Tu cycle. EF-Ts and EF-Tu are required for a transfer RNA to attach to the A site of the ribosome. At *top center,* we have EF-Tu attached to a GDP. The GDP is then displaced by EF-Ts, which in turn is displaced by GTP. A transfer RNA attaches and is brought to the ribosome. If the codon-anticodon fit is correct, the transfer RNA attaches at the A site with the help of the hydrolysis of GTP to GDP + P_i, allowing the EF-Tu to release. The EF-Tu is now back where we started. Since EF-Tu has a strong affinity for GDP, the role of EF-Ts is to displace the GDP, and later to be replaced by GTP.

BOX 11.2

Biomedical Applications

Antibiotics

Antibiotics, substances living organisms produce that are toxic to other living organisms, are of interest to us for two reasons: They have been extremely important in fighting the diseases that strike human beings and farm animals, and many are useful tools for analyzing protein synthesis. Some antibiotics impede the process of protein synthesis in a variety of ways, often poisoning bacteria selectively; the effectiveness of antibiotics normally derives from the metabolic differences between prokaryotes and eukaryotes. For example, an antibiotic that blocks a 70S bacterial ribosome without affecting an 80S human ribosome could be an excellent antibiotic. About 160 antibiotics are known.

PUROMYCIN

Puromycin resembles the 3′ end of an aminoacyl-tRNA (fig. 1). It is bound to the A site of the bacterial ribosome, where peptidyl transferase creates a bond from the nascent peptide attached to the transfer RNA in the P site to puromycin. Elongation can then no longer occur. The peptide chain is released prematurely, and protein synthesis at the ribosome terminates.

Experiments with puromycin helped demonstrate the existence of the A and P sites of the ribosome. It was found that puromycin could not bind to the ribosome if translocation factor EF-G were absent. With EF-G, translocation took place, and puromycin could then bind to the ribosome. Puromycin's ability to bind only after translocation indicates that a second site on the ribosome becomes available after translocation.

STREPTOMYCIN, TETRACYCLINE, AND CHLORAMPHENICOL

Streptomycin, which binds to one of the proteins (protein S12) of the 30S subunit of the prokaryotic ribosome, inhibits initiation of protein synthesis. Streptomycin also causes misreading of codons if chain initiation has already begun, presumably by altering the conformation of the ribosome so that transfer RNAs are less firmly bound to it. Bacterial mutants that are streptomycin resistant, as well as mutants that are streptomycin dependent (they cannot survive without the antibiotic), occur. Both types of mutants have altered 30S subunits, specifically changed in protein S12.

Tetracycline blocks protein synthesis by preventing an aminoacyl-tRNA from binding to the A site on the ribosome. Chloramphenicol blocks protein synthesis by binding to the 50S subunit of the prokaryotic ribosome, where it blocks the peptidyl transfer reaction. Chloramphenicol does not affect the eukaryotic ribosome. However, chloramphenicol, as well as several other antibiotics, is used cautiously because the mitochondrial ribosomes within eukaryotic cells are very similar to prokaryotic ribosomes. Some of the antibiotics that affect prokaryotic ribosomes thus also affect mitochondria. As was mentioned, the similarity between bacteria and mitochondria implies that mitochondria have a prokaryotic origin. (Similarities between cyanobacteria and chloroplasts also support the idea that chloroplasts have a prokaryotic origin.)

THE TROUBLE WITH ANTIBIOTICS

Over the years, antibiotics have virtually eliminated certain diseases from the industrialized world. They have also made modern surgery possible by preventing most serious infections that tend to follow operations. Antibiotics have been so successful that, in the 1980s, many pharmaceutical companies drastically cut back the development of new antibiotics. However, a disaster was in the making as we overprescribed antibiotics to people and farm animals: bacteria are not prepared to take this onslaught without fighting back.

Mutation takes place all the time at a low but dependable rate. Thus, resistant bacteria are constantly arising from sensitive strains. We can select for penicillin- and streptomycin-resistant strains of bacteria in the laboratory by allowing the antibiotic to act as a selective agent, removing all but the resistant individuals. The same sort of artificial selection that we can apply in the lab applies every time a person or animal takes an antibiotic. We may be at a point now where the ability of bacteria to develop resistance, and to pass that resistance to other strains, has put us on the verge of disaster. The process of evolution works amazingly fast in bacteria because of their ubiquity, large population sizes, and ability to transfer genetic material between individuals. We may shortly find ourselves as we were before World War II, when simple infections in hospitals were often lethal. Right now, only one antibiotic can keep the common—and potentially deadly—infectious bacterium *Staphylococcus* under control: vancomycin. Several types of disease-causing bacteria have already evolved a tolerance to vancomycin.

The answer to this potentially disastrous problem is to develop new antibiotics and reduce the irresponsible use of antibiotics in people and animals. Hopefully, the warning bell has sounded. At least a dozen new antibiotics that show promise are in the early stages of development by pharmaceutical companies.

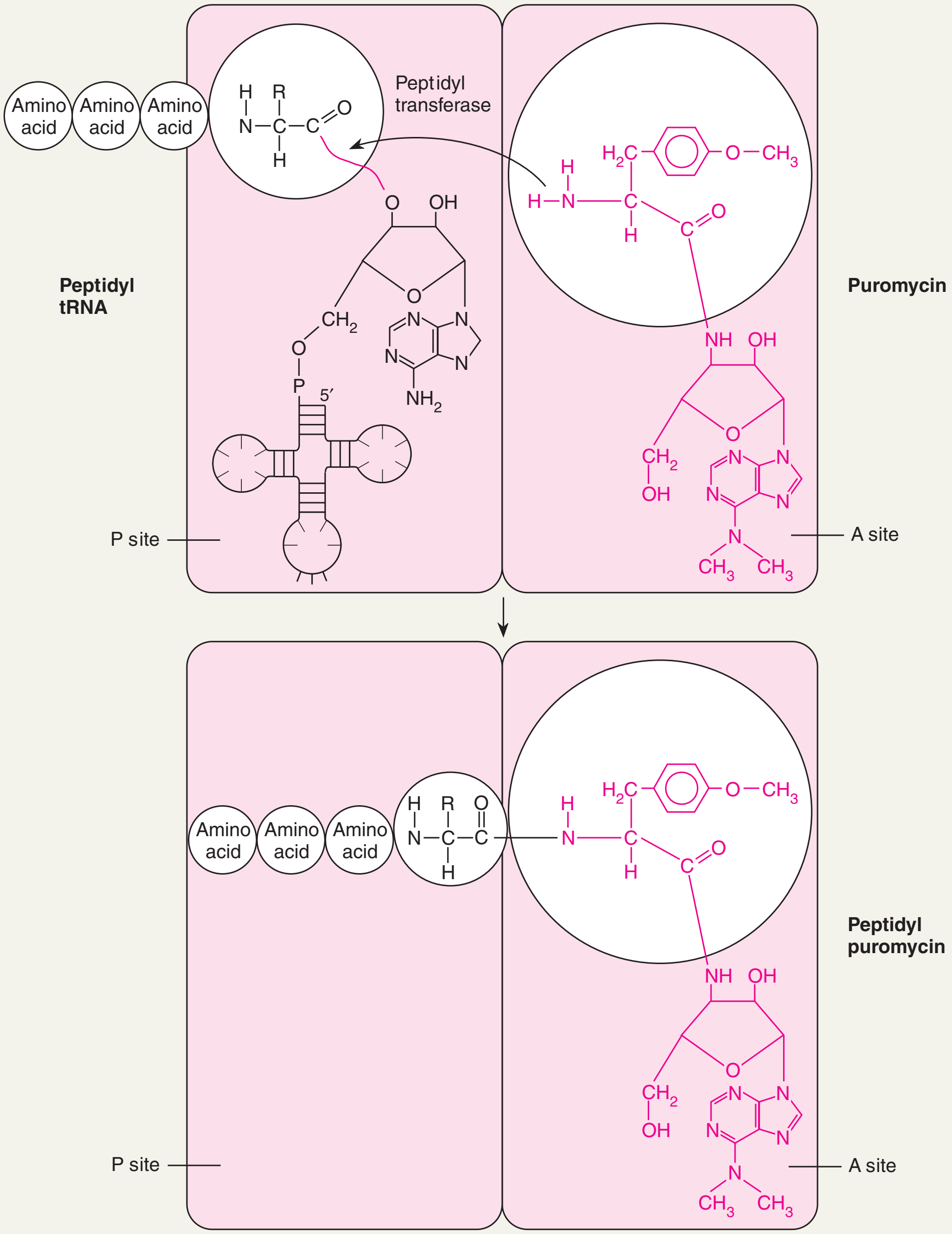


Figure 1 Puromycin is bound to the A site of the ribosome. A peptide bond then forms. Further elongation is prevented, and the chain is terminated.

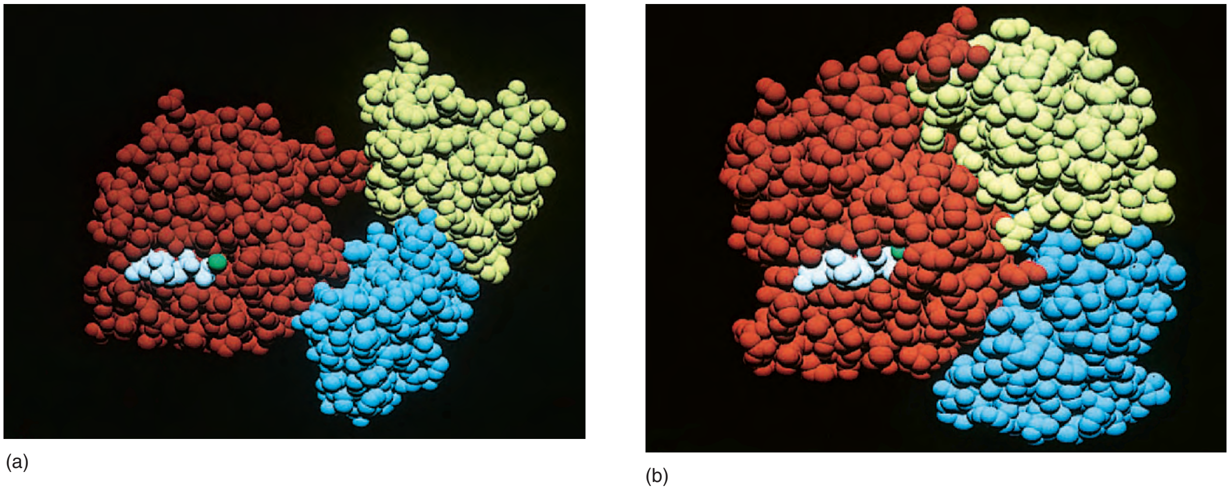


Figure 11.15 Space-filling model of EF-Tu bound with (*a*) GDP and (*b*) GTP, showing the change in the protein's structure. *Yellow, blue,* and *red* are domains of the protein. The GTP and GDP are in *white,* with a magnesium ion, Mg^{2+}, in *green.* When EF-Tu is bound with GDP, there is a visible hole in the molecule. The hole disappears when GTP is bound. The aminoacyl-transfer RNA is believed to bind between the *red* and *yellow* domains. (Courtesy of Rolf Hilgenfeld.)

