polymerase under standard conditions is 8.0×10^{-6} (per base duplicated), whereas for Pfu polymerase it is only 1.3×10^{-6}. To supply the commercial demand for thermostable DNA polymerases, the genes for these enzymes have been cloned into *E. coli*, allowing the enzymes to be produced in large quantities. www.microbiologyplace.com **Online Tutorial 6.2: Polymerase Chain Reaction (PCR)**

Applications and Sensitivity of PCR

PCR is a powerful tool. It is easy to perform, extremely sensitive and specific, and highly efficient. During each round of amplification the amount of product doubles, leading to an exponential increase in the DNA. This means not only that a large amount of amplified DNA can be produced in just a few hours, but that only a few molecules of target DNA need be present in the sample to start the reaction. The reaction is so specific that, with primers of 15 or so nucleotides and high annealing temperatures, there is almost no "false priming," and therefore the PCR product is virtually homogeneous.

PCR is extremely valuable for obtaining DNA for cloning genes or for sequencing purposes because the gene or genes of interest can easily be amplified if flanking sequences are known. PCR is also used routinely in comparative or phylogenetic studies to amplify genes from various sources. In these cases the primers are made for regions of the gene that are conserved in sequence across a wide variety of organisms. Because 16S rRNA, a molecule used for phylogenetic analyses, has both highly conserved and highly variable regions, primers specific for the 16S rRNA gene from various taxonomic groups can be synthesized. These may be used to survey different groups of organisms in any specific habitat. This technique is in widespread use in microbial ecology and has revealed the enormous diversity of the microbial world, much of it not yet cultured (↻ Section 22.5).

Because it is so sensitive, PCR can be used to amplify very small quantities of DNA. For example, PCR has been used to amplify and clone DNA from sources as varied as mummified human remains and fossilized plants and animals. The ability of PCR to amplify and analyze DNA from cell mixtures has also made it a common tool of diagnostic microbiology. For example, if a clinical sample shows evidence of a gene specific to a particular pathogen, then it can be assumed that the pathogen was present in the sample. Treatment of the patient can then begin without the need to culture the organism, a time-consuming and often fruitless process. PCR has also been used in forensics to identify human individuals from very small samples of their DNA.

MiniQuiz

- Why is a primer needed at each end of the DNA segment being amplified by PCR?
- From which organisms are thermostable DNA polymerases obtained?
- How has PCR improved diagnostic clinical medicine?

IV RNA Synthesis: Transcription

Transcription is the synthesis of ribonucleic acid (RNA) using DNA as a template. There are three key differences in the chemistry of RNA and DNA: (1) RNA contains the sugar ribose instead of deoxyribose; (2) RNA contains the base uracil instead of thymine; and (3) except in certain viruses, RNA is not double-stranded. The change from deoxyribose to ribose affects the chemistry of a nucleic acid; enzymes that act on DNA usually have no effect on RNA, and vice versa. However, the change from thymine to uracil does not affect base pairing, as these two bases pair with adenine equally well.

RNA plays several important roles in the cell. Three major types of RNA are involved in protein synthesis: **messenger RNA (mRNA)**, **transfer RNA (tRNA)**, and **ribosomal RNA (rRNA)**. Several other types of RNA also occur that are mostly involved in regulation (Chapter 8). These RNA molecules all result from the transcription of DNA. It should be emphasized that RNA operates at two levels, genetic and functional. At the genetic level, mRNA carries genetic information from the genome to the ribosome. In contrast, rRNA has both a functional and a structural role in ribosomes and tRNA has an active role in carrying amino acids for protein synthesis. Indeed, some RNA molecules including rRNA have enzymatic activity (ribozymes, ↻ Section 7.8). Here we focus on how RNA is synthesized in the *Bacteria*, using *Escherichia coli* as our model organism.

6.12 Overview of Transcription

Transcription is carried out by the enzyme **RNA polymerase**. Like DNA polymerase, RNA polymerase catalyzes the formation of phosphodiester bonds but between ribonucleotides rather than deoxyribonucleotides. RNA polymerase uses DNA as a template. The precursors of RNA are the ribonucleoside triphosphates ATP, GTP, UTP, and CTP. The mechanism of RNA synthesis is much like that of DNA synthesis. During elongation of an RNA chain, ribonucleoside triphosphates are added to the 3′-OH of the ribose of the preceding nucleotide. Polymerization is driven by the release of energy from the two energy-rich phosphate bonds of the incoming ribonucleoside triphosphates. In both DNA replication and RNA transcription the overall direction of chain growth is from the 5′ end to the 3′ end; thus the new strand is antiparallel to the template strand. Unlike DNA polymerase, however, RNA polymerase can initiate new strands of nucleotides on its own; consequently, no primer is necessary.

RNA Polymerases

The template for RNA polymerase is a double-stranded DNA molecule, but only one of the two strands is transcribed for any given gene. Nevertheless, genes are present on both strands of DNA and thus DNA sequences on both strands are transcribed, although at different locations. Although these principles are true for transcription in all organisms, there are significant differences among RNA polymerase from *Bacteria, Archaea*, and *Eukarya*. The following discussion deals only with RNA polymerase from *Bacteria*, which has the simplest structure and about which most is known (RNA polymerase in *Archaea* and *Eukarya* is discussed in Chapter 7).

RNA polymerase from *Bacteria* has five different subunits, designated β, β′, α, ω (omega), and σ (sigma), with α present in two copies. The β and β′ (beta prime) subunits are similar but not identical. The subunits interact to form the active enzyme, called the RNA polymerase holoenzyme, but the sigma factor is not as tightly bound as the others and easily dissociates, leading to the formation of the RNA polymerase core enzyme, $\alpha_2\beta\beta'\omega$. The core enzyme alone synthesizes RNA, whereas the sigma factor recognizes the appropriate site on the DNA for RNA synthesis to begin. The omega subunit is needed for assembly of the core enzyme but not for RNA synthesis. RNA synthesis is illustrated in **Figure 6.25**.

www.microbiologyplace.com **Online Tutorial 6.3: Transcription**

Promoters

RNA polymerase is a large protein and makes contact with many bases of DNA simultaneously. Proteins such as RNA polymerase can interact specifically with DNA because portions of the bases are exposed in the major groove. However, in order to initiate RNA synthesis correctly, RNA polymerase must first recognize the initiation sites on the DNA. These sites, called **promoters**, are recognized by the sigma factor (**Figure 6.26**).

Once the RNA polymerase has bound to the promoter, transcription can proceed. In this process, the DNA double helix at the promoter is opened up by the RNA polymerase to form a transcription bubble. As the polymerase moves, it unwinds the DNA in short segments. This transient unwinding exposes the template strand and allows it to be copied into the RNA complement. Thus, promoters can be thought of as pointing RNA polymerase in one direction or the other along the DNA. If a region of DNA has two nearby promoters pointing in opposite directions, then transcription from one will proceed in one direction (on one of the DNA strands) while transcription from the other promoter will proceed in the opposite direction (on the other strand).

Once a short stretch of RNA has been formed, the sigma factor dissociates. Elongation of the RNA molecule is then carried out by the core enzyme alone (Figure 6.25). Sigma is only needed to form the initial RNA polymerase–DNA complex at the promoter. As the newly made RNA dissociates from the DNA, the opened DNA closes back into the original double helix. Transcription stops at specific sites called transcription terminators (Section 6.14).

Unlike DNA replication, which copies entire genomes, transcription copies much smaller units of DNA, often as little as a single gene. This system allows the cell to transcribe different genes at different frequencies, depending on the needs of the cell for different proteins. In other words, gene expression is regulated. As we shall see in Chapter 8, regulation of transcription is an important and elaborate process that uses many different mechanisms and is very efficient at controlling gene expression and conserving cell resources.

MiniQuiz

- In which direction (5′ → 3′ or 3′ → 5′) along the template strand does transcription proceed?
- What is a promoter? What protein recognizes the promoters in *Escherichia coli*?
- What is the role of the omega subunit of RNA polymerase?

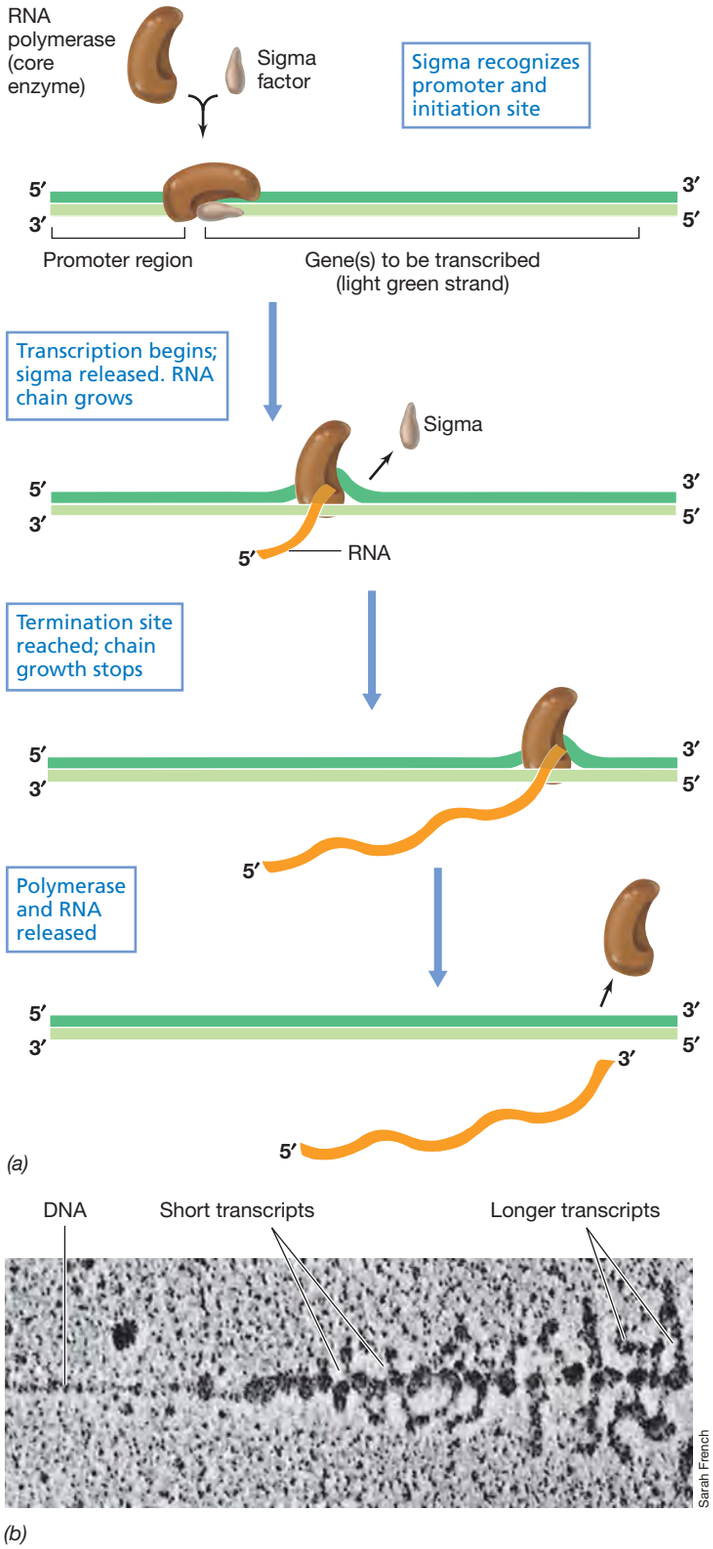

Figure 6.25 Transcription. *(a)* Steps in RNA synthesis. The initiation site (promoter) and termination site are specific nucleotide sequences on the DNA. RNA polymerase moves down the DNA chain, temporarily opening the double helix and transcribing one of the DNA strands. *(b)* Electron micrograph of transcription along a gene on the *Escherichia coli* chromosome. The region of active transcription is about 2 kb pairs of DNA. Transcription is proceeding from left to right, with the shorter transcripts on the left becoming longer as transcription proceeds.

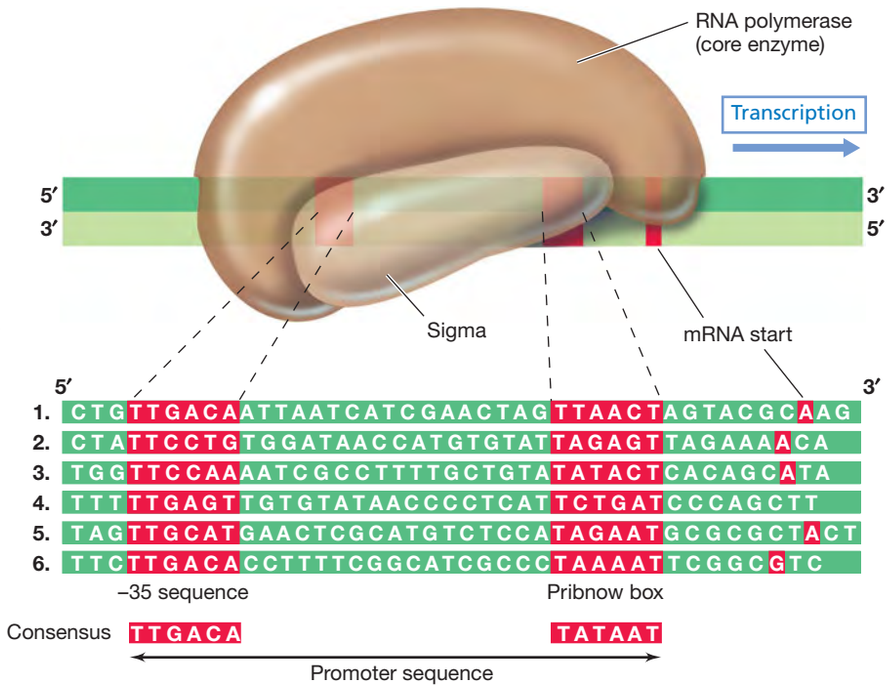

Figure 6.26 The interaction of RNA polymerase with the promoter. Shown below the RNA polymerase and DNA are six different promoter sequences identified in *Escherichia coli*, a species of *Bacteria*. The contacts of the RNA polymerase with the −35 sequence and the Pribnow box (−10 sequence) are shown. Transcription begins at a unique base just downstream from the Pribnow box. Below the actual sequences at the −35 and Pribnow box regions are consensus sequences derived from comparing many promoters. Note that although sigma recognizes the promoter sequences on the 5′ → 3′ (dark green) strand of DNA, the RNA polymerase core enzyme will actually transcribe the light green strand running 3′ → 5′ because core enzyme works only in a 5′ → 3′ direction.