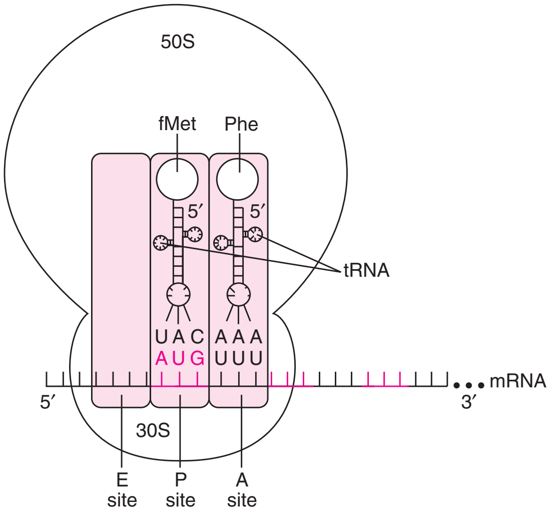


Figure 11.16 A ribosome with two transfer RNAs attached. In this case, the second codon (UUU) is for the amino acid phenylalanine. The two amino acids are next to each other.

protein synthesis, two GTPs are hydrolyzed per peptide bond: one GTP in the release of EF-Tu from the A site, and one GTP in the translocational process of the ribosome after the peptide bond has formed. In addition, every charged transfer RNA has had an amino acid attached at the expense of the hydrolysis of an ATP to AMP +PP. There is some evidence that the action of EF-Tu hydrolyzes two GTPs.

Termination

Nonsense Codons

Termination of protein synthesis in both prokaryotes and eukaryotes occurs when one of three **nonsense codons** appears in the A site of the ribosome. These codons are UAG (sometimes referred to as *amber*), UAA (*ochre*), and UGA (*opal*). ("Amber," or brown stone, is the English translation of the name Bernstein, a graduate student who took part in the discovery of UAG in R. H. Epstein's lab at the California Institute of Technology. "Ochre" and "opal" are tongue-in-cheek extensions of the first label.) In prokaryotes, three proteins called **release factors (RF)** are involved in termination, and a GTP is hydrolyzed to GDP + P_i.

When a nonsense codon enters the A site on the ribosome, a release factor recognizes it. RF1 and RF2 are class 1 release factors: They recognize stop codons and then promote hydrolysis of the bond between the terminal amino acid and its tRNA in the P site. Class 2 release factors (RF3) do not recognize stop codons, but they stimulate class 1 release factors to act. RF1 recognizes the stop codons UAA and UAG, and RF2 recognizes UAA and UGA (fig. 11.20). Both do so because they have tripep-

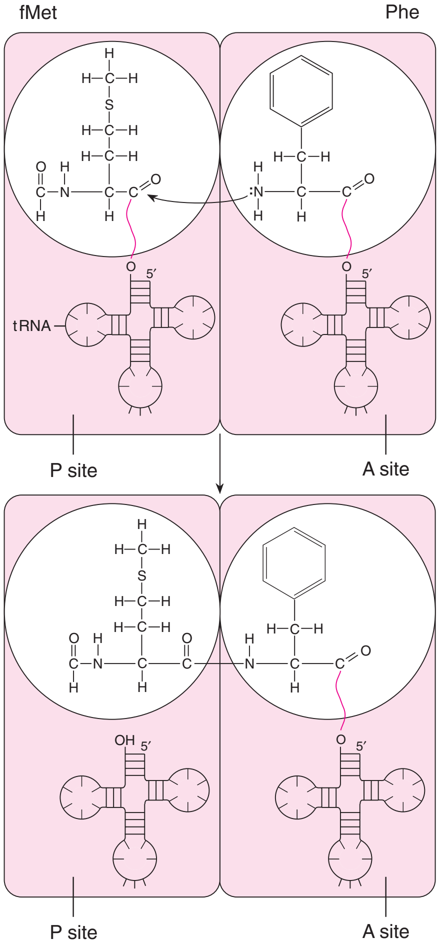


Figure 11.17 Peptide bond formation on the ribosome between N-formyl methionine and phenylalanine. The bond attaching the carboxyl end of the first amino acid to its tRNA is transferred to the amino end of the second amino acid. The first tRNA is now uncharged, whereas the second tRNA has a dipeptide attached.

tides that mimic anticodons to recognize the stop codons: proline-alanine-threonine in RF1 and serine-proline-phenylalanine in RF2. In this **molecular mimicry,** a protein mimics the shape of a nucleic acid in order to function properly.

The next base in the messenger RNA past the stop codon is usually an adenine, required for efficient termination. After the release factors act, with the hydrolysis of a GTP, the ribosome has completed its task of translating mRNA into a polypeptide. Final release of all factors and dissociation of the two subunits of the ribosome take place with the help of IF3, which rebinds to the 30S subunit, and a **ribosome recycling factor (RRF).** Table 11.1 compares prokaryotic and eukaryotic translation.

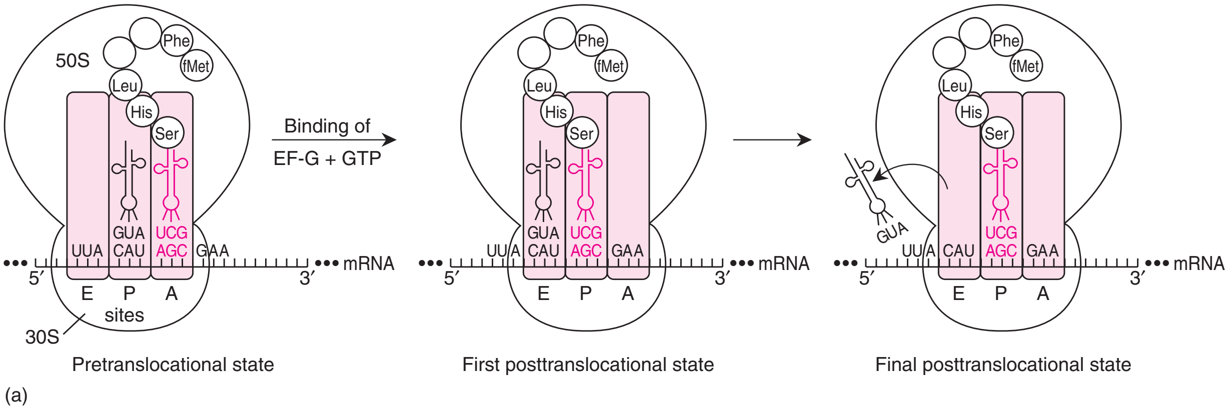
Rate and Cost of Translation

As mentioned, the average speed of protein synthesis is about fifteen peptide bonds per second in prokaryotes. Discounting the time for initiation and termination, an average protein of three hundred amino acids is synthesized in about twenty seconds (the released protein forms its final structure spontaneously or is modified with the aid of other proteins, as we shall see). An equivalent eukaryotic protein takes about 2.5 minutes to be synthesized. The energy cost is at least four high-energy phosphate bonds per peptide bond (two from an ATP during transfer RNA charging, and two from GTP hydrolysis during transfer RNA binding at the A site and translocation), or about twelve hundred high-energy bonds per protein. This cost is very high—about 90% of the energy production of an *E. coli* cell goes into protein synthesis. A high energy cost is presumably the price a living system has to pay for the speed and accuracy of its protein synthesis.

Coupling of Transcription and Translation

In prokaryotes, such as *E. coli,* in which no nuclear envelope exists, translation begins before transcription is completed. Figure 11.21 shows a length of an *E. coli* chromosome. An RNA polymerase is visible on the DNA, transcribing a gene. The messenger RNA, still being synthesized, can be seen extending away from the DNA. Attached to the messenger RNA are about a dozen ribosomes. Since translation starts at the end of the messenger that is synthesized first (5′), an initiation complex can form and translation can begin shortly after transcription begins. As translation proceeds along the messenger, its 5′ end will again become exposed, and a new initiation complex can form. The occurrence of several ribosomes translating the same messenger is referred to as a **polyribosome,** or simply a **polysome** (fig. 11.22).

In prokaryotes, most messenger RNAs contain the information for several genes. These RNAs are said to be **polycistronic** (fig. 11.23). (*Cistron,* another term for gene, is defined in chapter 12.) Each gene on the messenger RNA is translated independently: each has a Shine-Dalgarno sequence for ribosome recognition (see fig. 11.12) and an initiation codon (AUG) for fMet. The ribosome that completes the translation of the first gene may or may not continue to the second gene after dissociation. The translation of any gene follows all the steps we have outlined.



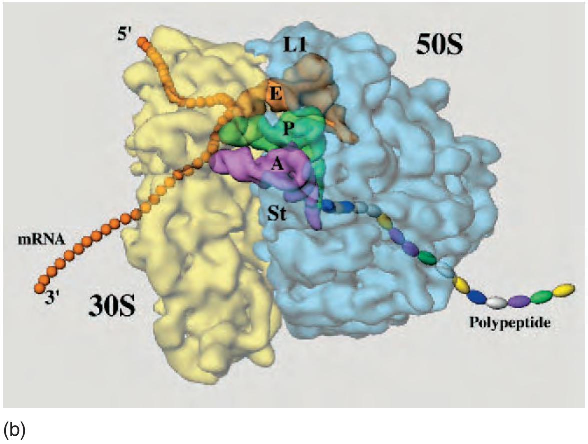
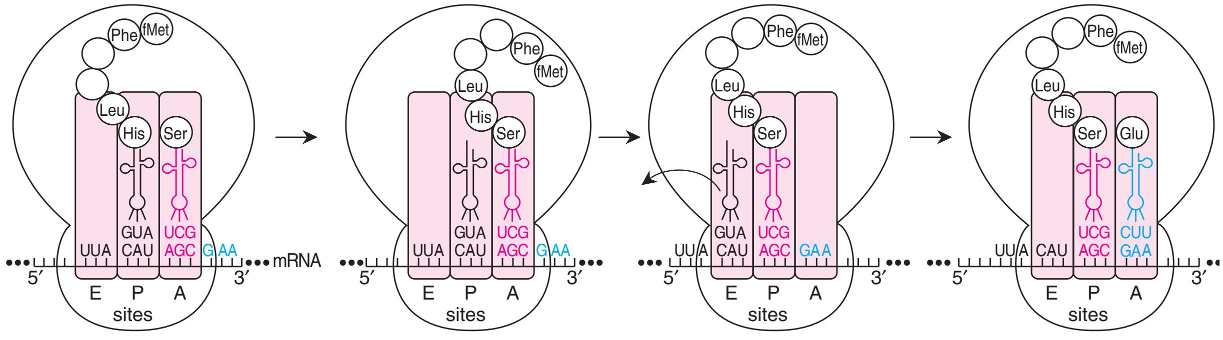


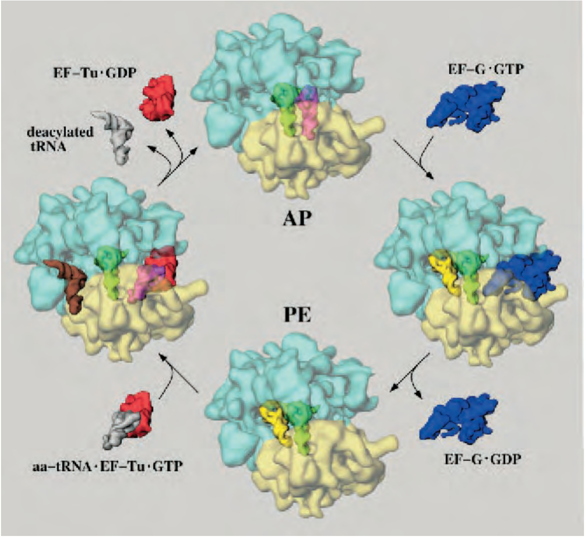
Figure 11.18 (*a*) EF-G's translocation of the ribosome converts it from a pretranslocational state (P and A sites occupied) to a posttranslocational state (E and P sites occupied). The uncharged transfer RNA in the E site is then ejected. (*b*) A see-through model of the 70S ribosome of *E. coli* with transfer RNAs in the A, P, and E sites. The structure was determined by cryoEM mapping, an electron microscopic technique using rapidly frozen specimens. The position of the messenger RNA is shown, as well as the stalk of the 50S subunit (St) and one of the polypeptides of the large subunit, L1. ([*b*] Courtesy of Joachim Frank, Howard Hughes Medical Institute.)

Table 11.1 **Some Comparisons Between Prokaryotic and Eukaryotic Translation**

	Prokaryotes	Eukaryotes
Initiation codon	AUG, occasionally GUG, UUG	AUG, occasionally GUG,CUG
Initiation amino acid	N-formyl methionine	Methionine
Initiation tRNA	tRNA_f^{Met}	tRNA_i^{Met}
Interior methionine tRNA	tRNA_m^{Met}	tRNA_m^{Met}
Initiation factors	IF1,IF2,IF3	eIF factors
Elongation factor	EF-Tu	eEF1α
Elongation factor	EF-Ts	eEF1βγ
Translocation factor	EF-G	eEF2
Release factors	RF1, RF2, RF3, RRF	eRF1, eRF3



(a)



(b)

Figure 11.19 Cycle of peptide bond formation and translocation on the ribosome. (*a*) After the peptide bond is transferred (fig. 11.17), the ribosome and messenger RNA move over one codon. Now the transfer RNA with the peptide is in the P site, and the A site is again open. In this example, the next transfer RNA that moves into the A site carries glutamic acid. (*b*) Three-dimensional model of the translocation process minus the mRNA and amino acids. The tRNA in the A site is *pale blue,* the tRNA in the P site is *green,* and the tRNA in the E site is *yellow,* then *brown* when ready to leave. Going clockwise from *a,* in which the A and P sites are occupied: EF-G translocates the ribosome after peptide bond formation and then evacuates the A site. Ef-Tu brings a new charged tRNA to the A site while the E site is emptied. ([*b*] Courtesy of Joachim Frank.)

In eukaryotes, however, almost all messenger RNAs contain the information for only one gene **(monocistronic).** Since most ribosomal recognition of eukaryotic genes depends on the 5′ cap, and since each eukaryotic messenger RNA has only one cap, usually only one polypeptide can be translated for any given messenger RNA. Exceptions occur when the messenger RNAs contain internal ribosome entry sites. Although it is certainly not the rule, the translated peptide can be modified or cleaved into smaller functional peptides. For example, in mice, a single messenger RNA codes for a protein that is later cleaved into epidermal growth factor and at least seven other related peptides. In addition, the same sequence can, in some cases, give rise to alternative proteins through alternative start codons, termination read-through, or alternative splicing.

More on the Ribosome

In the last chapter, we briefly discussed the shape and composition of the ribosomal subunits. All of the protein and RNA components have been isolated. Assembly pathways are known. We know approximately where the messenger RNA, initiation factors, and EF-Tu are located on the 30S subunit during translation (fig. 11.24; cf. fig. 11.18). We also know where peptidyl transferase activity and EF-G reside on the 50S subunit, which has a cleft leading into a tunnel that passes through the structure. At present, it seems that the nascent peptide passes through this

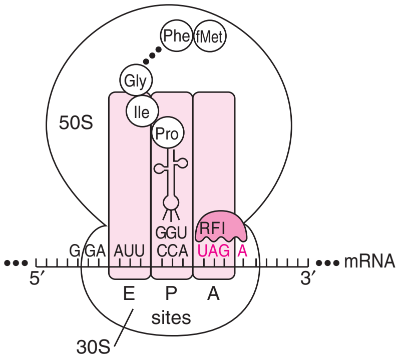


Figure 11.20 Chain termination at the ribosome. One of two release factors recognizes a nonsense codon in the A site. In this case, RF1 recognizes UAG. The complex then falls apart, releasing the peptide.

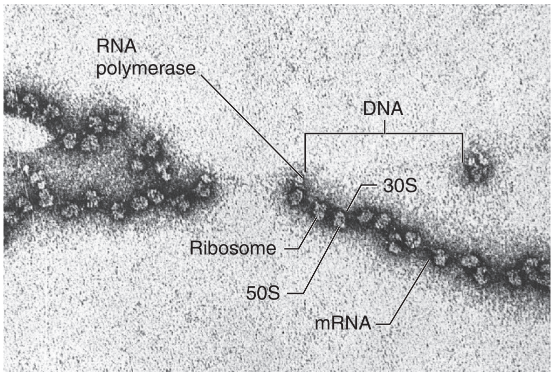
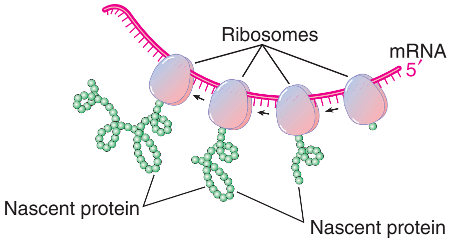
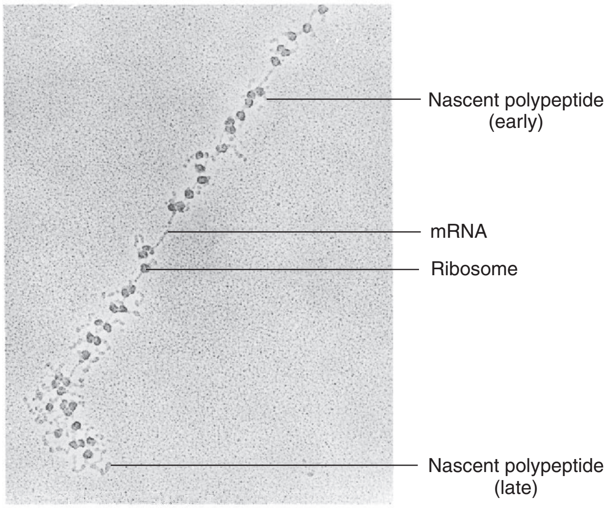


Figure 11.21 A polysome (i.e., multiple ribosomes on the same strand of mRNA). Each ribosome is approximately 250 Å units across. Also visible in this illustration are DNA and RNA polymerase. (Reproduced courtesy of Dr. Barbara Hamkalo, *International Review of Cytology,* (1972) 33:7, fig. 5. Copyright by Academic Press, Inc., Orlando, Florida.)



(a)



(b)

Figure 11.22 (*a*) Protein synthesis at a polysome. Nascent proteins exit from a tunnel in the 50S subunit. Messenger RNA is being translated by the ribosomes while the DNA is being transcribed. (*b*) A messenger RNA from the midge, *Chironomus tentans,* showing attached ribosomes and nascent polypeptides emerging from the ribosomes. Note the 5′ end of the messenger RNA at the *upper right* (small peptides). Magnification 165,000×. ([*b*] Courtesy of S. L. McKnight and O. L. Miller, Jr.)

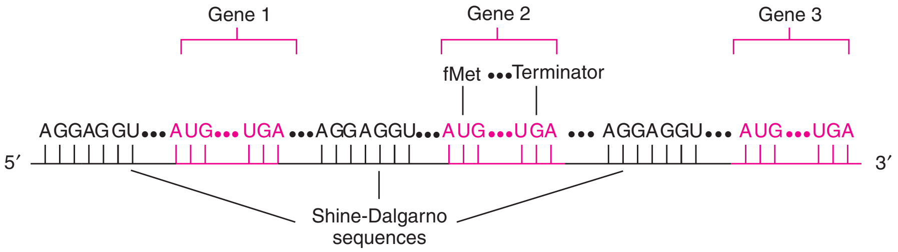


Figure 11.23 A prokaryotic polycistronic mRNA. Note the several Shine-Dalgarno sequences for ribosomal attachment and the initiation and termination codons marking each gene.

tunnel, emerging close to a membrane-binding site (fig. 11.24). The tunnel can hold a peptide length of about forty amino acids. Note that although every ribosome has a membrane-binding site, not all active ribosomes are bound to membranes.

The Signal Hypothesis

Ribosomes are either free in the cytoplasm or associated with membranes, depending on the type of protein being synthesized. Membrane-bound ribosomes, indistinguishable from free ribosomes, synthesize proteins that enter membranes. These proteins either become a part of the membrane or, in eukaryotes, either pass into membrane-bound organelles (e.g., the Golgi apparatus, mitochondria, chloroplasts, vacuoles) or are transported outside the cell membrane. The **signal hypothesis** of G. Blobel, a 1999 Nobel laureate, and his colleagues, explains the mechanism for membrane attachment. The mechanism applies to both prokaryotes and eukaryotes. Here, we describe it in mammals.

Gunter Blobel (1936–). (Courtesy of Dr. Gunter Blobel, Dept. of Cell Biology, Rockefeller University.)

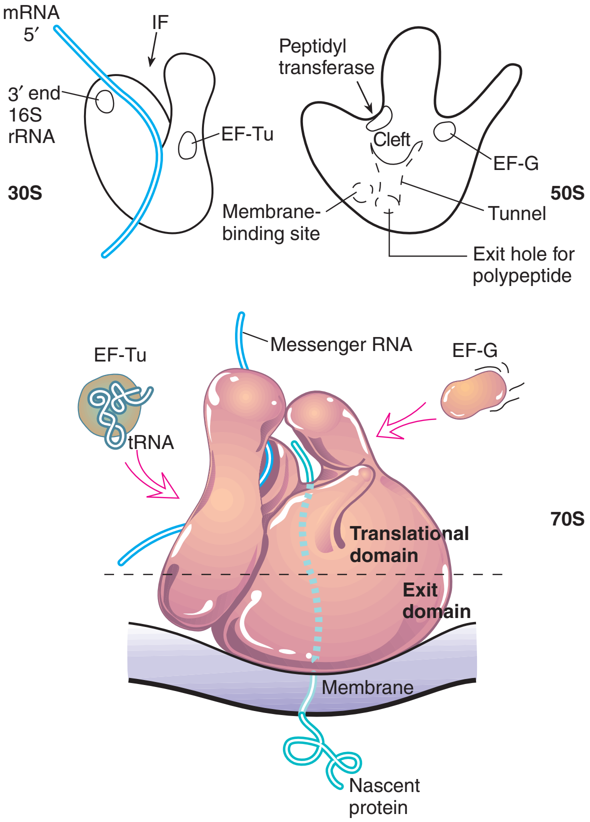


Figure 11.24 Functional sites on the prokaryotic ribosome. The ribosome is synthesizing a protein involved in membrane passage. Note the position of the messenger RNA on the 30S subunit and the cleft, tunnel, and membrane-binding site on the 50S subunit. (From C. Bernabeu and J. A. Lake, *Proceedings of the The National Academy of Sciences;* 79:3111–15, 1982. Reprinted by permission.)