6.13 Sigma Factors and Consensus Sequences

Promoters are specific DNA sequences that bind RNA polymerase. Figure 6.26 shows the sequence of several promoters from *Escherichia coli*. All these sequences are recognized by the same sigma factor, the major sigma factor in *E. coli*, called σ^{70} (the superscript 70 indicates the size of this protein, 70 kilodaltons). Although these sequences are not identical, two shorter sequences within the promoter region are highly conserved, and it is these that sigma recognizes.

Both conserved sequences are upstream of the transcription start site. One is 10 bases before the transcription start, the −10 region, or *Pribnow box*. Although promoters differ slightly, most bases are the same within the −10 region. Comparison of many −10 regions gives the consensus sequence: TATAAT. In our example, each promoter matches from three to five of these bases. The second conserved region is about 35 bases from the start of transcription. The consensus sequence in the −35 region is TTGACA (Figure 6.26). Again, most promoters differ slightly, but are very close to consensus.

In Figure 6.26, six alternative sequences are shown for only one strand of the DNA. This is conventional "shorthand" for writing DNA sequences. By convention, the strand shown is the one with its 5′ end upstream (this is the nontemplate strand for transcription). In reality, RNA polymerase binds to double-stranded DNA and then unwinds it. A single strand of the unwound DNA is then used as template by the RNA polymerase. Although it binds to both DNA strands, sigma makes most of its contacts with the nontranscribed strand where it recognizes the specific sequences in the −10 and −35 regions.

Some sigma factors in other bacteria are much more specific in regard to binding sequences than σ^{70} of *E. coli*. In such cases, very little leeway is allowed in the critical bases that are recognized. In *E. coli*, promoters that are most like the consensus sequence are usually more effective in binding RNA polymerase. Such promoters are called strong promoters and are very useful in genetic engineering, as discussed in Chapter 11.

Alternative Sigma Factors in *Escherichia coli*

Most genes in *E. coli* require the standard sigma factor, σ^{70} or RpoD, for transcription and have promoters like those in Figure 6.26. However, several alternative sigma factors are known that recognize different consensus sequences (**Table 6.4**). Each alternative sigma factor is specific for a group of genes required under special circumstances. Thus σ^{38}, also known as RpoS, recognizes a consensus sequence found in the promoters of genes expressed during stationary phase. Consequently, it is possible to control the expression of each family of genes by regulating the availability of the corresponding sigma factor. This may be done by changing either the rate of synthesis or the rate of degradation of the sigma factor. In addition, the activity of alternative sigma factors can be blocked by other proteins called *anti-sigma factors*. These may temporarily inactivate a particular sigma factor in response to environmental signals.

In total there are seven different sigma factors in *E. coli*, and each recognizes different consensus sequences (Table 6.4). Sigma factors were originally named according to their molecular weight. More recently, they have been named according to their roles, for example, RpoN stands for "*R*NA *po*lymerase—*N*itrogen." Most of these sigma factors have counterparts in other *Bacteria*. The endospore-forming bacterium *Bacillus subtilis* has 14 sigma factors, with 4 different sigma factors dedicated to the transcription of endospore-specific genes (↷ Section 8.12).

Table 6.4 *Sigma factors in* Escherichia coli

Name[a]	*Upstream recognition sequence*[b]	*Function*
σ^{70} RpoD	TTGACA	For most genes, major sigma factor for normal growth
σ^{54} RpoN	TTGGCACA	Nitrogen assimilation
σ^{38} RpoS	CCGGCG	Stationary phase, plus oxidative and osmotic stress
σ^{32} RpoH	TNTCNCCTTGAA[c]	Heat shock response
σ^{28} FliA	TAAA	For genes involved in flagella synthesis
σ^{24} RpoE	GAACTT	Response to misfolded proteins in periplasm
σ^{19} FecI	AAGGAAAAT	For certain genes in iron transport

[a]Superscript number indicates size of protein in kilodaltons. Many factors also have other names, for example, σ^{70} is also called σ^{D}.
[b]N = any nucleotide.

MiniQuiz

- What is a consensus sequence?
- To what parts of the promoter region does sigma bind?
- How are families of genes required during specialized conditions controlled as a group using sigma factors?

6.14 Termination of Transcription

Only those genes that need to be expressed should be transcribed. Therefore it is important to terminate transcription at the correct position. **Termination** of RNA synthesis is governed by specific base sequences on the DNA. In *Bacteria* a common termination signal on the DNA is a GC-rich sequence containing an inverted repeat with a central nonrepeating segment (Section 6.2). When such a DNA sequence is transcribed, the RNA forms a stem–loop structure by intra-strand base pairing (**Figure 6.27**). Such stem–loop structures, followed by a run of adenosines in the DNA template and therefore a run of uridines in the mRNA, are effective transcription terminators. This is due to the formation of a stretch of U:A base pairs that holds the RNA and DNA template together. This structure is very weak as U:A base pairs have only two hydrogen bonds each. The RNA polymerase pauses at the stem–loop, and the DNA and RNA come apart at the run of uridines. This terminates transcription. Sequence patterns that terminate transcription without the intervention of any extra factors are referred to as intrinsic terminators.

The other mechanism for transcription termination uses a specific protein factor, known in *Escherichia coli* as Rho. Rho does not bind to RNA polymerase or to the DNA, but binds tightly to RNA and moves down the chain toward the RNA polymerase–DNA complex. Once RNA polymerase has paused at a Rho-dependent termination site (a specific sequence on the DNA template), Rho causes both the RNA and RNA polymerase to be released from the DNA, thus terminating transcription. Although

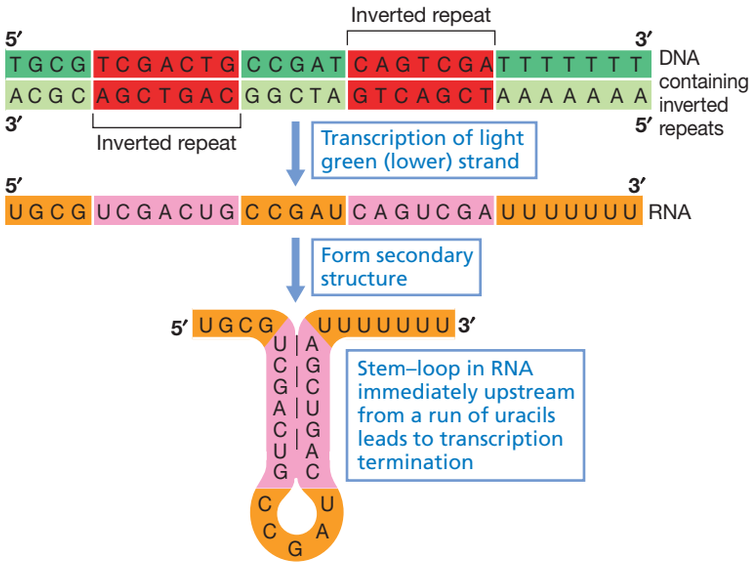

Figure 6.27 Inverted repeats and transcription termination. Inverted repeats in transcribed DNA form a stem–loop structure in the RNA that terminates transcription when followed by a run of uracils.

the termination sequences function at the level of RNA, remember that RNA is transcribed from DNA. Consequently, transcription termination is ultimately determined by specific nucleotide sequences on the DNA.

MiniQuiz

- What is a stem–sloop structure?
- What is an intrinsic terminator?
- How does Rho protein terminate transcription?

6.15 The Unit of Transcription

Genetic information on chromosomes is organized into transcription units. These are segments of DNA that are transcribed into a single RNA molecule. Each transcription unit is bounded by sites where transcription is initiated and terminated. Some units of transcription include only a single gene. Others contain two or more genes. These genes are said to be cotranscribed, yielding a single RNA molecule.

Ribosomal and Transfer RNAs and RNA Longevity

Most genes encode proteins, but others encode nontranslated RNAs, such as ribosomal RNA or transfer RNA. There are several different types of rRNA in an organism. Prokaryotes have three types: 16S rRNA, 23S rRNA, and 5S rRNA (with a ribosome having one copy of each; Section 6.19). As shown in **Figure 6.28**, transcription units exist that contain one gene for each of these rRNAs, and these genes are therefore cotranscribed. The situation is similar in eukaryotes. Therefore, in all organisms the unit of transcription for most rRNA is longer than a single gene. In prokaryotes tRNA genes are often cotranscribed with each other or even, as shown in Figure 6.28, with genes for rRNA.

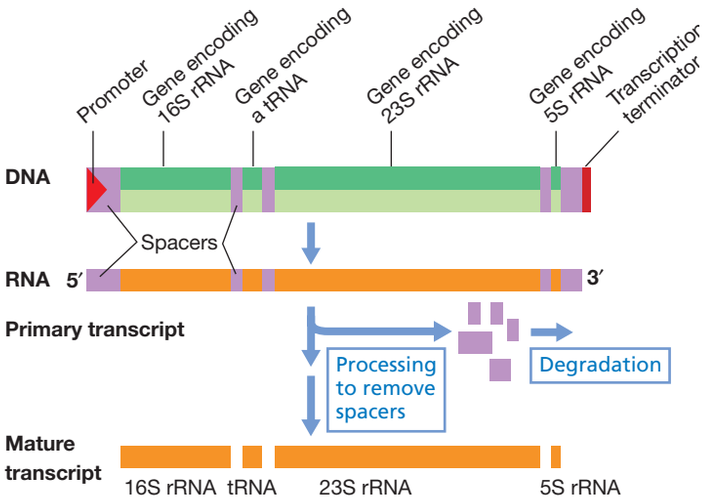

Figure 6.28 A ribosomal rRNA transcription unit from *Bacteria* and its subsequent processing. In *Bacteria* all rRNA transcription units have the genes in the order 16S rRNA, 23S rRNA, and 5S rRNA (shown approximately to scale). Note that in this particular transcription unit the spacer between the 16S and 23S rRNA genes contains a tRNA gene. In other transcription units this region may contain more than one tRNA gene. Often one or more tRNA genes also follow the 5S rRNA gene and are cotranscribed. *Escherichia coli* contains seven rRNA transcription units.

These cotranscribed transcripts must be processed by cutting into individual units to yield mature (functional) rRNAs or tRNAs. Overall, RNA processing is rare in prokaryotes but common in eukaryotes, as we will see later (Chapter 7).

In prokaryotes, most messenger RNAs have a short half-life (on the order of a few minutes), after which they are degraded by cellular ribonucleases. This is in contrast to rRNA and tRNA, which are stable RNAs. This stability is due to tRNAs and rRNAs forming highly folded structures that prevent them from being degraded by ribonucleases. By contrast, normal mRNA does not form such structures and is susceptible to ribonuclease attack. The rapid turnover of prokaryotic mRNAs permits the cell to quickly adapt to new environmental conditions and halt translation of messages whose products are no longer needed.

Polycistronic mRNA and the Operon

In prokaryotes, genes encoding related enzymes are often clustered together. RNA polymerase proceeds through such clusters and transcribes the whole group of genes into a single, long mRNA molecule. An mRNA encoding such a group of cotranscribed genes is called a *polycistronic mRNA*. When this is translated, several polypeptides are synthesized, one after another, by the same ribosome.

A group of related genes that are transcribed together to give a single polycistronic mRNA is known as an **operon**. Assembling genes for the same biochemical pathway or genes needed under the same conditions into an operon allows their expression to be coordinated. Despite this, eukaryotes do not have operons and polycistronic mRNA (Chapter 7). Often, transcription of an operon is controlled by a specific region of the DNA found just upstream of the protein-coding region of the operon. This is considered in more detail in Chapter 8.

MiniQuiz

- What is the role of messenger RNA (mRNA)?
- What is a transcription unit?
- What is a polycistronic mRNA?
- What are operons and why are they useful to prokaryotes?

V Protein Structure and Synthesis

6.16 Polypeptides, Amino Acids, and the Peptide Bond

Proteins play major roles in cell function. Two major classes of proteins are *catalytic* proteins (enzymes) and *structural* proteins. **Enzymes** are the catalysts for chemical reactions that occur in cells. Structural proteins are integral parts of the major structures of the cell: membranes, walls, ribosomes, and so on. Regulatory proteins control most cell processes by a variety of mechanisms, including binding to DNA. However, all proteins show certain basic features in common.