The signal for membrane insertion is coded into the first one to three dozen amino acids of membrane-bound proteins. This **signal peptide** takes part in a chain of events that leads the ribosome to attach to the membrane and to the insertion of the protein. The first step occurs when the signal peptide becomes accessible outside of the ribosome. A ribonucleoprotein particle called the **signal recognition particle** (SRP), which consists of six different proteins and a 7S RNA about three hundred nucleotides long, recognizes the signal peptide. The complex of signal recognition particle, ribosome, and signal peptide then passes, or diffuses, to a membrane, where the SRP binds to a receptor called a **docking protein** (DP) or signal recognition particle receptor (fig. 11.25). During this time, protein synthesis halts. The ribosome is brought into direct contact with the membrane, and other proteins of the membrane help anchor the ribosome. Protein synthesis then resumes, with the nascent protein usually passing directly into a **translocation channel (translocon).** Once through the membrane, the signal peptide is cleaved from the protein by an enzyme called *signal peptidase.* A striking verification of this hypothesis came about through recombinant DNA techniques (chapter 13). A signal sequence was placed in front of the α-globin gene, whose protein product is normally not transported through a membrane. When this gene was translated, the ribosome became membrane bound, and the protein passed through the membrane.

Since different proteins enter different membrane-bound compartments (e.g., the Golgi apparatus), some mechanism must direct a nascent protein to its proper

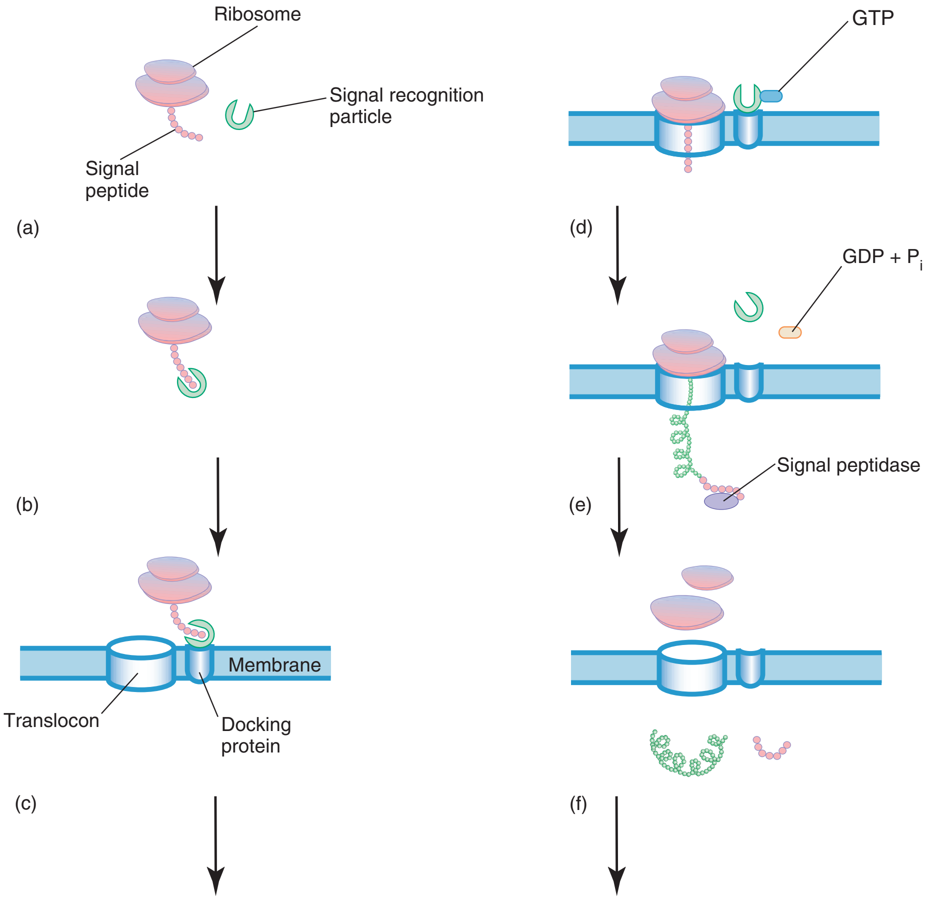


Figure 11.25 The signal hypothesis. A signal recognition particle recognizes a ribosome with a signal peptide, then draws the ribosome to a docking protein located near a translocon in the membrane. With the addition of GTP, the signal recognition particle releases the signal peptide; hydrolysis of the GTP to GDP + P_i causes the signal recognition particle to leave the docking protein. Peptide synthesis then resumes, with the newly synthesized peptide passing through the translocon in the membrane. A signal peptidase on the other side of the membrane removes the signal peptide. When translation is completed, the ribosome dissociates and drops free of the translocon.

membrane. This specificity seems to depend on the exact signal sequence and membrane-bound glycoproteins called *signal-sequence receptors.* Apparently, after the ribosome binds to the docking protein, the signal peptide interacts with a signal-sequence receptor, which presumably determines whether that protein is specific for that membrane. If it is, the remaining processes continue. If not, the ribosome may be released from the membrane.

The signal peptide does not seem to have a consensus sequence like the transcription or translation recognition boxes. Rather, similarities (at least for the endoplasmic reticulum and bacterial membrane-bound proteins) include a positively charged (basic) amino acid (commonly lysine or arginine) near the beginning (N-terminal end), followed by about a dozen hydrophobic (nonpolar) amino acids, commonly alanine, isoleucine, leucine, phenylalanine, and valine (table 11.2).

Table 11.2 **The Signal Peptide of the Bovine Prolactin Protein***

NH_2 − Met Asp Ser Lys Gly Ser Ser Gln Lys Gly Ser Arg Leu
Leu Leu Leu Leu Val Val Ser Asn Leu Leu Leu Cys Gln Gly Val
Val Ser | Thr Pro Val...Asn Asn Cys − COOH

Source: From Sasavage et al., *Journal of Biological Chemistry,* 257:678–81, 1982. Reprinted with permission.

* The vertical line separates the signal peptide from the rest of the protein, which consists of 199 residues.

Figure 11.26 Electron micrograph of a chaperone protein (GroEL) from *E. coli.* Note the hollow, barrel shape of the protein. (Courtesy of Dr. R. W. Hendrix.)

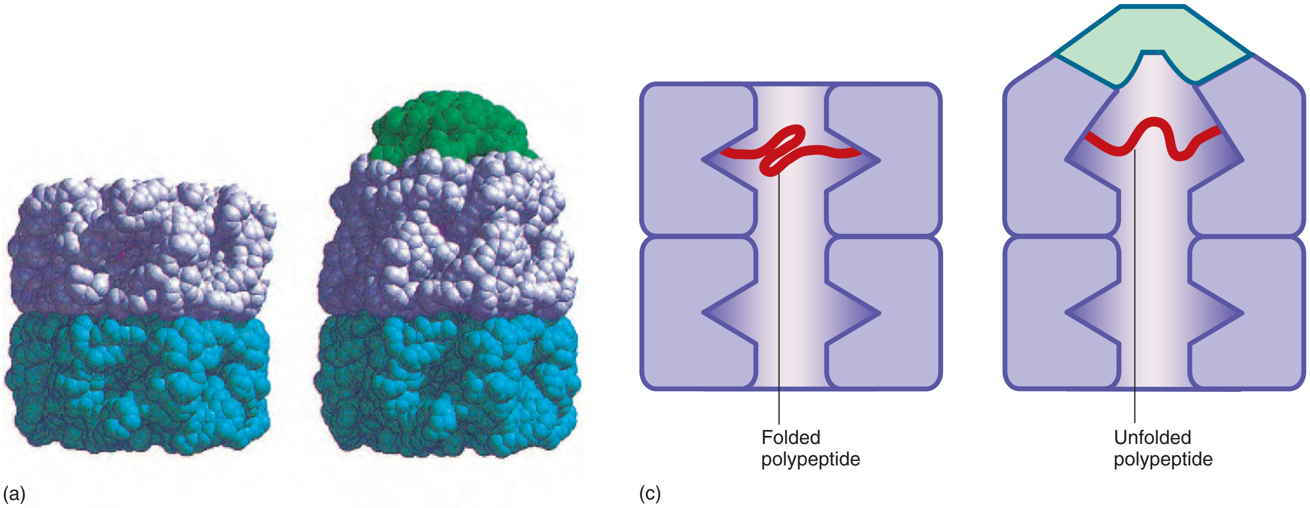
The mitochondrion, which needs to import upwards of one thousand proteins through both inner and outer membranes, poses a specific problem. Recent research has revealed a family of translocation proteins (called Tom proteins) in the outer membrane and a different set of translocation proteins (called Tim proteins) in the inner membrane. These proteins control the passage of proteins synthesized in the cytoplasm into the mitochondrion.

The Protein-Folding Problem

Since biochemist Christian Anfinsen won a 1972 Nobel Prize for showing that the enzyme ribonuclease refolds to its original shape after denaturation in vitro, scientists have believed that the final protein shape (secondary and tertiary structure) forms spontaneously. Recently it has been shown, however, that many proteins do not normally form their final active shape in vivo without the help of proteins called **chaperones** or **molecular chaperones.** The chaperones do not provide the three-dimensional structure of the proteins they help, but rather bind to a protein in the early stages of folding to prevent unproductive folding or to allow denatured proteins to refold correctly. Like human chaperones, they prevent or undo "incorrect interactions," according to J. Ellis. That is, many proteins have a large number of different structures they could fold into. Many of these structures would have no enzymatic activity or would form functionless aggregates with other proteins. Molecular chaperones allow proteins to fold into a thermodynamically stable and functional configuration. Each cycle of refolding requires ATP energy.

A well-studied class of chaperones is known as the *chaperonins,* or Hsp60 proteins, because they are heat shock proteins about 60 kilodaltons (60,000 daltons) in size. They occur in bacteria, chloroplasts, and mitochondria. One of the best studied of these chaperonins is the protein GroE of *E. coli.* This protein in its active form is composed of two components, GroEL and GroES. GroEL (Hsp60) is made up of two disks, each composed of seven copies of a polypeptide. GroES (Hsp10) is a smaller component composed of seven copies of a small subunit. GroEL forms a barrel in which protein folding takes place (fig. 11.26). The barrel is shaped in such a way that entering proteins of a certain size make contact at interior points in either the upper or lower ring of GroEL (upper ring shown in fig. 11.27). The attachment of GroES, the cap, causes the ring to open outward at the top, stretching the protein inside. This stretching takes energy from the hydrolysis of ATP molecules located inside the rings. When GroES dissociates, the protein can fold into a new, more functional, configuration. If it doesn't, the cycle repeats.

There are several classes of molecular chaperones, proteins of different sizes and shapes that recognize different groups of proteins or protein conformations. GroEL recognizes about 300 different proteins, small enough to fit into the barrel (20–60 kilodaltons) and having hydrophobic surfaces. These include many proteins in the transcription and translation machinery of the cell. Hsp90, another heat shock protein, recognizes proteins involved in signal transduction, discussed in chapter 16. Hsp70 recognizes hydrophobic regions in polypeptide side chains, many of which extend across membranes.



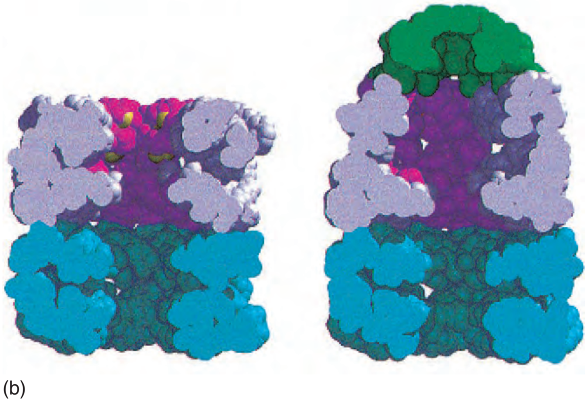


Figure 11.27 The change in structure of GroEL with GroES attached explains how the chaperonin can unfold a partially folded polypeptide to allow it to refold in a different way. (*a*) A space-filling model of GroEL is shown without (*left*) and with (*right*) GroES. GroEL's rings are *blue* and *magenta,* and GroES is *green.* (*b*) The same structures are seen in a cutaway view. (*c*) This diagram shows how the attachment of GroES causes the top part of the top ring of GroEL to pull apart an improperly folded polypeptide. ([*a* & *b*]: Reprinted from Bernd Bakau and Arthur L. Horwich, "The Hsp70 and Hsp60 Chaperone Machines" in *Cell*, vol. 92, 351–366. Copyright 1988, with permission from Elsevier Science.)

THE GENETIC CODE

Researchers in the mid-1950s assumed that the genetic code consisted of simple sequences of nucleotides specifying particular amino acids. They sought answers to questions such as: Is the code overlapping? Are there nucleotides between code words (punctuation)? How many letters make up a code word (codon)? Logic, along with genetic experiments, supplied some of the answers, but only with the rapidly improving techniques of biochemistry did they eventually decode the genetic language.

Triplet Nature of the Code

Several lines of evidence seemed to indicate that the nature of the code was triplet (three bases in messenger RNA specifying one amino acid). If codons contained only one base, they would only be able to specify four amino acids since there are only four different bases in DNA (or messenger RNA). A couplet code would have $4 \times 4 = 16$ two-base words, or codons, which is still not enough to specify uniquely twenty different amino acids. A triplet code would allow for $4 \times 4 \times 4 = 64$ codons, which are more than enough to specify twenty amino acids.

Evidence for the Triplet Nature of the Code

The experimental manipulation of mutant genes, primarily by Francis Crick and his colleagues, reinforced the triplet code concept. In these experiments, a chemical mutagen, the acridine dye proflavin, was used to cause inactivation of the rapid lysis (*rIIB*) gene of the bacteriophage T4. Proflavin inactivates the gene by either adding or deleting a nucleotide from the DNA (see chapter 12). The *rII* gene controls the plaque morphology of this bacteriophage growing on *E. coli* cells. Rapid-lysis mutants produce large plaques; the wild-type form of the gene, rII^+, results in normal plaque morphology.

Figure 11.28 shows the consequences of adding or deleting a nucleotide. From the point of addition or deletion onward, a **frameshift** causes codons to be read in different groups of three. If a deletion is combined with an addition to produce a double-mutant gene, the frameshift occurs only in the region between the two mutants. If this region is small enough or does not contain coding for vital amino acids, the function of the gene may be restored. Two deletions or two insertions combined will not restore the reading frame. However, Crick and his colleagues found that the combination of three additions or three deletions did restore gene function. This finding led to the conclusion that the genetic code was triplet, because a triplet code would be put back into the reading frame by three additions or three deletions (fig. 11.29).

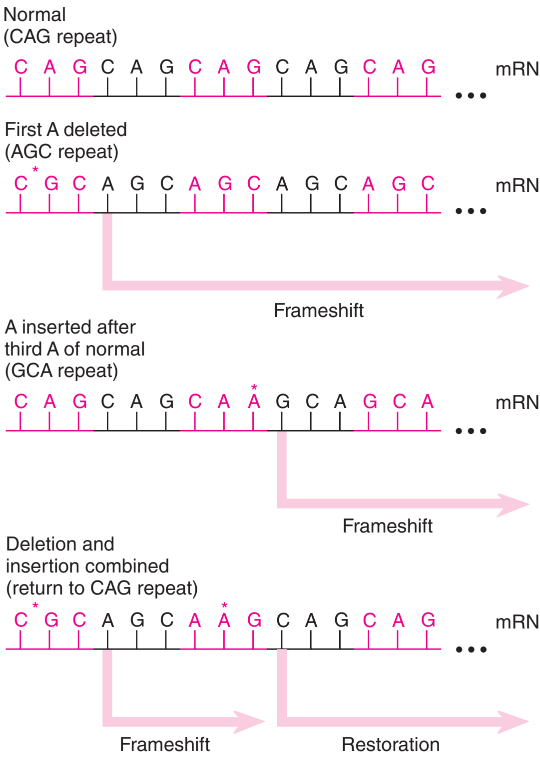


Figure 11.28 Frameshift mutations in a gene result from the addition or deletion of one or several nucleotides (any number other than a multiple of three) in the DNA. The messenger RNA shown here normally has a CAG repeat. A single-base deletion shifts the three-base reading frame to a series of AGC repeats. A later insertion restores the reading frame. *Asterisks* (*) indicate points of deletion or insertion.

Overlap and Punctuation in the Code

Questions still remained: was the code overlapping? Did it have punctuation? Several logical arguments favored a no-punctuation, nonoverlapping model (fig. 11.30). An overlapping code would be subject to two restrictions. First, a change in one base (a mutation) could affect more than one codon and thus affect more than one amino acid. But studies of amino acid sequences almost always showed that only one amino acid was changed, which argued against codon overlap. Second, certain restrictions affected which amino acids occurred next to each other in proteins. For example, the amino acid UUU coded could never be adjacent to the amino acid coded by AAA because one or both (depending on the number of bases overlapped) of the overlap codons UUA and UAA would always insert other amino acids between them. Overlap, then, seemed to be ruled out since every amino acid appears next to every other amino acid in one protein or another.

Punctuation between codons was also tentatively ruled out. The messenger RNA in the tobacco necrosis satellite virus has just about enough codons to specify its coat protein with no room left for a punctuating base or bases between each codon.

Breaking the Code

Once geneticists had figured out that the genetic code is in nonoverlapping triplets, they turned their attention to the sixty-four codons. They wondered which amino acid, for example, does ACU specify? The work was done in two stages. In the first stage, M. W. Nirenberg, S. Ochoa, and their colleagues made long artificial messenger RNAs and determined which amino acids these messenger RNAs incorporated into protein. In the second stage, specific triplet RNA sequences were synthesized. The researchers then determined the amino acid-transfer RNA complex that was bound by each sequence.

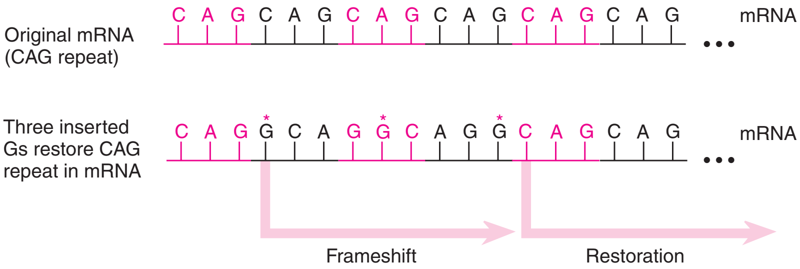
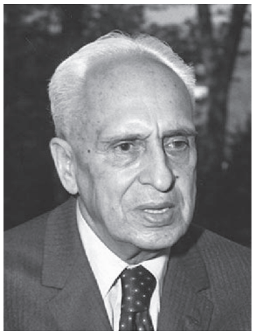
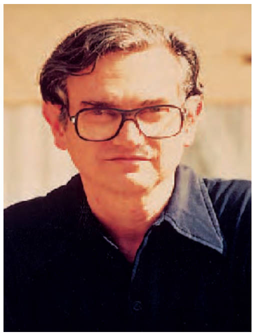


Figure 11.29 The coding frame of CAG repeats is first shifted and then restored by three additions (insertions). *Asterisks* (*) indicate insertions.



Severo Ochoa (1905–1993). (Courtesy of Dr. Severo Ochoa.)



Marshall W. Nirenberg (1934–). (Courtesy of Dr. Marshall W. Nirenberg.)

Synthetic Messenger RNAs

The ability to synthesize long-chain messenger RNAs resulted from the 1955 discovery of M. Grunberg-Manago and Ochoa of the enzyme **polynucleotide phosphorylase,** which joins diphosphate nucleotides into long-chain, single-stranded polynucleotides. Unlike a polymerase, polynucleotide phosphorylase does not need a primer on which to act. This enzyme is found in all bacteria. (Its main function in the cell is probably the reverse of its use here. It most likely serves as an exonuclease, degrading messenger RNA.) In 1961, Nirenberg and J. H. Matthei added artificially formed RNA polynucleotides of known composition to an *E. coli* ribosomal system and looked for the incorporation of amino acids into proteins.

The system just described is called a **cell-free system,** a mixture primarily of the cytoplasmic components of cells, such as *E. coli,* but missing nucleic acids and membrane components. These systems are relatively easy to create by disrupting and then fractionating whole cells. The systems hold the advantage of containing virtually all the components needed for protein synthesis except the messenger RNAs. Their disadvantages are that they are relatively short-lived (several hours) and are relatively inefficient in translation. However, an added benefit to the *E. coli* cell-free system is that it will translate, albeit inefficiently, RNAs that normally are not translated in vivo because they lack translation initiation signals. This feature allowed these scientists to use artificial messenger RNAs that contained no Shine-Dalgarno sequence for ribosomal binding.

Nirenberg and Matthei found that when the enzyme polynucleotide phosphorylase in the *E. coli* cell-free system made uridine diphosphates into a poly-U messenger RNA, phenylalanine residues were incorporated into a polypeptide. Thus, the first code word established was UUU for phenylalanine. Nirenberg and Ochoa and their associates continued the work. They found that AAA was the code word for lysine, CCC was the code word for proline, and GGG was the code word for glycine.

They then made synthetic messenger RNAs by using mixtures of the various diphosphate nucleotides in known proportions. Table 11.3 gives an example. From their experiments, it was possible to determine the bases used in many of the code words, but not their specific order. For example, cysteine, leucine, and valine are all coded by two Us and a G, but the experiment could not sort out the order of these bases (5′-UUG-3′, 5′-UGU-3′, or 5′-GUU-3′) for any one of them. Determining the order required an extra step in sophistication—that is, being able to synthesize known trinucleotides.