Proteins are polymers of **amino acids**. All amino acids contain an amino group ($-NH_2$) and a carboxylic acid group (–COOH) that are attached to the α-carbon (**Figure 6.29*a***). Linkages between the carboxyl carbon of one amino acid and the amino nitrogen of a second (with elimination of water) are known as **peptide bonds** (**Figure 6.30**). Two amino acids bonded by peptide linkage constitute a dipeptide; three amino acids, a tripeptide; and so on. When many amino acids are linked they form a **polypeptide**. A protein consists of one or more polypeptides. The number of amino acids differs greatly from one protein to another, from as few as 15 to as many as 10,000.

Each amino acid has a unique side chain (abbreviated R). These vary considerably, from as simple as a hydrogen atom in the amino acid glycine to aromatic rings in phenylalanine, tyrosine, and tryptophan (Figure 6.29*b*). Amino acids exist as pairs of **enantiomers**. These are optical isomers that have the same molecular and structural formulas, except that they are mirror images and are designated as either D or L, depending on whether a pure solution rotates light to the right or left, respectively. Natural proteins employ L-amino acids only. Nevertheless, D-amino acids are occasionally found in cells, most notably in the cell wall polymer peptidoglycan (Section 3.6) and in certain peptide antibiotics (Section 26.9). Cells can interconvert certain enantiomers by enzymes called *racemases*.

The chemical properties of an amino acid are governed by its side chain. Amino acids with similar chemical properties are grouped into related "families" (Figure 6.29*b*). For example, the side chain may contain a carboxylic acid group, as in aspartic acid or glutamic acid, rendering the amino acid acidic. Others contain additional amino groups, making them basic. Several amino

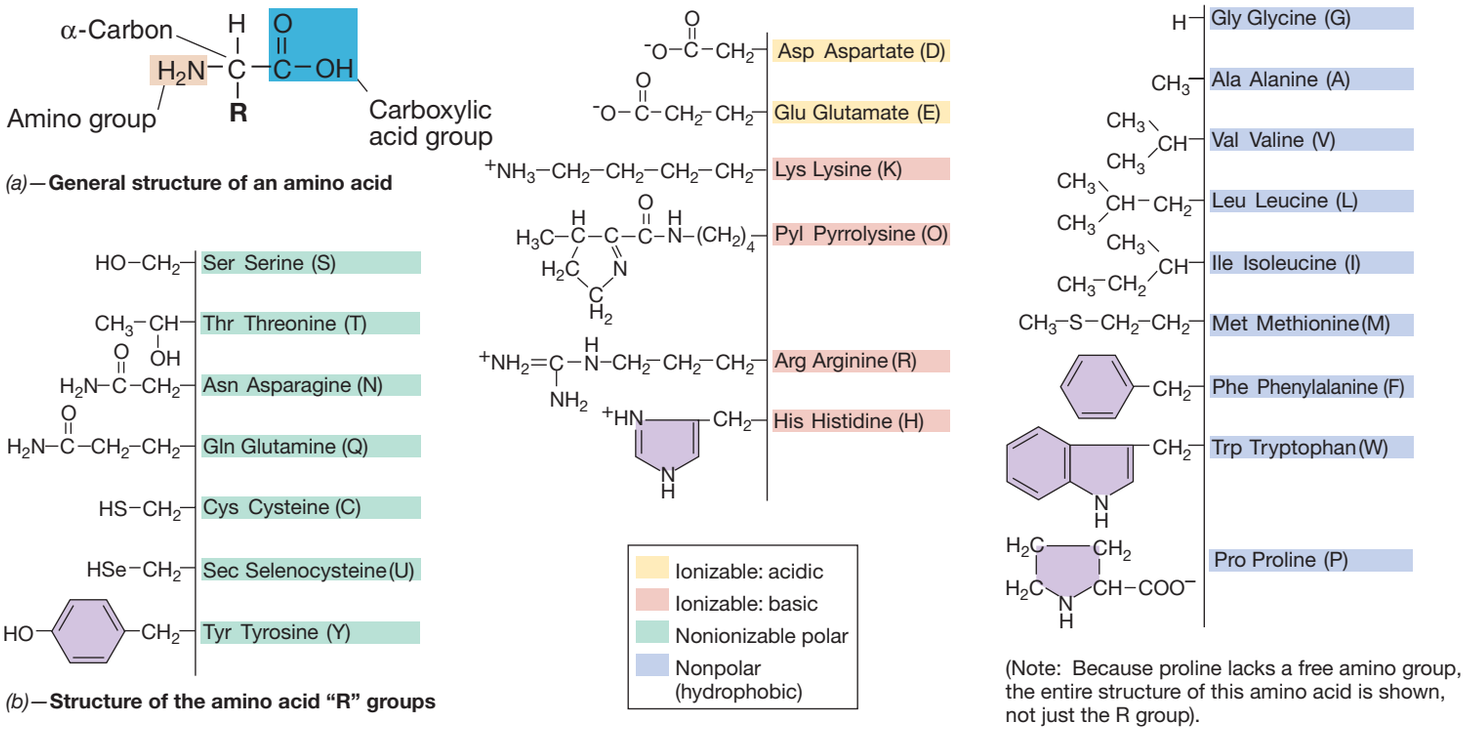

Figure 6.29 **Structure of the 22 genetically encoded amino acids.** *(a)* General structure. *(b)* R group structure. The three-letter codes for the amino acids are to the left of the names, and the one-letter codes are in parentheses to the right of the names. Pyrrolysine has thus far been found only in certain methanogenic *Archaea* (↻ Section 19.3).

acids contain hydrophobic side chains and are known as nonpolar amino acids. Cysteine contains a sulfhydryl group (–SH). Using their sulfhydryl groups, two cysteines can form a disulfide linkage (R–S–S–R) that connects two polypeptide chains.

The diversity of chemically distinct amino acids makes possible an enormous number of unique proteins with widely different biochemical properties. If one assumes that an average polypeptide contains 300 amino acids, there are 22^{300} different polypeptide sequences that are theoretically possible. No cell has anywhere near this many different proteins. In practice, a cell of *Escherichia coli* contains around 2000 different kinds of proteins.

The linear sequence of amino acids in a polypeptide is the **primary structure**. This, ultimately, determines the further folding of the polypeptide, which in turn determines the biological activity. The two ends of a polypeptide are designated as the "C-terminus" and "N-terminus" depending on whether a free carboxylic acid group or a free amino group is found (Figure 6.30).

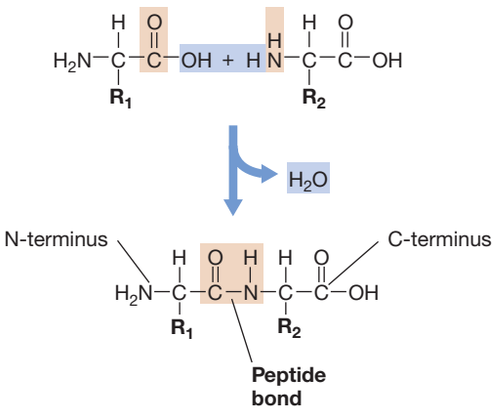

Figure 6.30 **Peptide bond formation.** R_1 and R_2 refer to the side chains of the amino acids. Note that, following peptide bond formation, a free OH group is present at the C-terminus for formation of the next peptide bond.

MiniQuiz

- What chemical groups do amino acids contain?
- Draw the structure of a dipeptide containing the amino acids alanine and tyrosine. Outline the peptide bond.
- Which enantiomeric form of amino acids is found in proteins?
- Glycine does not have two different enantiomers; why?

6.17 Translation and the Genetic Code

In the first two steps in biological information transfer, replication and transcription, nucleic acids are synthesized on nucleic acid templates. The last step, **translation**, also uses a nucleic acid as template, but in this case the product is a protein rather than a nucleic acid. The heart of biological information transfer is the correspondence between the nucleic acid template and the amino acid sequence of the polypeptide product. This is known

Table 6.5 *The genetic code as expressed by triplet base sequences of mRNA*

Codon	Amino acid	Codon	Amino acid	Codon	Amino acid	Codon	Amino acid
UUU	Phenylalanine	UCU	Serine	UAU	Tyrosine	UGU	Cysteine
UUC	Phenylalanine	UCC	Serine	UAC	Tyrosine	UGC	Cysteine
UUA	Leucine	UCA	Serine	UAA	None (stop signal)	UGA	None (stop signal)
UUG	Leucine	UCG	Serine	UAG	None (stop signal)	UGG	Tryptophan
CUU	Leucine	CCU	Proline	CAU	Histidine	CGU	Arginine
CUC	Leucine	CCC	Proline	CAC	Histidine	CGC	Arginine
CUA	Leucine	CCA	Proline	CAA	Glutamine	CGA	Arginine
CUG	Leucine	CCG	Proline	CAG	Glutamine	CGG	Arginine
AUU	Isoleucine	ACU	Threonine	AAU	Asparagine	AGU	Serine
AUC	Isoleucine	ACC	Threonine	AAC	Asparagine	AGC	Serine
AUA	Isoleucine	ACA	Threonine	AAA	Lysine	AGA	Arginine
AUG (start)[a]	Methionine	ACG	Threonine	AAG	Lysine	AGG	Arginine
GUU	Valine	GCU	Alanine	GAU	Aspartic acid	GGU	Glycine
GUC	Valine	GCC	Alanine	GAC	Aspartic acid	GGC	Glycine
GUA	Valine	GCA	Alanine	GAA	Glutamic acid	GGA	Glycine
GUG	Valine	GCG	Alanine	GAG	Glutamic acid	GGG	Glycine

[a]AUG encodes *N*–formylmethionine at the beginning of polypeptide chains of *Bacteria*.

as the **genetic code**. A triplet of three bases called a *codon* encodes each specific amino acid. The 64 possible codons (four bases taken three at a time = 4^3) of mRNA are shown in **Table 6.5**. The genetic code is written as RNA rather than as DNA because it is mRNA that is translated. Note that in addition to the codons for amino acids, there are also specific codons for starting and stopping translation.

Properties of the Genetic Code

There are 22 amino acids that are encoded by the genetic information carried on mRNA. (A variety of others are created by modification of these after translation.) Consequently, because there are 64 codons, many amino acids are encoded by more than one codon. Although knowing the codon at a given location unambiguously identifies the corresponding amino acid, the reverse is not true. Knowing the amino acid does not mean that the codon at that location is known. A code such as this that lacks one-to-one correspondence between "word" (that is, the amino acid) and code is called a degenerate code. However, knowing the DNA sequence and the correct reading frame, one can specify the amino acid sequence of a protein. This permits the determination of amino acid sequences from DNA base sequences and is at the heart of genomics (Chapter 12). In most cases where multiple codons encode the same amino acid, the multiple codons are closely related in base sequence (Table 6.5).

A codon is recognized by specific base pairing with a complementary sequence of three bases called the **anticodon**, which is found on tRNAs. If this base pairing were always the standard pairing of A with U and G with C, then at least one specific tRNA would be needed to recognize each codon. In some cases, this is true. For instance, there are six different tRNAs in *Escherichia coli* for the amino acid leucine, one for each codon (Table 6.5). By contrast, some tRNAs can recognize more than one codon. Thus, although there are two lysine codons in *E. coli*, there is only one lysyl tRNA whose anticodon can base-pair with either AAA or AAG. In these special cases, tRNA molecules form standard base pairs at only the first two positions of the codon while tolerating irregular base pairing at the third position. This phenomenon is called **wobble** and is illustrated in **Figure 6.31**, where a pairing between G and U (rather than G with C) is illustrated at the wobble position.

Stop and Start Codons

A few codons do not encode any amino acid (Table 6.5). These codons (UAA, UAG, and UGA) are the **stop codons**, and they signal the termination of translation of a protein-coding sequence on the mRNA. Stop codons are also called **nonsense codons**, because they interrupt the "sense" of the growing polypeptide when they terminate translation.

A coding sequence on messenger RNA is translated beginning with the **start codon (AUG)**, which encodes a chemically modified methionine, *N*-formylmethionine. Although AUG at the beginning of a coding region encodes *N*-formylmethionine, AUG

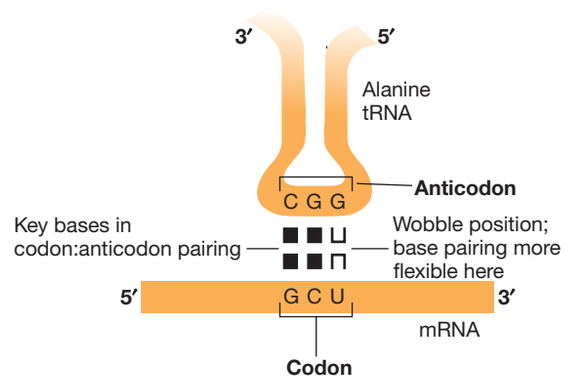

Figure 6.31 **The wobble concept.** Base pairing is more flexible for the third base of the codon than for the first two. Only a portion of the tRNA is shown here.