Synthetic Codons

Once trinucleotides of known composition could be manufactured, Nirenberg and P. Leder in 1964 developed a "binding assay." They found that isolated *E. coli* ribosomes, in the presence of high-molarity magnesium chlo-

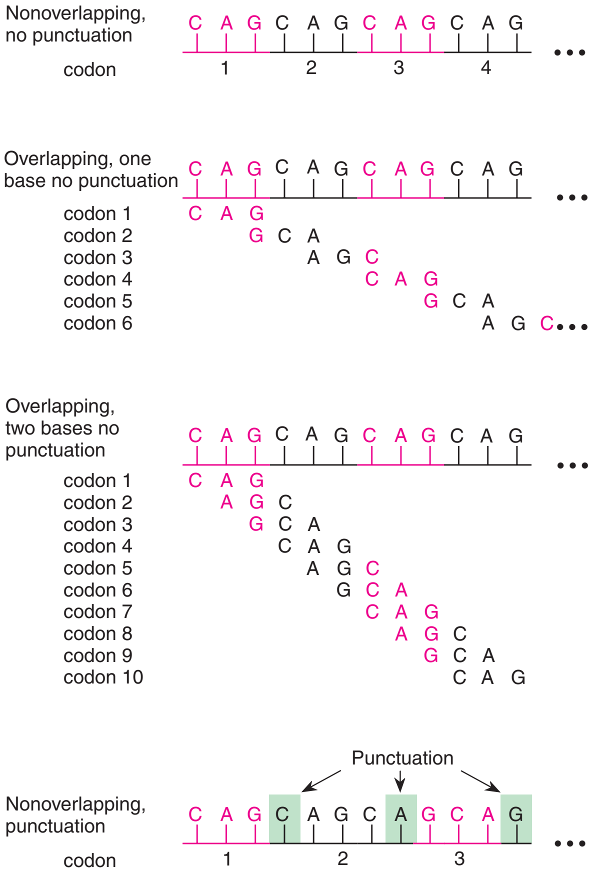


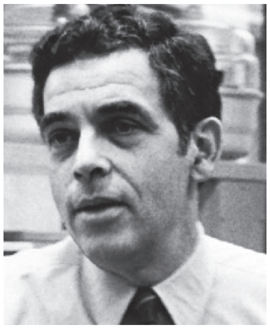
Figure 11.30 The genetic code is read as a nonoverlapping code with no punctuation (*top*). Before that was proven, it was suggested that the code could overlap by one or two bases (*middle*) or have noncoded bases (punctuation) between code words (*bottom*).

Table 11.3 **Structure of Artificial mRNA Made by Randomly Assembling Uracil- and Guanine-Containing Ribose Diphosphate Nucleotides with a Ratio of 5U:1G**

Codon	Frequency of Occurrence
UUU	$(5/6)^3 = 0.58$
UUG	$(5/6)^2(1/6) = 0.12$
UGU	$(5/6)^2(1/6) = 0.12$
GUU	$(5/6)^2(1/6) = 0.12$
UGG	$(5/6)(1/6)^2 = 0.02$
GUG	$(5/6)(1/6)^2 = 0.02$
GGU	$(5/6)(1/6)^2 = 0.02$
GGG	$(1/6)^3 = 0.005$

ride, could bind trinucleotides as if they were messenger RNAs. Also bound was the transfer RNA that carried the anticodon complementary to the trinucleotide. It was thus possible, using radioactive amino acids, to determine which messenger RNA trinucleotide coded for a particular amino acid. A given synthetic trinucleotide was mixed with ribosomes and aminoacyl-tRNAs, including one radioactively labeled amino acid. The reaction mixture was passed over a filter that would allow everything except the large trinucleotide + ribosome + aminoacyl-tRNA complex to pass through. If the radioactivity passed through the filter, it meant that the radioactive amino acid was not associated with the ribosome. The experiment was then repeated with another labeled amino acid. When the radioactivity appeared on the filter, the investigators knew that the amino acid was affiliated with the ribosome. Thus, that amino acid was coded by the selected trinucleotide codon. In other words, the radioactive amino acid was attached to a transfer RNA whose anticodon was complementary to the trinucleotide codon and thus bound at the ribosome.

Figure 11.31 shows an example. In the figure, the trinucleotide is 5′-CUG-3′. The transfer RNA with the anticodon 3′-GAC-5′ is charged with leucine. The mixture is passed through a filter. If threonine, or any other amino acid except leucine, is radioactive, the radioactivity passes through the filter. When the experiment is repeated with radioactive leucine, the leucine, and hence the radioactivity, is trapped by the filter. In a short period of time, all of the codons were deciphered (table 11.4).



Phillip Leder (1934–).
(Courtesy of Dr. Phillip Leder.)

Wobble Hypothesis

The genetic code is a **degenerate code,** meaning that a given amino acid may have more than one codon. As you can see from table 11.4, eight of the sixteen boxes contain just one amino acid per box. (A box is determined by the first and second positions; e.g., the UUX box, in which X is any of the four bases.) Therefore, for these eight amino acids, the codon need only be read in the first two positions because the same amino acid will be represented regardless of the third base of the codon. These eight groups of codons are termed **unmixed families** of codons. An unmixed family is the four codons beginning with the same two bases that specify a single amino acid. For example, the codon family GUX codes for valine. **Mixed families** code for two amino acids or for stop signals and one or two amino acids.

Six of the mixed-family boxes are split in half so that the codons are differentiated by the presence of a purine or a pyrimidine in the third base. For example, CAU and

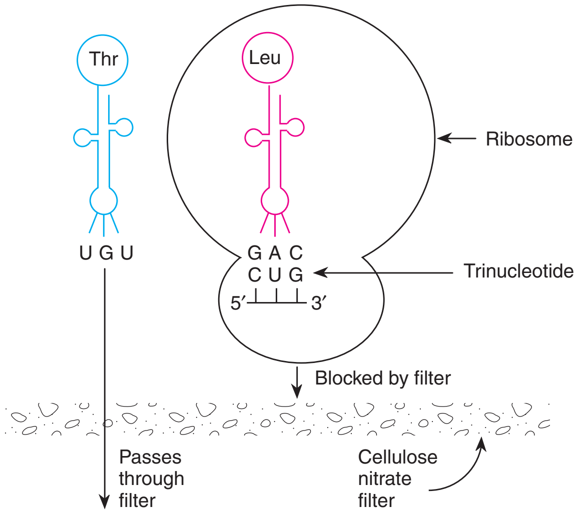


Figure 11.31 The binding assay determines the amino acid associated with a given trinucleotide codon. Transfer RNAs with noncomplementary codons pass through the membrane. Transfer RNAs with anticodons complementary to the trinucleotide bind to the ribosome and do not pass through the filter. When the transfer RNA is charged with a radioactive amino acid, the radioactivity is trapped on the filter.

Table 11.4 **The Genetic Code**

	Second Position				
First Position (5′ End)	**U**	**C**	**A**	**G**	**Third Position (3′ End)**
U	Phe	Ser	Tyr	Cys	**U**
	Phe	Ser	Tyr	Cys	**C**
	Leu	Ser	*stop*	*stop*	**A**
	Leu	Ser	*stop*	Trp	**G**
C	Leu	Pro	His	Arg	**U**
	Leu	Pro	His	Arg	**C**
	Leu	Pro	Gln	Arg	**A**
	Leu	Pro	Gln	Arg	**G**
A	Ile	Thr	Asn	Ser	**U**
	Ile	Thr	Asn	Ser	**C**
	Ile	Thr	Lys	Arg	**A**
	Met (*start*)	Thr	Lys	Arg	**G**
G	Val	Ala	Asp	Gly	**U**
	Val	Ala	Asp	Gly	**C**
	Val	Ala	Glu	Gly	**A**
	Val	Ala	Glu	Gly	**G**

CAC both code for histidine; in both, the third base, U (uracil) or C (cytosine), is a pyrimidine. Only two of the families of codons are split differently.

The lesser importance of the third position in the genetic code ties in with two facts about transfer RNAs. First, although there would seem to be a need for sixty-two transfer RNAs—since there are sixty-one codons specifying amino acids and an additional codon for initiation—there are actually only about fifty different transfer RNAs in an *E. coli* cell. Second, a rare base such as inosine can appear in the anticodon, usually in the position that is complementary to the third position of the codon. These two facts lead researchers to believe that some kind of conservation of transfer RNAs is occurring and that rare bases may be involved.

We should mention, to avoid confusion, that both messenger RNA and transfer RNA bases are usually numbered from the 5′ side. Thus, the number-one base of the codon is complementary to the number-three base of the anticodon (fig. 11.32). Thus, the codon base of lesser importance is the number-three base, whereas its complement in the anticodon is the number-one base.

Since the first position of the anticodon (5′) is not as constrained as the other two positions, a given base at that position may be able to pair with any of several bases in the third position of the codon. Crick characterized this ability as **wobble** (fig. 11.33). Table 11.5 shows the possible pairings that would produce a transfer RNA system compatible with the known code. For example, if an isoleucine transfer RNA has the anticodon 3′-UAI-5′, it is compatible with the three codons for that amino acid (see table 11.4): 5′-AUU-3′, 5′-AUC-3′, and 5′-AUA-3′. That is, inosine in the first (5′) position of the anticodon can recognize U, C, or A in the third (3′) position of the codon, and thus one transfer RNA complements all three codons for isoleucine.

Universality of the Genetic Code

Until 1979, scientists concluded that the genetic code was universal. That is, the codon dictionary (see table 11.4) was the same for *E. coli,* human beings, and oak trees, as well as all other species studied up to that time. The universality of the code was demonstrated, for example, by taking the ribosomes and messenger RNA from rabbit reticulocytes and mixing them with the aminoacyl-tRNAs and other translational components of *E. coli.* Rabbit hemoglobin was synthesized.

In 1979 and 1980, however, researchers noted discrepancies when sequencing mitochondrial genes for struc-

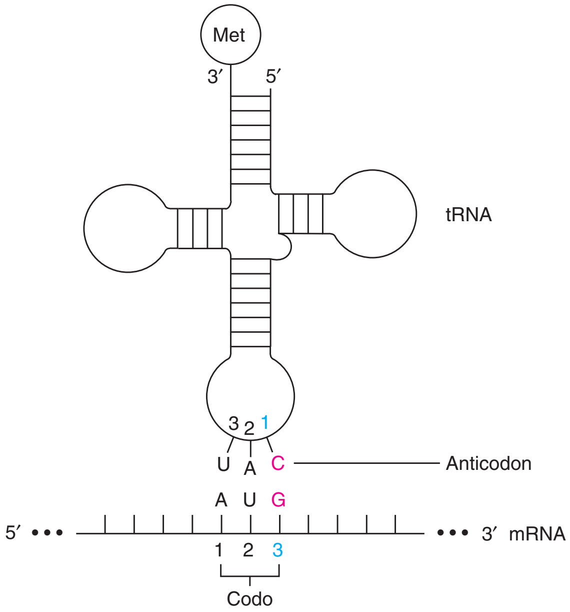


Figure 11.32 Codon and anticodon base positions are numbered from the 5′ end. The 3′ position in the codon (5′ in the anticodon) is the wobble base.

tural proteins (see chapters 13 and 17). It was discovered that there were two kinds of deviations from universality in the way mitochondrial transfer RNAs read the code. First, fewer transfer RNAs were needed to read the code. Second, there were several instances in which the mitochondrial and cellular systems interpreted a codon differently.

According to Crick's wobble rules (see table 11.5), thirty-two transfer RNAs (including one for initiation) can complement all sixty-one nonterminating codons. Unmixed families require two transfer RNAs, and mixed families require one, two, or three transfer RNAs, depending on the family. The yeast mitochondrial coding system apparently needs only twenty-four transfer RNAs. The reduction in numbers is accomplished primarily by having only one transfer RNA recognize each unmixed family (table 11.6; cf. table 11.4). Because mitochondrial transfer RNAs for unmixed families of codons have a U in the first (wobble) position of the anticodon, apparently, given the structure of the mitochondrial transfer RNAs, the U can pair with U, C, A, or G. Presumably, evolutionary pressure has minimized the number of transfer RNA genes in the DNA of the mitochondrion, in keeping with its small size. Reduction from thirty-two to twenty-four is a 25% savings. (Recent evidence suggests that mammalian mitochondria may need only twenty-two transfer RNAs.)

Table 11.5 **Pairing Combinations at the Third Codon Position**

Number-one Base in tRNA (5′ End)	Number-three Base in mRNA (3′ End)
G	U *or* C
C	G
A	U
U	A *or* G
I	A, U, *or* C

It has also been found that yeast mitochondria read the CUX family as threonine rather than as leucine (tables 11.4 and 11.6) and the terminator UGA (opal) as tryptophan rather than as termination. However, there appear to be differences among different groups of organisms reading the CUX family. Human and *Neurospora* mitochondria appear to read the CUX codons as leucine, just as cellular systems do. Of the groups so far analyzed, only yeast reads the CUX family as threonine. Similarly, human and *Drosophila* mitochondria read AGA and AGG as stop signals rather than as arginine (table 11.7).

In 1985, it was discovered that *Paramecium* species read the UAA and UAG stop codons as glutamine within the cell. In addition, a prokaryote (*Mycoplasma capricolum*) reads UGA as tryptophan. We do not yet know how general this finding is: scientists have scrutinized the genetic code of very few species. We can thus conclude that the genetic code seems to have universal tendencies among prokaryotes, eukaryotes, and viruses. Mitochondria, however, read the code slightly differently: different wobble rules apply, and mitochondria and cells read at least one terminator and one unmixed family of codons differently. Also, the mitochondrial discrepancies are not universal among all types of mitochondria. Further work, involving the sequencing of more mitochondrial DNAs, should elucidate the pattern of discrepancies among the mitochondria of diverse species. We also now know that not every organism reads all codons in the same way. Ciliated protozoa and a mycoplasma read some stop signals as coding for amino acids. Nuclear variants are known in the following codons: CUG, AUA, UAA, UAG, UGA, CGG, and AGA. Mitochondrial variants are known in CUX, AUA, UAA, UAG, AAA, UGA, CGX, AGA, and AGG.

One other type of variation of codon reading occurs: *site-specific variation,* in which the interpretation of a codon depends on its specific location. We are already familiar with the fact that GUG and, rarely, UUG can serve as prokaryotic initiation codons. This means that they are recognized by $\text{tRNA}_\text{f}^\text{Met}$. However, they are not recognized by $\text{tRNA}_\text{m}^\text{Met}$ (i.e., GUG and UUG are not misread internally in messenger RNAs). In some cases, two of the termination codons (UGA and UAG, but not UAA) are misinterpreted as codons for amino acids. That is, termination will not occur at the normal place, resulting in a longer-than-usual protein. In some cases, these "read-through" proteins are vital—the

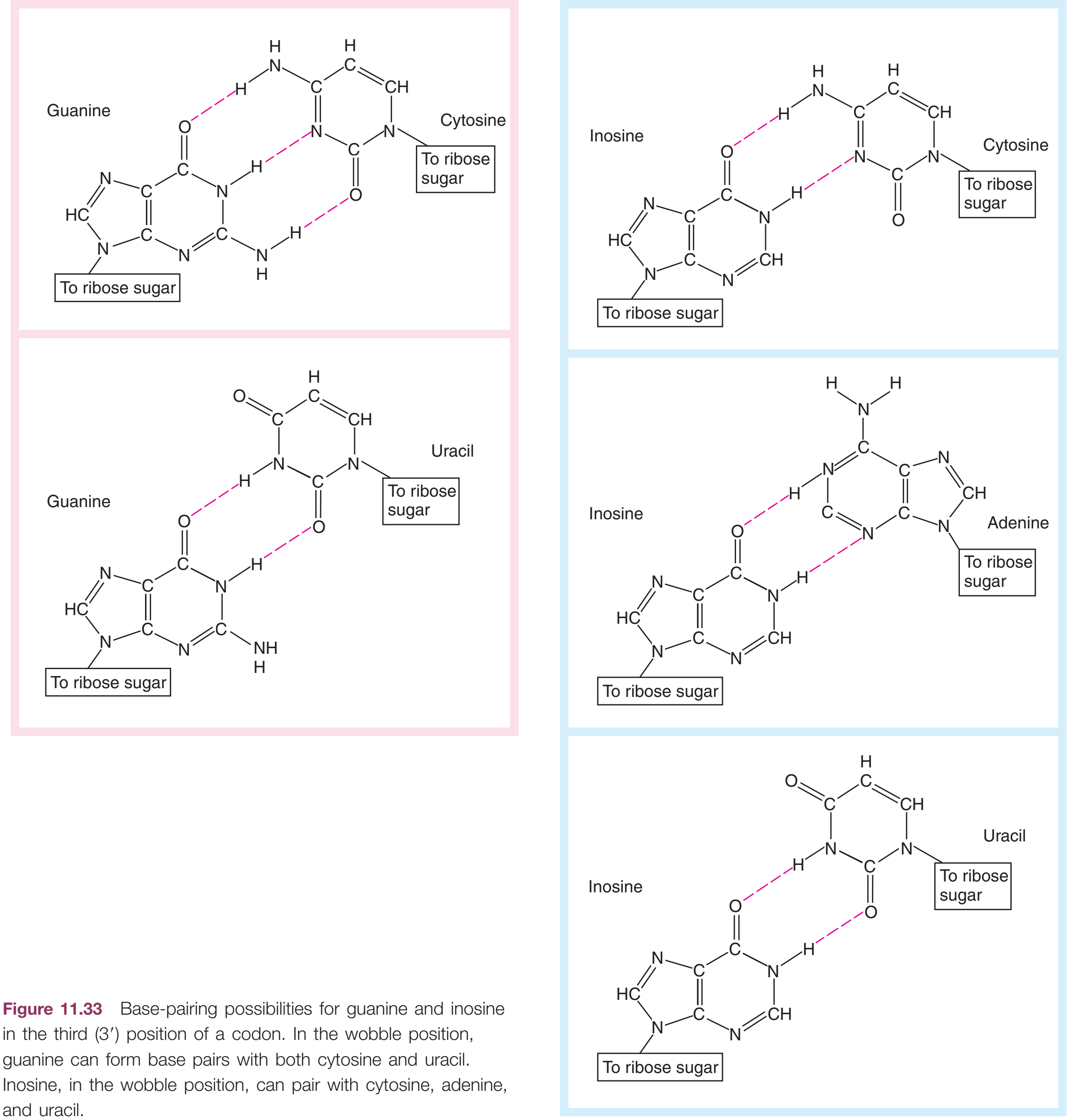


Figure 11.33 Base-pairing possibilities for guanine and inosine in the third (3′) position of a codon. In the wobble position, guanine can form base pairs with both cytosine and uracil. Inosine, in the wobble position, can pair with cytosine, adenine, and uracil.

organism depends on their existence. For example, in the phage Qβ, the coat-protein gene is read through about 2% of the time. Without this small number of read-through proteins, the phage coat cannot be constructed properly.

One last example of site-specific variation involves the amino acid selenocysteine (cysteine with a selenium atom replacing the sulfur; see fig. 11.1). Although many proteins have unusual amino acids, almost all are due to posttranslational modifications of normal amino acids. However, the amino acid selenocysteine is inserted directly into some proteins, such as formate dehydrogenase in *E. coli,* which has selenium in its active site. Selenocysteine is inserted into the protein by a novel transfer RNA that recognizes the termination codon, UGA, if that codon is involved in a particular stem-loop secondary structure in the messenger RNA. The selenocysteine transfer RNA is originally charged with a serine that is then modified to a selenocysteine. In addition to the

Table 11.6 The Genetic Code Dictionary of Yeast Mitochondria*

	Second Position				
First Position (5′ End)	**U**	**C**	**A**	**G**	**Third Position (3′ End)**
U	Phe AAG	Ser AGU	Tyr AUG	Cys ACG	**U**
					C
	Leu AAU		*stop*	Trp ACU	**A**
					G
C	Thr GAU	Pro GGU	His GUG	Arg GCA	**U**
					C
			Gln GUU		**A**
					G
A	Ile UAG	Thr UGU	Asn UUG	Ser UCG	**U**
					C
	Met UAC		Lys UUU	Arg UCU	**A**
					G
G	Val CAU	Ala CGU	Asp CUG	Gly CCU	**U**
					C
			Glu CUU		**A**
					G

Source: Data from S. Bonitz, et al., "Codon recognition rules in yeast mitochondria," *Proceedings of the National Academy of Sciences* 77:3167–70, 1980.

* Anticodons (3′ → 5′) are given within boxes. (The ACU Trp anticodon is predicted.)

Table 11.7 Common and Alternative Meanings of Codons

Codon	General Meaning	Alternative Meaning
CUX	Leu	Thr in yeast mitochondria
AUA	Ile	Met in mitochondria of yeast, *Drosophila,* and vertebrates
UGA	*Stop*	Trp in mycoplasmas and mitochondria other than higher plants
AGA/AGG	Arg	*Stop* in mitochondria of yeast and vertebrates
		Ser in mitochondria of *Drosophila*
CGG	Arg	Trp in mitochondria of higher plants
UAA/UAG	*Stop*	Gln in ciliated protozoa
UAG	*Stop*	Ala or Leu in mitochondria of some higher plants

stem-loop structure 3′ (downstream) from the amber codon (UAG), a selenocysteine elongation factor (SELB) is also needed at the ribosome. This same mechanism may occur in eukaryotes, but not all of the components have yet been identified.

Evolution of the Genetic Code

It has been theorized that the genetic code has wobble in it because it originally arose from a code in which only the first two bases were needed for the small number of amino acids in use several billion years ago. As new amino acids with useful properties became available, they were incorporated into proteins by a code modified by the third base, albeit with less specificity. This view has support from the fact that codons starting with the same nucleotide come from the same biosynthetic pathway. This indicates that in early evolution, as biosynthetic pathways were extended to new amino acids, the newcomers were incorporated by use of the second and third bases of the code.

However, the question remains as to whether the genetic code is highly evolved or just a "frozen accident." In other words, is there a relationship between the codons and the amino acids they code for, or is the code just one of many random possibilities? Recent computer simulations of random codes indicate that the current genetic code is far outside the range of random in its ability to protect the organism from mutation. This suggests that the genetic code is not a frozen accident, but rather is highly evolved. Numerous examples in the current code support this view.

For example, in the unmixed codon family 5′-CUX-3′, any mutation in the third position produces another codon for the same amino acid. Wobble in the third position and codon arrangement ensures that less than half of the mutations in the third codon position result in the specification of a different amino acid.