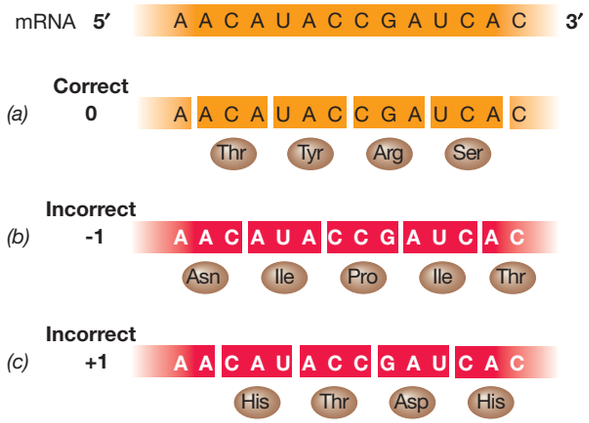

Figure 6.32 **Possible reading frames in an mRNA.** An interior sequence of an mRNA is shown. *(a)* The amino acids that would be encoded if the ribosome is in the correct reading frame (designated the "0" frame). *(b)* The amino acids that would be encoded by this region of the mRNA if the ribosome were in the −1 reading frame. *(c)* The amino acids that would be encoded if the ribosome were in the +1 reading frame.

within the coding region encodes methionine. Two different tRNAs are involved in this process (Section 6.19). With a triplet code it is critical for translation to begin at the correct nucleotide. If it does not, the whole reading frame of the mRNA will be shifted and thus an entirely different protein will be made. If the shift introduces a stop codon into the reading frame, the protein will terminate prematurely. By convention the reading frame that is translated to give the protein encoded by the gene is called the 0 frame. As can be seen in **Figure 6.32**, the other two possible reading frames (−1 and +1) do not encode the same amino acid sequence. Therefore it is essential that the ribosome finds the correct start codon to begin translation and, once it has, that it moves down the mRNA exactly three bases at a time. How is the correct reading frame ensured?

Reading frame fidelity is governed by interactions between mRNA and rRNA within the ribosome. Ribosomal RNA recognizes a specific AUG on the mRNA as a start codon with the aid of an upstream sequence in the mRNA called the Shine–Dalgarno sequence. This alignment requirement explains why occasional mRNA from *Bacteria* can use other start codons, such as GUG. However, even these unusual start codons direct the incorporation of *N*-formylmethionine as the initiator amino acid.

Open Reading Frames

One common method of identifying protein-encoding genes is to examine each strand of the DNA sequence for **open reading frames (ORFs)**. Remember that RNA is transcribed from DNA, so that if one knows the sequence of DNA, one also knows the sequence of mRNA that is transcribed from it. If an mRNA can be translated, it contains an open reading frame: a start codon (typically AUG) followed by a number of codons and then a stop codon in the same reading frame as the start codon. In practice, only ORFs long enough to encode a protein of realistic length are accepted as true coding sequences. Although most functional proteins are at least 100 amino acids in length, a few protein hormones and regulatory peptides are much shorter. Consequently, it is not always possible to tell from sequence data alone whether a relatively short ORF is merely due to chance or encodes a genuine, albeit short, protein.

A computer can be programmed using the above guidelines to scan long DNA base sequences to look for open reading frames. In addition to looking for start and stop codons, the search may include promoters and Shine–Dalgarno ribosome-binding sequences as well. The search for ORFs is very important in genomics (Chapter 12). If an unknown piece of DNA has been sequenced, the presence of an ORF implies that it can encode protein.

Codon Bias

Several amino acids are encoded by multiple codons. One might assume that such multiple codons would be used at equal frequencies. However, this is not so, and sequence data show major **codon bias**. In other words, some codons are greatly preferred over others even though they encode the same amino acid. Moreover, this bias is organism-specific. In *E. coli*, for instance, only about 1 out of 20 isoleucine residues in proteins is encoded by the isoleucine codon AUA, the other 19 being encoded by the other isoleucine codons, AUU and AUC (Table 6.5). Codon bias is correlated with a corresponding bias in the concentration of different tRNA molecules. Thus a tRNA corresponding to a rarely used codon will be in relatively short supply.

The origin of codon bias is unclear, but it is easily recognized and may be taken into account in practical uses of gene sequence information. For example, a gene from one organism whose codon usage differs dramatically from that of another may not be translated efficiently if the gene is cloned into the latter using genetic engineering (Chapter 11). This is due to a shortage of the tRNA for codons that are rare in the host but frequent in the cloned gene. However, this problem can be corrected or at least compensated for by genetic manipulation.

Modifications to the Genetic Code

All cells appear to use the same genetic code. Therefore, the genetic code is a universal code. However, this view has been tempered a bit by the discovery that some organelles and a few cells use genetic codes that are slight variations of the "universal" genetic code.

Alternative genetic codes were first discovered in the genomes of animal mitochondria. These modified codes typically use nonsense codons as sense codons. For example, animal (but not plant) mitochondria use the codon UGA to encode tryptophan instead of using it as a stop codon (Table 6.5). Several organisms are known that also use slightly different genetic codes. For example, in the genus *Mycoplasma* (*Bacteria*) and the genus *Paramecium* (*Eukarya*), certain nonsense codons encode amino acids. These organisms simply have fewer nonsense codons because one or two of them are used as sense codons. In a few rare cases, nonsense codons encode unusual amino acids rather than one of the 20 common amino acids (Section 6.20).

MiniQuiz

- What are stop codons and start codons?
- Why is it important for the ribosome to read "in frame"?
- What is codon bias?
- If you were given a nucleotide sequence, how would you find ORFs?

6.18 Transfer RNA

A transfer RNA carries the anticodon that base-pairs with the codon on mRNA. In addition, each tRNA is specific for the amino acid that corresponds to its own anticodon (that is, the cognate amino acid). The tRNA and its specific amino acid are linked by specific enzymes called **aminoacyl-tRNA synthetases**. These ensure that a particular tRNA receives the correct amino acid and must thus recognize both the tRNA and its cognate amino acid.

General Structure of tRNA

There are about 60 different tRNAs in bacterial cells and 100–110 in mammalian cells. Transfer RNA molecules are short, single-stranded molecules that contain extensive secondary structure and have lengths of 73–93 nucleotides. Certain bases and secondary structures are constant for all tRNAs, whereas other parts are variable. Transfer RNA molecules also contain some purine and pyrimidine bases that differ somewhat from the standard bases found in RNA because they are chemically modified. These modifications are made to the bases after transcription. These unusual bases include pseudouridine (ψ), inosine, dihydrouridine (D), ribothymidine, methyl guanosine, dimethyl guanosine, and methyl inosine. The mature and active tRNA also contains extensive double-stranded regions within the molecule. This secondary structure forms by internal base pairing when the single-stranded molecule folds back on itself (**Figure 6.33**).

The structure of a tRNA can be drawn in a cloverleaf fashion, as in Figure 6.33*a*. Some regions of tRNA secondary structure are named after the modified bases found there (the TψC and D loops) or after their functions (anticodon loop and acceptor stem). The three-dimensional structure of a tRNA is shown in Figure 6.33*b*. Note that bases that appear widely separated in the cloverleaf model may actually be much closer together when viewed in three dimensions. This allows some of the bases in one loop to pair with bases in another loop.

The Anticodon and the Amino Acid–Binding Site

One of the key variable parts of the tRNA molecule is the anticodon, the group of three bases that recognizes the codon on the mRNA. The anticodon is found in the anticodon loop (Figure 6.33). The three nucleotides of the anticodon recognize the codon by specifically pairing with its three bases. By contrast, other portions of the tRNA interact with both the rRNA and protein components of the ribsome, nonribosomal translation proteins, and the aminoacyl synthetase enzyme.

At the 3′ end, or acceptor stem, of all tRNAs are three unpaired nucleotides. The sequence of these three nucleotides is always cytosine-cytosine-adenine (CCA), and they are absolutely essential for function. Curiously, in most organisms these three nucleotides are not encoded by the tRNA genes on the chromosome. Instead they are added, one after another, by an enzyme

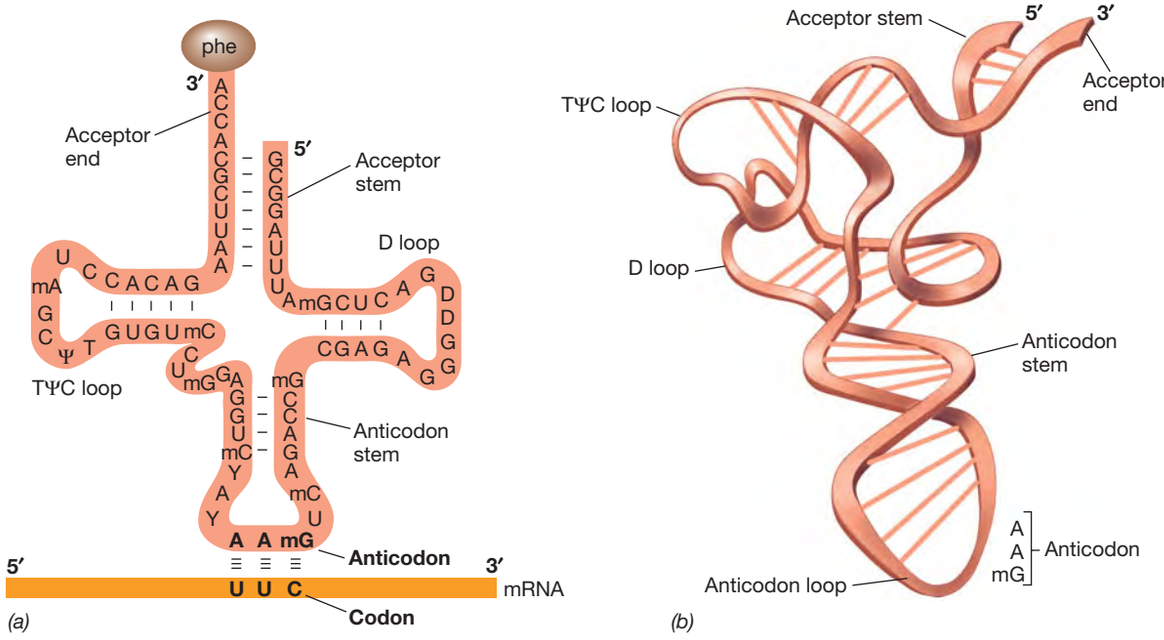

Figure 6.33 **Structure of a transfer RNA.** *(a)* The conventional cloverleaf structural drawing of yeast phenylalanine tRNA. The amino acid is attached to the ribose of the terminal A at the acceptor end. A, adenine; C, cytosine; U, uracil; G, guanine; T, thymine; ψ, pseudouracil; D, dihydrouracil; m, methyl; Y, a modified purine. *(b)* In fact, the tRNA molecule folds so that the D loop and TψC loops are close together and associate by hydrophobic interactions.

called the CCA-adding enzyme, using CTP and ATP as substrates. The cognate amino acid is covalently attached to the terminal adenosine of the CCA end by an ester linkage to the ribose sugar. As we shall see, from this location on the tRNA, the amino acid is incorporated into the growing polypeptide chain on the ribosome by a mechanism described in the next section.

Recognition, Activation, and Charging of tRNAs

Recognition of the correct tRNA by an aminoacyl-tRNA synthetase involves specific contacts between key regions of the tRNA and the synthetase (**Figure 6.34**). As might be expected because of its unique sequence, the anticodon of the tRNA is important in recognition by the synthetase. However, other contact sites between the tRNA and the synthetase are also important. Studies of tRNA binding to aminoacyl-tRNA synthetases, in which specific tRNA bases have been changed by mutation, have shown that only a small number of key nucleotides in tRNA are involved in recognition. These other key recognition nucleotides are often part of the acceptor stem or D loop of the tRNA (Figure 6.33). It should be emphasized that the fidelity of this recognition process is crucial, for if the wrong amino acid is attached to the tRNA, it will be inserted into the growing polypeptide, likely leading to the synthesis of a faulty protein.

The specific reaction between amino acid and tRNA catalyzed by the aminoacyl-tRNA synthetase begins with activation of the amino acid by reaction with ATP:

$$\text{Amino acid} + \text{ATP} \longleftrightarrow \text{aminoacyl—AMP} + \text{P—P}$$

The aminoacyl-AMP intermediate formed normally remains bound to the enzyme until collision with the appropriate tRNA molecule. Then, as shown in Figure 6.34*a*, the activated amino acid is attached to the tRNA to form a charged tRNA:

$$\text{Aminoacyl—AMP} + \text{tRNA} \longleftrightarrow \text{aminoacyl—tRNA} + \text{AMP}$$

The pyrophosphate (PP_i) formed in the first reaction is split by a pyrophosphatase, giving two molecules of inorganic phosphate.

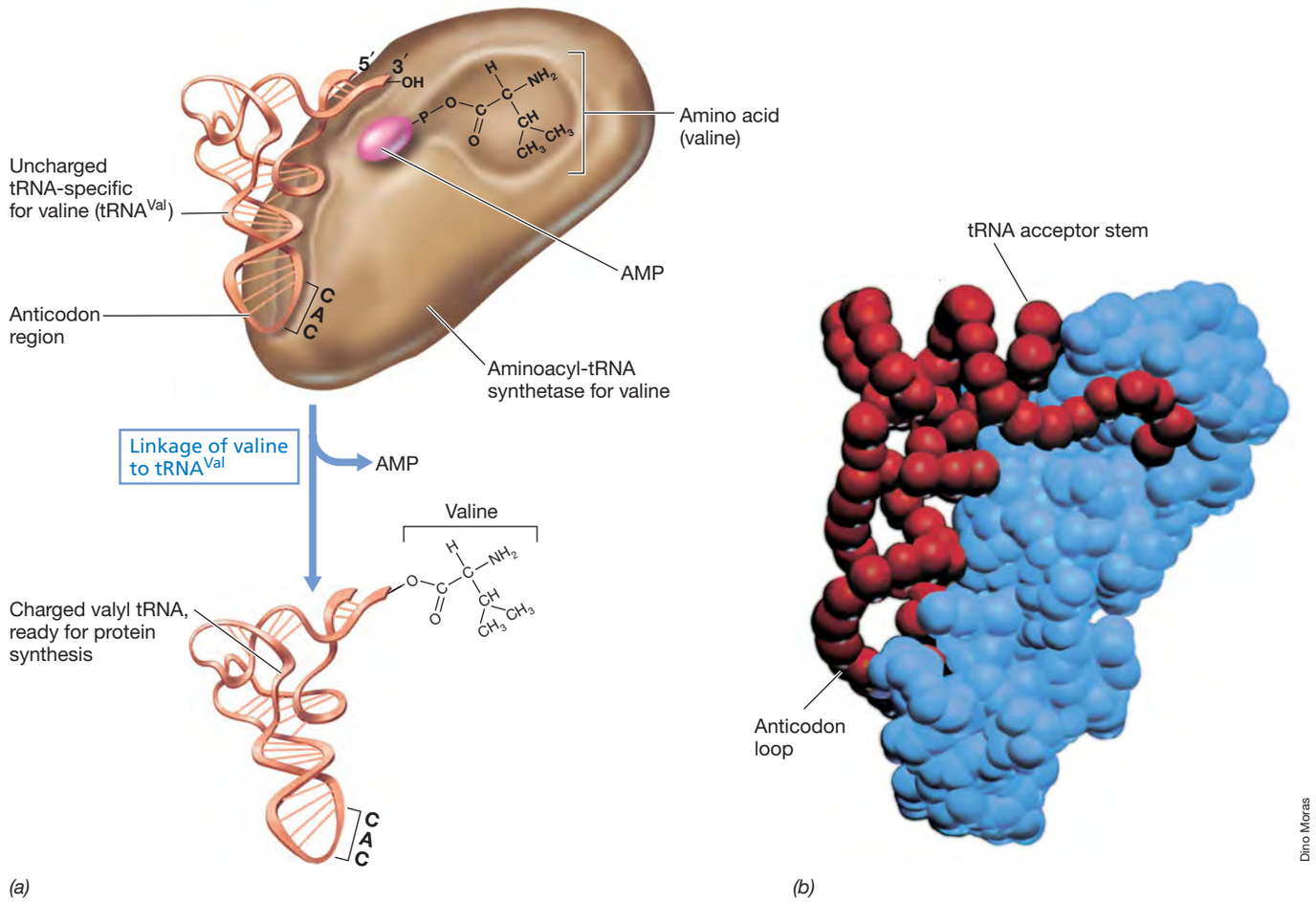

Figure 6.34 **Aminoacyl-tRNA synthetase.** *(a)* Mode of activity of an aminoacyl-tRNA synthetase. Recognition of the correct tRNA by a particular synthetase involves contacts between specific nucleic acid sequences in the D loop and acceptor stem of the tRNA and specific amino acids of the synthetase. In this diagram, valyl-tRNA synthetase is shown catalyzing the final step of the reaction, where the valine in valyl-AMP is transferred to tRNA. *(b)* A computer model showing the interaction of glutaminyl-tRNA synthetase (blue) with its tRNA (red). Reprinted with permission from M. Ruff et al. 1991. *Science* 252: 1682–1689.

Because ATP is used and AMP is formed in these reactions, a total of two energy-rich phosphate bonds are needed to charge a tRNA with its cognate amino acid. After activation and charging, the aminoacyl-tRNA leaves the synthetase and travels to the ribosome where the polypeptide is synthesized.

MiniQuiz

- What is the function of the anticodon of a tRNA?
- What is the function of the acceptor stem of a tRNA?

6.19 Steps in Protein Synthesis

It is vital for proper functioning of proteins that the correct amino acids are inserted at the proper locations in the polypeptide chain. This is the task of the protein-synthesizing machinery, the ribosome. Although protein synthesis is a continuous process, it can be broken down into a number of steps: initiation, elongation, and termination. In addition to mRNA, tRNA, and ribosomes, the process requires a number of proteins designated initiation, elongation, and termination factors. The energy-rich compound guanosine triphosphate (GTP) provides the necessary energy for the process. The key steps in protein synthesis are shown in Figure 6.35.

Ribosomes

Ribosomes are the sites of protein synthesis. A cell may have many thousand ribosomes, the number increasing at higher growth rates. Each ribosome consists of two subunits (**Figure 6.35**). Prokaryotes possess 30S and 50S ribosomal subunits, yielding intact 70S ribosomes. The S-values are Svedberg units, which refer to the sedimentation coefficients of ribosomal subunits (30S and 50S) or intact ribosomes (70S) when subjected to centrifugal force in an ultracentrifuge. (Although larger particles do have larger S-values, the relationship is not linear and S-values cannot be added together.)

Each ribosomal subunit contains specific ribosomal RNAs and ribosomal proteins. The 30S subunit contains 16S rRNA and 21 proteins, and the 50S subunit contains 5S and 23S rRNA and 31 proteins. Thus, in *Escherichia coli*, there are 52 distinct ribosomal proteins, most present at one copy per ribosome. The ribosome is a dynamic structure whose subunits alternately associate and dissociate and also interact with many other proteins. There are several proteins that are essential for ribosome function and interact with the ribosome at various stages of translation. These are regarded as "translation factors" rather than "ribosomal proteins" per se.

Initiation of Translation

In *Bacteria*, such as *E. coli*, initiation of protein synthesis begins with a free 30S ribosomal subunit. From this, an initiation complex forms consisting of the 30S subunit, plus mRNA, formylmethionine tRNA, and several initiation proteins called IF1, IF2, and IF3. GTP is also required for this step. Next, a 50S ribosomal subunit is added to the initiation complex to form the active 70S ribosome. At the end of the translation process, the ribosome separates again into 30S and 50S subunits.

Just preceding the start codon on the mRNA is a sequence of three to nine nucleotides called the Shine–Dalgarno sequence or ribosome-binding site that helps bind the mRNA to the ribosome. The ribosome-binding site is toward the 5′ end of the mRNA and is complementary to base sequences in the 3′ end of the 16S rRNA. Base pairing between these two molecules holds the ribosome–mRNA complex securely together in the correct reading frame. Polycistronic mRNA has multiple Shine–Dalgarno sequences, one upstream of each coding sequence. This allows bacterial ribosomes to translate several genes on the same mRNA because the ribosome can find each initiation site within a message by binding to its Shine–Dalgarno site.

Translational initiation always begins with a special initiator aminoacyl-tRNA binding to the start codon, AUG. In *Bacteria* this is formylmethionyl-tRNA. After polypeptide completion, the formyl group is removed. Consequently, the N-terminal amino acid of the completed protein will be methionine. However, in many proteins this methionine is removed by a specific protease. Because the Shine–Dalgarno sequences (and other possible interactions between the rRNA and the mRNA) direct the ribosome to the proper start site, prokaryotic mRNA can use a start codon other than AUG. The most common alternative start codon is GUG. When used in this context, however, GUG calls for formylmethionine initiator tRNA (and not valine, see Table 6.5).

Elongation, Translocation, and Termination

The mRNA threads through the ribosome primarily bound to the 30S subunit. The ribosome contains other sites where the tRNAs interact. Two of these sites are located primarily on the 50S subunit, and they are termed the A site and the P site (Figure 6.35). The A site, the acceptor site, is the site on the ribosome where the incoming charged tRNA first attaches. Loading of tRNA into the A site is assisted by the elongation factor EF-Tu.

The P site, the peptide site, is the site where the growing polypeptide chain is held by the previous tRNA. During peptide bond formation, the growing polypeptide chain moves to the tRNA at the A site as a new peptide bond is formed. Several nonribosomal proteins are required for elongation, especially the elongation factors, EF-Tu and EF-Ts, as well as more GTP (to simplify Figure 6.35, the elongation factors are omitted and only part of the ribosome is shown).

Following elongation, the tRNA holding the polypeptide is translocated (moved) from the A site to the P site, thus opening the A site for another charged tRNA (Figure 6.35). Translocation requires the elongation factor EF-G and one molecule of GTP for each translocation event. At each translocation step the ribosome advances three nucleotides, exposing a new codon at the A site. Translocation pushes the now empty tRNA to a third site, called the E site. It is from this exit site that the tRNA is actually released from the ribosome. The precision of the translocation step is critical to the accuracy of protein synthesis. The ribosome must move exactly one codon at each step. Although mRNA appears to be moving through the ribosome complex, in reality, the ribosome is moving along the mRNA. Thus, the three sites on the ribosome shown in Figure 6.35 are not static locations but are moving parts of a complex biomolecular machine.

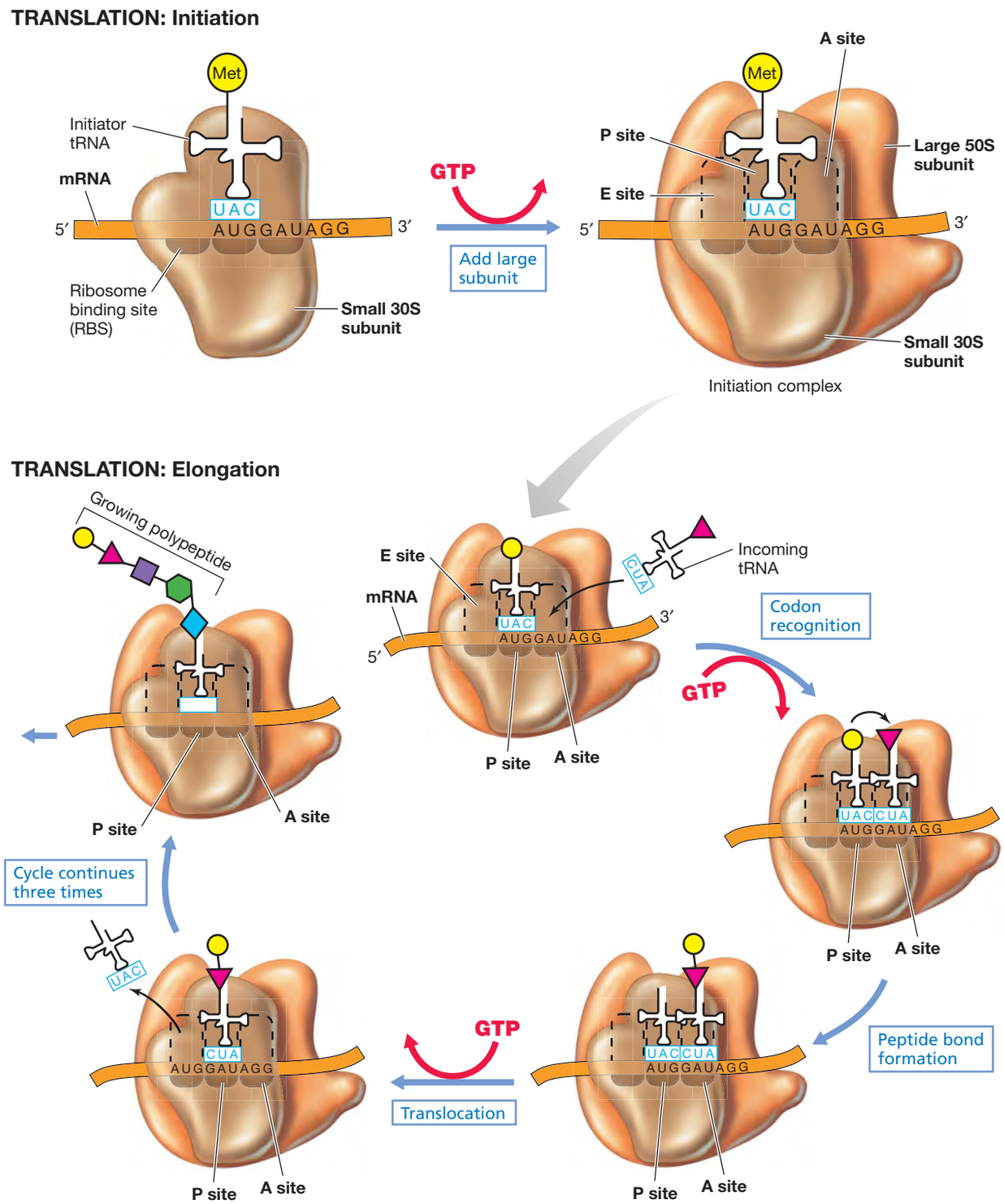

Figure 6.35 The ribosome and protein synthesis. *Initiation* of protein synthesis. The mRNA and initiator tRNA, carrying N-formylmethionine ("Met"), bind first to the small subunit of the ribosome. Initiation factors (not shown) use energy from GTP to promote the addition of the large ribosomal subunit. The initiator tRNA starts out in the P site. *Elongation* cycle of translation. Elongation factors (not shown) use GTP to install the incoming tRNA into the A site. Peptide bond formation is then catalyzed by the 23S rRNA. Translocation of the ribosome along the mRNA from one codon to the next requires hydrolysis of another GTP. The outgoing tRNA is released from the E site. The next charged tRNA binds to the A site and the cycle repeats. The genetic code, expressed in the language of mRNA, is shown in Table 6.5.