There are also patterns in the genetic code in which the mutation of one codon to another results in an amino acid of similar properties. A high probability exists that such a mutation will produce a functional protein. All the codons with U as the middle base, for example, are for amino acids that are hydrophobic (phenylalanine, leucine, isoleucine, methionine, and valine). Mutation in the first or third positions for any of these codons still codes a hydrophobic amino acid. Both of the two negatively charged amino acids, aspartic acid and glutamic acid, have codons that start with GA. All of the aromatic amino acids—phenylalanine, tyrosine, and tryptophan (see fig. 11.1)—have codons that begin with uracil. Such patterns minimize the negative effects of mutation.

This chapter completes the discussion of the mechanics of gene expression. The next chapter deals with recombinant DNA technology, followed by several chapters concerned with the control of gene expression in both prokaryotes and eukaryotes.

SUMMARY

STUDY OBJECTIVE 1: To study the mechanism of protein biosynthesis, in which organisms, using the information in DNA, string together amino acids to form proteins 281–303

A charged transfer RNA has an anticodon at one end and a specific amino acid at the other end. The transfer RNAs are charged with the proper amino acid by aminoacyl-tRNA synthetase enzymes that incorporate the energy of ATP into amino acid-tRNA bonds. Hence, no additional source of energy is needed during peptide bond formation. During protein synthesis, the translation apparatus at the ribosome recognizes the transfer RNA. Through complementarity, the anticodon pairs with a messenger RNA codon.

An initiation complex forms at the start of translation. In prokaryotes, this complex consists of the messenger RNA, the 30S subunit of the ribosome, the initiator transfer RNA with N-formyl methionine (fMet-tRNA_f^{Met}), and the initiation factors IF1, IF2, and IF3. The 50S ribosomal subunit is then added and A and P sites form in the resulting 70S ribosome. The charged N-formyl methionine transfer RNA is in the P site. A GTP is hydrolyzed, and the initiation factors are released.

A transfer RNA enters the A site, which requires the involvement of elongation factors EF-Ts and EF-Tu (in *E. coli*). At least one GTP hydrolysis releases the elongation factor, EF-Tu, which had originally brought the charged transfer RNA to the ribosome. Peptidyl transferase, which appears to be a ribozymic component of the 50S ribosomal subunit, transfers the amino acid from the transfer RNA in the P site to the amino end of the amino acid on the transfer RNA in the A site.

With the help of elongation factor G (EF-G), the ribosome translocates in relation to the messenger RNA. The depleted transfer RNA is moved from the P site to the E site, where it is released; the transfer RNA with the growing peptide is moved into the P site. EF-G is then released. Elongation and translocation continue until a nonsense codon enters the A site. With the aid of the release factors RF1 and RF2, the protein is released, and the messenger RNA-ribosome complex dissociates. Eukaryotes have slightly more complex processes involving several more proteins.

Proteins pass through membranes with the help of a signal peptide synthesized at their N-terminal ends. Proteins fold into their final, functional configurations with the help of molecular chaperones, proteins that aid the folding process.

Many antibiotics interfere with translation in prokaryotes. Puromycin, streptomycin, tetracycline, and chloramphenicol all act at the ribosome. Studying the mode of action of these antibiotics has provided insights into the mechanism of the translation process.

STUDY OBJECTIVE 2: To examine the genetic code 304–312

The genetic code was first assumed to be triplet because of logical arguments regarding the minimum size of codons. With his work on deletion and insertion mutants, Crick provided evidence that the code was triplet. Part of the code was worked out initially with the synthesis of long, artificial messenger RNAs and then the synthesis of specific trinucleotide codons. Crick's wobble hypothesis accounts for the fact that fewer than sixty-one transfer RNAs can read the entire genetic code. Fewer transfer RNAs are needed because additional complementary base pairings occur in the third position (3′) of the codon.

The rule of universality of the genetic code has to be modified in light of findings regarding mitochondrial transfer RNAs; only twenty-four are needed to read the code. In addition, some sense codons are interpreted differently in mitochondrial systems; some nonmitochondrial systems read stop codons differently (a mycoplasma and ciliated protozoan); and some site-specific variation in codon reading also occurs. The structure of the code in both cells and mitochondria seems to protect the cell against a good deal of potential mutation.

SOLVED PROBLEMS

PROBLEM 1: What is the energy requirement of protein biosynthesis?

Answer: The cost of adding one amino acid to a growing polypeptide is four or five high-energy bonds: two from an ATP during the charging of the transfer RNA, and two or three from the hydrolysis of GTPs during transfer RNA binding to the A site of the ribosome and during translocation. Thus, for an average protein of three hundred amino acids, there is a cost of 1,200 to 1,500 high-energy bonds.

PROBLEM 2: What are the start and stop signals of translation?

Answer: Once a messenger RNA is attached at the ribosome, the start signal is the methionine initiation codon (usually AUG), whereas the stop signal is one of the three nonsense codons (UAA, UAG, and UGA). Binding to the ribosome in order to position the messenger RNA in relation to the A and P sites differs in prokaryotes and eukaryotes. In prokaryotes, the Shine-Dalgarno sequence allows the 16S ribosomal RNA and the messenger RNA to form hydrogen bonds, locating the beginning of the messenger RNA at the ribosome. In eukaryotes, the 5′ cap is usually recognized by the ribosome, and the ribosome then proceeds to scan the messenger RNA for the initiation codon.

PROBLEM 3: What amino acids could replace methionine if a one-base mutation occurred?

Answer: The codon for methionine (internal as well as initiation) is AUG. If the A is replaced, we would get UUG (Leu), CUG (Leu), or GUG (Val); if the U is replaced, we would get AAG (Lys), ACG (Thr), or AGG (Arg); and if the G is replaced, we would get AUA (Ile), AUU (Ile), or AUC (Ile). Hence, a one-base change in the codon for methionine could result in any of six different amino acids.

EXERCISES AND PROBLEMS*

INFORMATION TRANSFER

1. Given the following end part of a gene, which will be transcribed and then translated into a pentapeptide, provide the base sequence for its messenger RNA. Give the anticodons on the transfer RNAs by making use of wobble rules. What amino acids are incorporated? Draw the actual structure of the pentapeptide.

   3′-TACAATGGCCCTTTTATC-5′
   5′-ATGTTACCGGGAAAATAG-3′

2. Give an alternative translation mechanism that would require only one transfer RNA site on the ribosome.
3. Draw the details of a moment in time at the ribosome during the translation of the messenger RNA produced in problem 1. Include in the diagram the ribosomal sites, the transfer RNAs, and the various nonribosomal proteins involved.
4. How do prokaryotic and eukaryotic ribosomes recognize the 5′ end of messenger RNAs? Could eukaryotic messenger RNAs be polycistronic?
5. How many aminoacyl-tRNA synthetases are there? What do they use for recognition signals?
6. What are the similarities and differences among the three nonsense codons? Using the wobble rules, what are their theoretical anticodons?
7. Describe an experiment that demonstrates that the transfer RNA, and not its amino acid, is recognized at the ribosome during translation.
8. Other than the antibiotics named in the chapter, suggest five "theoretical" antibiotics that could interfere with the prokaryotic translation process.
9. How many single-base deletions are required to restore the reading frame of a messenger RNA? Give an example.
10. A "nonsense mutation" is one in which a codon for an amino acid changes to one for chain termination. Give an example. What are its consequences?
11. The reverse situation to problem 10 is a mutation from a nonsense codon to a codon for an amino acid. Give an example. What are its consequences?
12. What are the consequences when an internal methionine codon recognizes a prokaryotic initiation transfer RNA?
13. What role does EF-Ts play in elongation? EF-Tu? What are their eukaryotic equivalents?
14. What roles do RF1 and RF2 play in chain termination? What are their eukaryotic equivalents?
15. What is a signal peptide? What role does it play in eukaryotes? What is its fate?
16. Why doesn't puromycin disrupt eukaryotic translation?

---

*Answers to selected exercises and problems are on page A-12.

17. A peptide, fifteen amino acids long, is digested by two methods, and each segment is sequenced according to the Edman degradation technique (see box 11.1). The fifteen amino acids are denoted by the letters A through O, with F as the N-terminal amino acid. If the segments are as follows, what is the sequence of the original peptide?

Method 1: CABHLN; FGKI; OEDJM

Method 2: KICAB; JM; FG; HLNOED

18. In human hemoglobin, the β chain is 146 amino acids long. What is the minimum length of RNA needed to make this protein?

19. Part of a DNA strand to be transcribed has the following sequence:

3′-TACTAACTTACGCTCGCCTCA-5′

- **a.** What is the sequence of RNA transcribed from this part of the strand?
- **b.** What sequence of amino acids does the RNA produce?

THE GENETIC CODE

20. If DNA contained only the bases cytosine and guanine, how long would a code word have to be? How could we tell if this DNA were double-stranded?

21. If an artificial messenger RNA contains two parts uracil to one of cytosine, name the amino acids and the proportions in which they should be incorporated into protein.

22. What would be proved or disproved if an organism were discovered that did not follow any of the rules of the codon dictionary? Would we expect organisms from another galaxy (if they exist) to use our codon dictionary?

23. What would the genetic code dictionary (see table 11.4) look like if wobble occurred in the second position rather than the third (i.e., if an unmixed family of codons were of the form GXU)?

24. In experiments using repeating polymers, $(GCGC)_n$ incorporates alanine and arginine into polypeptides, and $(CGGCGG)_n$ incorporates arginine, glycine, and alanine. What codon can probably be assigned to glycine?

25. If poly-G is used as a messenger RNA in an incorporation experiment, glycine is incorporated into a polypeptide. If poly-C is used, proline is incorporated. If both poly-G and poly-C are used, no amino acids are incorporated into protein. Why?

26. A protein has leucine at a particular position. If the codon for leucine is CUC, how many different amino acids might appear as the result of a single-base substitution?

27. Polymers of $(GUA)_n$ result in the incorporation of only two different amino acids rather three, as for most other three-base polymers. Why?

28. The sixth amino acid in the β chain of normal human hemoglobin is glutamate. Two different mutations of this codon substitute valine and lysine. What is the likely codon for glutamate?

29. A normal protein has the following C-terminal amino acid sequence: *ser-thr-lys-leu*-COOH. A mutant is isolated with the following sequence: *ser-thr-lys-leu-leu-phe-arg*-COOH. What has probably happened to produce the mutant protein?

30. A segment of a normal protein and three different mutants appears as follows:

normal	_____gly-ala-ser-his-cys-leu-phe_____
mutant 1	_____gly-ala-ser-his
mutant 2	_____gly-ala-ser-leu-cys-leu-phe_____
mutant 3	_____gly-val-ala-ile-ala-ser_____

What is the probable sequence of bases in the normal RNA?

31. A normal protein has histidine in a given position. Four mutants are isolated and determined to have either tyrosine, glutamine, proline, or leucine in place of histidine. What are the possible codon assignments, and what codon is probably used for histidine?

CRITICAL THINKING QUESTIONS

1. How do transcriptional and translational signals interact?

2. Why is the E site necessary in translation?

Suggested Readings for chapter 11 are on page B-8.