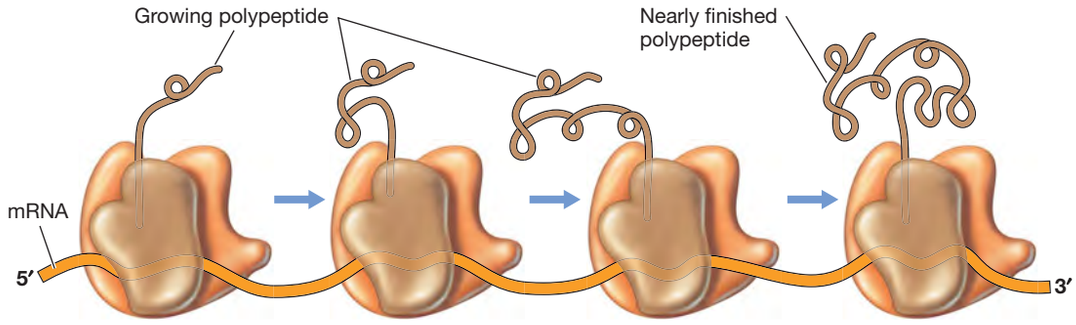

Figure 6.36 **Polysomes.** Translation by several ribosomes on a single messenger RNA forms the polysome. Note how the ribosomes nearest the 5′ end of the message are at an earlier stage in the translation process than ribosomes nearer the 3′ end, and thus only a relatively short portion of the final polypeptide has been made.

Several ribosomes can simultaneously translate a single mRNA molecule, forming a complex called a polysome (**Figure 6.36**). Polysomes increase the speed and efficiency of translation, and because the activity of each ribosome is independent of that of its neighbors, each ribosome in a polysome complex makes a complete polypeptide. Note in Figure 6.36 how ribosomes closest to the 5′ end (the beginning) of the mRNA molecule have short polypeptides attached to them because only a few codons have been read, while ribosomes closest to the 3′ end of the mRNA have nearly finished polypeptides.

Protein synthesis terminates when the ribosome reaches a stop codon (nonsense codon). No tRNA binds to a stop codon. Instead, specific proteins called *release factors* (RFs) recognize the stop codon and cleave the attached polypeptide from the final tRNA, releasing the finished product. Following this, the ribosomal subunits dissociate, and the 30S and 50S subunits are then free to form new initiation complexes and repeat the process.

Role of Ribosomal RNA in Protein Synthesis

Ribosomal RNA plays vital roles in all stages of protein synthesis, from initiation to termination. The role of the many proteins present in the ribosome, although less clear, may be to act as a scaffold to position key sequences in the ribosomal RNAs.

In *Bacteria* it is clear that 16S rRNA is involved in initiation through base pairing with the Shine–Dalgarno sequence on the mRNA. There are also other mRNA–rRNA interactions during elongation. On either side of the codons in the A and P sites, the mRNA is held in position by binding to 16S rRNA and ribosomal proteins. Ribosomal RNA also plays a role in ribosome subunit association, as well as in positioning tRNA in the A and P sites on the ribosome (Figure 6.35). Although charged tRNAs that enter the ribosome recognize the correct codon by codon–anticodon base pairing, they are also bound to the ribosome by interactions of the anticodon stem–loop of the tRNA with specific sequences within 16S rRNA. Moreover, the acceptor end of the tRNA (Figure 6.35) base-pairs with sequences in 23S rRNA.

In addition to all of this, the actual formation of peptide bonds is catalyzed by rRNA. The peptidyl transferase reaction happens on the 50S subunit of the ribosome and is catalyzed by the 23S rRNA itself, rather than by any of the ribosomal proteins. The 23S rRNA also plays a role in translocation, and the EF proteins are known to interact specifically with 23S rRNA. Thus, besides its role as the structural backbone of the ribosome, ribosomal RNA plays a major catalytic role in the translation process as well. www.microbiologyplace.com **Online Tutorial 6.4: Translation**

Freeing Trapped Ribosomes

A defective mRNA that lacks a stop codon causes a problem in translation. Such a defect may arise, for example, from a mutation that removed the stop codon, defective synthesis of the mRNA, or partial degradation of the mRNA. If a ribosome reaches the end of an mRNA molecule and there is no stop codon, release factor cannot bind and the ribosome cannot be released from the mRNA. The ribosome is trapped.

Bacterial cells contain a small RNA molecule, called *tmRNA*, that frees stalled ribosomes (**Figure 6.37**). The "tm" in its name refers to the fact that tmRNA mimics both tRNA, in that it carries the amino acid alanine, and mRNA, in that it contains a short stretch of RNA that can be translated. When tmRNA collides with a stalled ribosome, it binds alongside the defective mRNA. Protein synthesis can then proceed, first by adding the alanine on the tmRNA and then by translating the short tmRNA message. Finally, tmRNA contains a stop codon that allows release factor

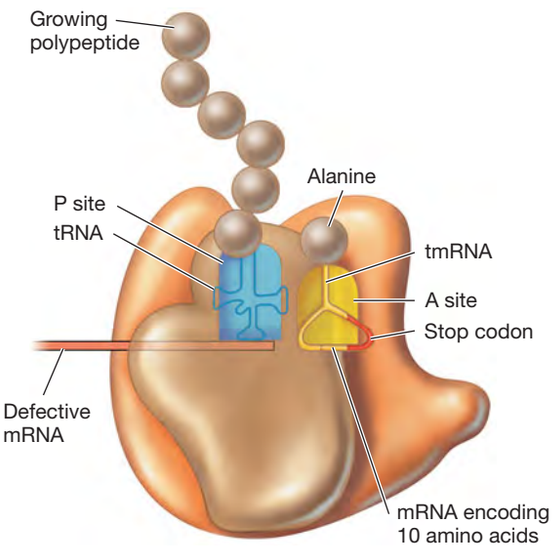

Figure 6.37 **Freeing of a stalled ribosome by tmRNA.** A defective mRNA lacking a stop codon stalls a ribosome that has a partly synthesized polypeptide attached to a tRNA (blue) in the P site. Binding of tmRNA (yellow) in the A site releases the polypeptide. Translation then continues up to the stop codon provided by the tmRNA.

to bind and disassemble the ribosome. The protein made as a result of this rescue operation is defective and is subsequently degraded. The short sequence of amino acids encoded by tmRNA and added to the end of the defective protein is a signal for a specific protease to degrade the protein. Thus, through the activity of tmRNA, stalled ribosomes are freed up to participate in protein synthesis once again.

Effect of Antibiotics on Protein Synthesis

A large number of antibiotics inhibit protein synthesis by interacting with the ribosome. These interactions are quite specific, and many involve rRNA. Some antibiotics are useful research tools because they are specific for different steps in protein synthesis. For instance, streptomycin inhibits initiation, whereas puromycin, chloramphenicol, cycloheximide, and tetracycline inhibit elongation. Several of these antibiotics are clinically useful. Many antibiotics specifically inhibit ribosomes of organisms from only one or two of the phylogenetic domains. For example, chloramphenicol and streptomycin are specific for the ribosomes of *Bacteria* and cycloheximide for ribosomes of *Eukarya*. The mode of action of these and other antibiotics will be discussed in Chapter 26.

MiniQuiz

- What are the components of a ribosome?
- What functional roles does rRNA play in protein synthesis?
- What roles do the initiation and elongation factors play in protein synthesis?
- How is a completed polypeptide chain released from the ribosome?
- How does tmRNA free stalled ribosomes?

6.20 The Incorporation of Selenocysteine and Pyrrolysine

The universal genetic code has codons for 20 amino acids (Table 6.5). However, many proteins contain other amino acids. In fact, more than 100 different amino acids have been found in various proteins. Most of these are made by modifying a standard amino acid after it is incorporated into a protein, a process called *post-translational modification*. However, two nonstandard amino acids are genetically encoded, although in an unusual manner, and are thus inserted during protein synthesis itself. These exceptions are selenocysteine and pyrrolysine, the 21st and 22nd genetically encoded amino acids (Figure 6.29).

Selenocysteine has the same structure as cysteine except it contains selenium instead of sulfur. It is formed by modifying serine after it has been attached to selenocysteine tRNA. Pyrrolysine is a lysine derivative with an extra aromatic ring. Pyrrolysine is fully synthesized and only then attached to pyrrolysyl tRNA.

Both selenocysteine and pyrrolysine are encoded by stop codons (UGA and UAG, respectively). Both have their own tRNAs that contain anticodons that read these stop codons. Both selenocysteine and pyrrolysine also have specific aminoacyl-tRNA synthetases to charge the tRNA with the amino acids. Most stop codons in organisms that use selenocysteine and pyrrolysine do indeed indicate stop. However, occasional stop codons are recognized as encoding selenocysteine or pyrrolysine. For selenocysteine this depends on a recognition sequence just downstream of the special UGA codon. This forms a stem–loop that binds the SelB protein. The SelB protein also binds charged selenocysteine tRNA and brings it to the ribosome when needed. Similarly, pyrrolysine incorporation relies on a recognition sequence just downstream of the pyrrolysine-encoding UAG codon.

Selenocysteine and pyrrolysine are both relatively rare. *Escherichia coli* makes only a handful of proteins with selenocysteine, including two different formate dehydrogenase enzymes. It was sequencing the genes for these enzymes that led to the discovery of selenocysteine. Most organisms, including plants and animals, have a few proteins that contain selenocysteine. Pyrrolysine is rarer still. It has been found in certain *Archaea* and *Bacteria* but was first discovered in species of methanogenic *Archaea*, organisms that generate methane (Section 19.3). In certain methanogens the enzyme methylamine methyltransferase contains a pyrrolysine residue. Whether there are yet other genetically encoded amino acids is unlikely but remains a possibility.

MiniQuiz

- Explain the term posttranslational modification.
- What specific components (apart from a ribosome and a stop codon) are needed for the insertion of selenocysteine into a growing polypeptide chain?

6.21 Folding and Secreting Proteins

For a protein to function properly it must be folded correctly and it must also end up in the correct location in the cell. Here we briefly discuss these two related processes.

Levels of Protein Structure

Once formed, a polypeptide does not remain linear; instead it folds to form a more stable structure. Hydrogen bonding, between the oxygen and nitrogen atoms of two peptide bonds, generates the **secondary structure** (**Figure 6.38*a***). One common type of secondary structure is the *α-helix*. To envision an α-helix, imagine a linear polypeptide wound around a cylinder (Figure 6.38*b*). This positions peptide bonds close enough to allow hydrogen bonding. The large number of such hydrogen bonds gives the α-helix its inherent stability. In the β-sheet, the polypeptide chain folds back and forth upon itself instead of forming a helix. However, as in the α-helix, the folding in a β-sheet positions peptide bonds so that they can undergo hydrogen bonding (Figure 6.38*c*).

Many polypeptides contain regions of both α-helix and β-sheet secondary structure, the type of folding and its location in the molecule being determined by the primary structure and the available opportunities for hydrogen bonding. A typical protein is thus made up of many folded subdomains.

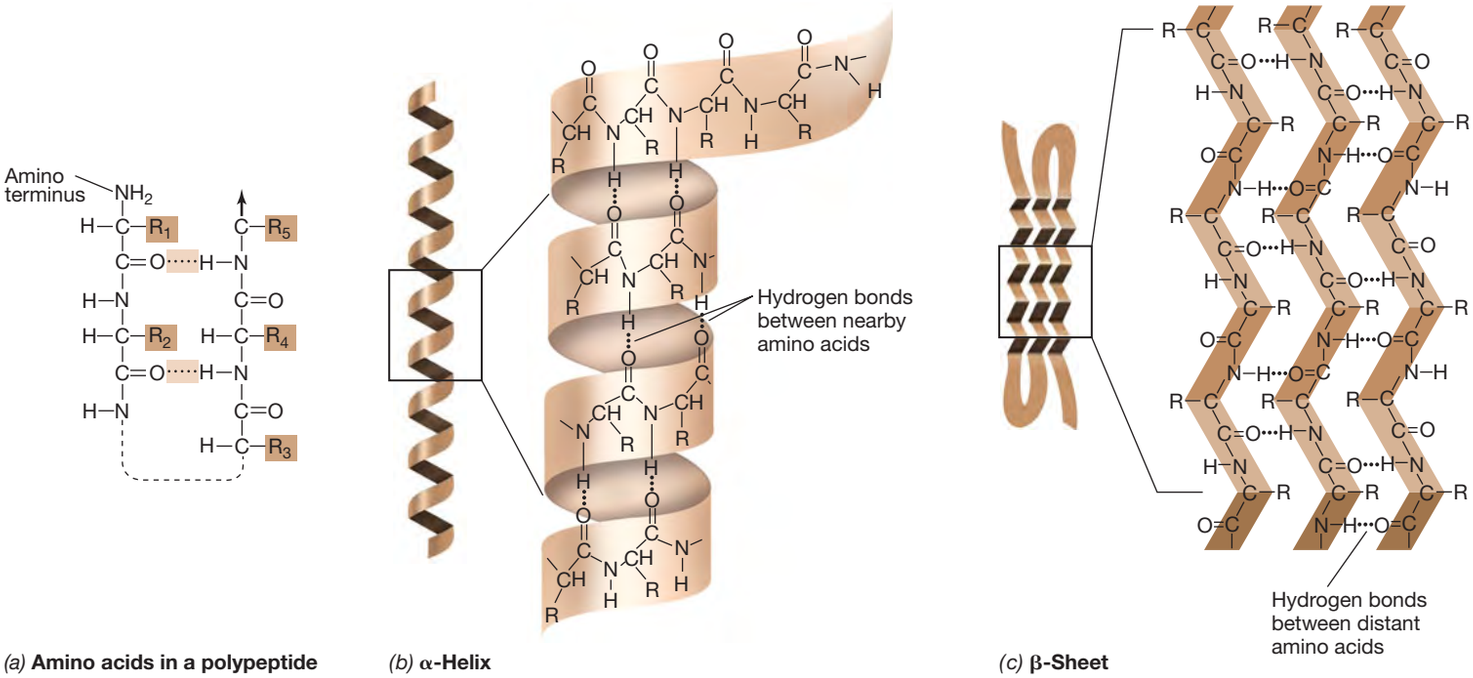

Figure 6.38 **Secondary structure of polypeptides.** *(a)* Hydrogen bonding in protein secondary structure. R represents the side chain of the amino acid. *(b)* α-Helix secondary structure. *(c)* β-Sheet secondary structure. Note that the hydrogen bonding is between atoms in the peptide bonds and does not involve the R groups.

Interactions between the R groups of the amino acids in a polypeptide generate two further levels of structure. The **tertiary structure** depends largely on hydrophobic interactions, with lesser contributions from hydrogen bonds, ionic bonds, and disulfide bonds. The tertiary folding generates the overall three-dimensional shape of each polypeptide chain (**Figure 6.39**). Many proteins consist of two or more polypeptide chains. The **quaternary structure** refers to the number and type of polypeptides that form the final protein. In proteins with quaternary structure, each polypeptide is called a *subunit* and has its own primary, secondary, and tertiary structure. Some proteins have multiple copies of a single subunit. A protein with two identical subunits, for example, is called a *homodimer*. Other proteins may contain nonidentical subunits, each present in one or more copies (a *heterodimer*, for example, has one copy each of two different polypeptides). The subunits are held together by the same forces as for tertiary structure.

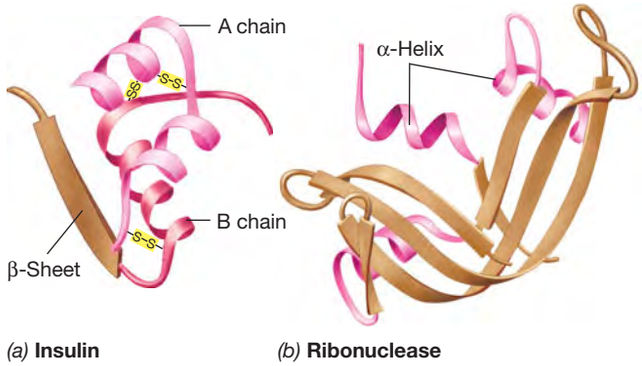

Figure 6.39 **Tertiary structure of polypeptides.** *(a)* Insulin, a protein containing two polypeptide chains; note how the B chain contains both α-helix and β-sheet secondary structure and how disulfide linkages (S–S) help in dictating folding patterns (tertiary structure). *(b)* Ribonuclease, a large protein with several regions of α-helix and β-sheet secondary structure.

Both tertiary and quaternary structures may be stabilized by disulfide bonds between two adjacent sulfhydryl groups of appropriately positioned cysteine residues. If the two cysteine residues are located in different polypeptides, the disulfide bond covalently links the two molecules. Alternatively, a single polypeptide chain can fold and bond to itself if a disulfide bond can form within the molecule.

Chaperonins Assist Protein Folding

Most polypeptides fold spontaneously into their active form while they are being synthesized. However, some do not and require assistance from other proteins called **chaperonins** (also known as **molecular chaperones**) for proper folding or for assembly into larger complexes. The chaperonins themselves do not become part of the assembly but only assist in folding. Indeed, one important function of chaperonins is to prevent improper aggregation of proteins.

There are several different kinds of chaperonins. Some help newly synthesized proteins fold correctly. Other chaperonins are very abundant in the cell, especially under growth conditions that put protein stability at risk (for example, high temperatures).

Chaperonins are widespread in all domains of life, and their sequences are highly conserved among all organisms.

Four key chaperonins in *Escherichia coli* are the proteins DnaK, DnaJ, GroEL, and GroES. DnaK and DnaJ are ATP-dependent enzymes that bind to newly formed polypeptides and keep them from folding too abruptly, a process that increases the risk of improper folding (**Figure 6.40**). Slower folding thus improves the chances of correct folding. If the DnaKJ complex is unable to fold the protein properly, it may transfer the partially folded protein to the two multi-subunit proteins GroEL and GroES. The protein first enters GroEL, a large barrel-shaped protein that uses the energy of ATP hydrolysis to fold the protein properly. GroES assists in this (Figure 6.40). It is estimated that only about 100 of the several thousand proteins of *E. coli* need help in folding from the GroEL–GroES complex, and of these approximately a dozen are essential for survival of the bacteria.

In addition to folding newly synthesized proteins, chaperonins can also refold proteins that have partially denatured in the cell. A protein may denature for many reasons, but often it is because the organism has temporarily experienced high temperatures. Chaperonins are thus one type of *heat shock protein*, and their synthesis is greatly accelerated when a cell is stressed by excessive heat (Section 8.11). The heat shock response is an attempt by the cell to refold its partially denatured proteins for reuse before proteases recognize them as improperly folded and destroy them. Refolding is not always successful, and cells contain proteases whose function is to specifically target and destroy misfolded proteins, freeing their amino acids to make new proteins.

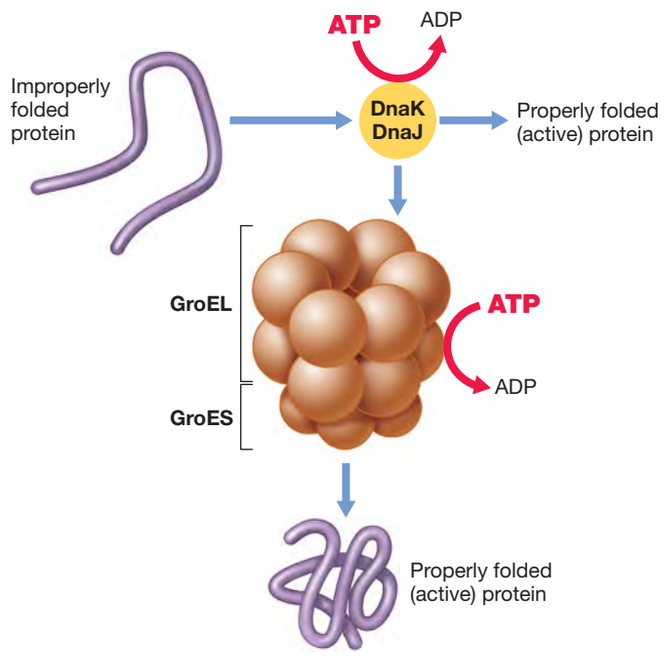

Figure 6.40 **The activity of molecular chaperones.** An improperly folded protein can be refolded by either the DnaKJ complex or by the GroEL–GroES complex. In both cases, energy for refolding comes from ATP.

Denaturation

When proteins are exposed to extremes of heat or pH or to certain chemicals that affect their folding, they may undergo **denaturation**. This results from the polypeptide chain unfolding, so destroying the higher-order (secondary, tertiary, and quaternary) structure of the protein. Depending on the severity of the denaturing conditions, the polypeptide may refold after the denaturant is removed. Typically, however, denatured proteins unfold to expose their hydrophobic regions. They then stick together to form protein aggregates that lack biological activity.

The biological properties of a protein are usually lost when it is denatured. Peptide bonds are not broken, so a denatured molecule retains its primary structure. This shows that biological activity is a function of the uniquely folded form of the protein as ultimately directed by primary structure. Denaturation of proteins is a major means of destroying microorganisms. Alcohols such as phenol and ethanol are effective disinfectants because they readily penetrate cells and irreversibly denature their proteins. Such chemical disinfectants have enormous practical value in household, hospital, and industrial applications. We discuss disinfectants, along with other agents used to destroy microorganisms, in Chapter 26.

Protein Secretion and the Signal Recognition Particle

Many proteins are located in the cytoplasmic membrane, in the periplasm of gram-negative cells, or even outside the cell proper. Such proteins must get from their site of synthesis on ribosomes into or through the cytoplasmic membrane. How is it possible for a cell to selectively transfer some proteins across a membrane while leaving most proteins in the cytoplasm?

Most proteins that must be transported into or through membranes are synthesized with an amino acid sequence of about 15–20 residues, called the **signal sequence**, at the beginning of the protein molecule. Signal sequences are quite variable, but typically they have a few positively charged residues at the beginning, a central region of hydrophobic residues, and then a more polar region. The signal sequence "signals" the cell's secretory system that this particular protein is to be exported and also helps prevent the protein from completely folding, a process that could interfere with its secretion. Because the signal sequence is the first part of the protein to be synthesized, the early steps in export may actually begin before the protein is completely synthesized (**Figure 6.41**).

Proteins to be exported are identified by their signal sequences either by the *SecA protein* or the *signal recognition particle* (*SRP*) (Figure 6.41). Generally, SecA binds proteins that are fully exported across the membrane into the periplasm whereas the SRP binds proteins that are inserted into the membrane but are not released on the other side. SRPs are found in all cells. In *Bacteria*, they contain a single protein and a small noncoding RNA molecule (4.5S RNA). Both SecA and the SRP deliver proteins to be secreted to the membrane secretion complex. In *Bacteria* this is normally the Sec system, whose channel consists of the three proteins SecYEG. The protein is exported across the cytoplasmic membrane through this channel. It may then either

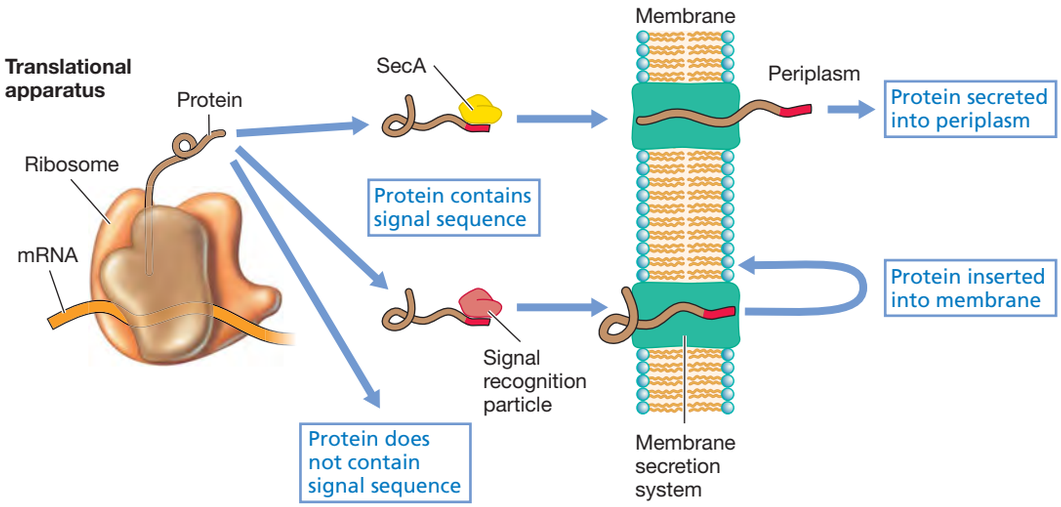

Figure 6.41 Export of proteins via the major secretory system. The signal sequence is recognized either by SecA or by the signal recognition particle, which carries the protein to the membrane secretion system. The signal recognition particle binds proteins that are inserted into the membrane whereas SecA binds proteins that are secreted across the cytoplasmic membrane.

remain in the membrane or be released into the periplasm or the environment (Figure 6.41). After crossing the membrane, the signal sequence is removed by a protease.

Secretion of Folded Proteins: The Tat System

In the Sec system for protein export, the transported proteins are threaded through the cytoplasmic membrane in an unfolded state and only fold afterward (Figure 6.41). However, there are a few proteins that must be transported outside the cell after they have already folded. Usually this is because they contain small cofactors that must be inserted into the protein as it folds into its final form. Such proteins fold in the cytoplasm and then are exported by a transport system distinct from Sec, called the *Tat protein export system*.

The acronym Tat stands for "twin arginine translocase" because the transported proteins contain a short signal sequence containing a pair of arginine residues. This signal sequence on a folded protein is recognized by the TatBC proteins, which carry the protein to TatA, the membrane transporter. The energy required for transport is supplied by the proton motive force. A wide variety of proteins are transported by the Tat system, especially proteins required for energy metabolism that function in the periplasm. This includes iron–sulfur proteins and several other redox-coupled proteins (Section 4.9). In addition, the Tat pathway transports proteins needed for outer membrane biosynthesis and a few proteins that do not contain cofactors but can only fold properly within the cytoplasm.

MiniQuiz

- Define the terms primary, secondary, and tertiary structure with respect to proteins.
- How does a polypeptide differ from a protein?
- Describe the number and kinds of polypeptides present in a homotetrameric protein.
- What is a molecular chaperone?
- Why do some proteins have a signal sequence?
- What is a signal recognition particle?

In this chapter we have covered the essentials of the key molecular processes that occur in *Bacteria*. We next consider how archaeal and eukaryotic cells carry out the same processes. There are many similarities but also some major differences, in replication, transcription, and translation, among organisms in the three domains of life.

Big Ideas

6.1

The informational content of a nucleic acid is determined by the sequence of nitrogenous bases along the polynucleotide chain. Both RNA and DNA are informational macromolecules, as are the proteins they encode. RNA can fold into various configurations to generate secondary structure. The three key processes of macromolecular synthesis are: (1) DNA replication; (2) transcription (the synthesis of RNA from a DNA template); and (3) translation (the synthesis of proteins using messenger RNA as template).

6.2

DNA is a double-stranded molecule that forms a helix. Its length is measured in terms of numbers of base pairs. The two strands in the double helix are antiparallel, but inverted repeats allow for the formation of secondary structure. The strands of a

double-helical DNA molecule can be denatured by heat and allowed to reassociate following cooling.

6.3
Very long DNA molecules can be packaged into cells because they are supercoiled. In prokaryotes this supercoiling is brought about by enzymes called topoisomerases. DNA gyrase is a key enzyme in prokaryotes and introduces negative supercoils into the DNA.

6.4
In addition to the chromosome, a number of other genetic elements exist in cells. Plasmids are DNA molecules that exist separately from the chromosome of the cell. Viruses contain a genome, either DNA or RNA, that controls their own replication. Transposable elements exist as a part of other genetic elements.

6.5
The *Escherichia coli* chromosome has been mapped using conjugation, transduction, molecular cloning, and sequencing. *E. coli* has been a useful model organism, and a considerable amount of information has been obtained from it, not only about gene structure but also about gene function and regulation.

6.6
Plasmids are small circular or linear DNA molecules that carry nonessential genes. Although a cell can contain more than one plasmid, these cannot be closely related genetically. Although they have no extracellular form, plasmids can be transferred by the process of conjugation.

6.7
The genetic information that plasmids carry is not essential for cell function under all conditions but may confer a selective growth advantage under certain conditions. Examples include antibiotic resistance, enzymes for degradation of unusual organic compounds, and special metabolic pathways. Virulence factors of many pathogenic bacteria are plasmid encoded.

6.8
Both strands of the DNA helix serve as templates for the synthesis of two new strands (semiconservative replication). The two progeny double helices each contain one parental strand and one new strand. The new strands are elongated by addition of deoxyribonucleotides to the 3′ end. DNA polymerases require a primer made of RNA.

6.9
DNA synthesis begins at a unique location called the origin of replication. The double helix is unwound by helicase and is stabilized by single-strand binding protein. Extension of the DNA occurs continuously on the leading strand but discontinuously on the lagging strand. The fragments of the lagging strand are joined together later.

6.10
Starting from a single origin, DNA synthesis proceeds in both directions around circular chromosomes. Therefore, there are two replication forks in operation simultaneously. The proteins at the replication fork form a large complex known as the replisome. Most errors in base pairing that occur during replication are corrected by the proofreading functions of DNA polymerases. Incorrect nucleotides are removed and replaced. Finally, DNA replication terminates when the replication forks meet at a special terminus region on the chromosome.

6.11
The polymerase chain reaction is a procedure for amplifying DNA *in vitro* and employs heat-stable DNA polymerases. Heat is used to denature the DNA into two single-stranded molecules, each of which is copied by the polymerase. After each cycle, the newly formed DNA is denatured and a new round of copying proceeds. After each cycle, the amount of target DNA doubles.

6.12
The three major types of RNA are messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA). Transcription of RNA from DNA is due to the enzyme RNA polymerase, which adds nucleotides onto 3′ ends of growing chains. Unlike DNA polymerase, RNA polymerase needs no primer and recognizes a specific start site on the DNA called the promoter.

6.13
In *Bacteria*, promoters are recognized by the sigma subunit of RNA polymerase. Regions of DNA recognized by a particular DNA-binding protein have very similar sequences. Alternative sigma factors allow joint regulation of large families of genes in response to growth conditions.

6.14
RNA polymerase stops transcription at specific sites called transcription terminators. Although encoded by DNA, these terminators function at the level of RNA. Some are intrinsic terminators and require no accessory proteins beyond RNA polymerase itself. In *Bacteria* these sequences are usually stem–loops followed by a run of uridines. Other terminators require proteins such as Rho.

6.15
The unit of transcription in prokaryotes often contains more than a single gene. Several genes are then transcribed into a single mRNA molecule that contains information for more than one polypeptide. A cluster of genes that are transcribed together from a single promoter constitute an operon. In all organisms, genes encoding rRNA are cotranscribed but then processed to form the final rRNA species.

6.16
Polypeptide chains contain 22 different genetically encoded amino acids that are linked via peptide bonds. Mirror-image (enantiomeric) forms of amino acids exist, but only the L-form is found in proteins. The primary structure of a protein is its amino acid sequence, but the folding (higher-order structure) of the polypeptide determines how the protein functions in the cell.

6.17
The genetic code is expressed as RNA, and a single amino acid may be encoded by several different but related codons. In addition to the nonsense codons, there is also a specific start codon that signals where the translation process should begin.

6.18
One or more tRNAs exist for each amino acid incorporated into proteins by the ribosome. Enzymes called aminoacyl-tRNA synthetases attach amino acids to their cognate tRNAs. Once the correct amino acid is attached to its tRNA, further specificity resides primarily in the codon–anticodon interaction.

6.19
The ribosome plays a key role in the translation process, bringing together mRNA and aminoacyl-tRNAs. There are three sites on the ribosome: the acceptor site, where the charged tRNA first combines; the peptide site, where the growing polypeptide chain is held; and an exit site. During each step of amino acid addition, the ribosome advances three nucleotides (one codon) along the mRNA, and the tRNA that is in the acceptor site moves to the peptide site. Protein synthesis terminates when a nonsense codon, which does not encode an amino acid, is reached.

6.20
Many nonstandard amino acids are found in proteins as a result of posttranslational modification. In contrast, the two rare amino acids selenocysteine and pyrrolysine are inserted into growing polypeptide chains during protein synthesis. They are both encoded by special stop codons that have a nearby recognition sequence that is specific for insertion of selenocysteine or pyrrolysine.

6.21
Proteins must be properly folded in order to function correctly. Folding may occur spontaneously but may also involve other proteins called molecular chaperones. Many proteins also need to be transported into or through membranes. Such proteins are synthesized with a signal sequence that is recognized by the cellular export apparatus and is removed either during or after export.

Review of Key Terms

Amino acid one of the 22 different monomers that make up proteins; chemically, contains a carboxylic acid, an amino group and a characteristic side chain all attached to the α-carbon

Aminoacyl-tRNA synthetase an enzyme that catalyzes attachment of an amino acid to its cognate tRNA

Anticodon a sequence of three bases in a tRNA molecule that base-pairs with a codon during protein synthesis

Antiparallel in reference to double-stranded DNA, the two strands run in opposite directions (one runs 5′ → 3′ and the complementary strand 3′ → 5′)

Bacteriocin a toxic protein secreted by bacteria that kills other, related bacteria

Chaperonin or molecular chaperone a protein that helps other proteins fold or refold from a partly denatured state

Chromosome a genetic element, usually circular in prokaryotes, carrying genes essential to cellular function

Codon a sequence of three bases in mRNA that encodes an amino acid

Codon bias nonrandom usage of multiple codons encoding the same amino acid

Complementary nucleic acid sequences that can base-pair with each other

Denaturation loss of the correct folding of a protein, leading (usually) to protein aggregation and loss of biological activity

DNA (deoxyribonucleic acid) a polymer of deoxyribonucleotides linked by phosphodiester bonds that carries genetic information

DNA gyrase an enzyme found in most prokaryotes that introduces negative supercoils in DNA

DNA ligase an enzyme that seals nicks in the backbone of DNA

DNA polymerase an enzyme that synthesizes a new strand of DNA in the 5′ → 3′ direction using an antiparallel DNA strand as a template

Enantiomer a form of a molecule that is the mirror image of another form of the same molecule

Enzyme a protein or an RNA that catalyzes a specific chemical reaction in a cell

Gene a segment of DNA specifying a protein (via mRNA), a tRNA, an rRNA, or any other noncoding RNA

Genetic code the correspondence between nucleic acid sequence and amino acid sequence of proteins

Genetic element a structure that carries genetic information, such as a chromosome, a plasmid, or a virus genome

Genome the total complement of genes contained in a cell or virus

Informational macromolecule any large polymeric molecule that carries genetic information, including DNA, RNA, and protein

Lagging strand the new strand of DNA that is synthesized in short pieces and then joined together later

Leading strand the new strand of DNA that is synthesized continuously during DNA replication

Messenger RNA (mRNA) an RNA molecule that contains the genetic information to encode one or more polypeptides

Nonsense codon another name for a stop codon

Nucleic acid DNA or RNA

Nucleoside a nitrogenous base (adenine, guanine, cytosine, thymine, or uracil) plus a sugar (either ribose or deoxyribose) but lacking phosphate

Nucleotide a monomer of a nucleic acid containing a nitrogenous base (adenine, guanine, cytosine, thymine, or uracil), one or more molecules of phosphate, and a sugar, either ribose (in RNA) or deoxyribose (in DNA)

Open reading frame (ORF) a sequence of DNA or RNA that could be translated to give a polypeptide

Operon a cluster of genes that are cotranscribed as a single messenger RNA

Peptide bond a type of covalent bond linking amino acids in a polypeptide

Phosphodiester bond a type of covalent bond linking nucleotides together in a polynucleotide

Plasmid an extrachromosomal genetic element that has no extracellular form

Polymerase chain reaction (PCR) artificial amplification of a DNA sequence by repeated cycles of strand separation and replication

Polynucleotide a polymer of nucleotides bonded to one another by covalent bonds called phosphodiester bonds

Polypeptide a polymer of amino acids bonded to one another by peptide bonds

Primary structure the precise sequence of monomers in a macromolecule such as a polypeptide or a nucleic acid
Primase the enzyme that synthesizes the RNA primer used in DNA replication
Primer an oligonucleotide to which DNA polymerase attaches the first deoxyribonucleotide during DNA synthesis
Promoter a site on DNA to which RNA polymerase binds to commence transcription
Protein a polypeptide or group of polypeptides forming a molecule of specific biological function
Purine one of the nitrogenous bases of nucleic acids that contain two fused rings; adenine and guanine
Pyrimidine one of the nitrogenous bases of nucleic acids that contain a single ring; cytosine, thymine, and uracil
Quaternary structure in proteins, the number and types of individual polypeptides in the final protein molecule
Replication synthesis of DNA using DNA as a template
Replication fork the site on the chromosome where DNA replication occurs and where the enzymes replicating the DNA are bound to untwisted, single-stranded DNA
Ribosomal RNA (rRNA) types of RNA found in the ribosome; some participate actively in protein synthesis
Ribosome a cytoplasmic particle composed of ribosomal RNA and protein, whose function is to synthesize proteins
RNA (ribonucleic acid) a polymer of ribonucleotides linked by phosphodiester bonds that plays many roles in cells, in particular, during protein synthesis
RNA polymerase an enzyme that synthesizes RNA in the 5′ → 3′ direction using a complementary and antiparallel DNA strand as a template
Secondary structure the initial pattern of folding of a polypeptide or a polynucleotide, usually dictated by opportunities for hydrogen bonding
Semiconservative replication DNA synthesis yielding two new double helices, each consisting of one parental and one progeny strand
Signal sequence a special N-terminal sequence of approximately 20 amino acids that signals that a protein should be exported across the cytoplasmic membrane
Start codon a special codon, usually AUG, that signals the start of a protein
Stop codon a codon that signals the end of a protein
Termination stopping the elongation of an RNA molecule at a specific site
Tertiary structure the final folded structure of a polypeptide that has previously attained secondary structure
Transcription the synthesis of RNA using a DNA template
Transfer RNA (tRNA) a small RNA molecule used in translation that possesses an anticodon at one end and has the corresponding amino acid attached to the other end
Translation the synthesis of protein using the genetic information in RNA as a template
Wobble a less rigid form of base pairing allowed only in codon–anticodon pairing

Review Questions

1. Describe the central dogma of molecular biology (Section 6.1).
2. Genes were discovered before their chemical nature was known. First, define a gene without mentioning its chemical nature. Then name the chemicals that compose a gene (Section 6.1).
3. Inverted repeats can give rise to stem–loops. Show this by giving the sequence of a double-stranded DNA containing an inverted repeat and show how the transcript from this region can form a stem–loop (Section 6.2).
4. Is the sequence 5′-GCACGGCACG-3′ an inverted repeat? Explain your answer (Section 6.2).
5. DNA molecules that are AT-rich separate into two strands more easily when the temperature rises than do DNA molecules that are GC-rich. Explain this based on the properties of AT and GC base pairing (Section 6.2).
6. Describe how DNA, which is many times the length of a cell when linearized, fits into the cell (Section 6.3).
7. List the major genetic elements known in microorganisms (Section 6.4).
8. What is the size of the *Escherichia coli* chromosome? About how many proteins can it encode? How much noncoding DNA is present in the *E. coli* chromosome (Section 6.5)?
9. How do plasmids replicate and how does this differ from chromosomal replication (Section 6.6)?
10. What are R plasmids and why are they of medical concern (Section 6.7)?
11. A structure commonly seen in circular DNA during replication is the theta structure. Draw a diagram of the replication process and show how a theta structure could arise (Sections 6.9 and 6.10).
12. Why are errors in DNA replication so rare? What enzymatic activity, in addition to polymerization, is associated with DNA polymerase III and how does it reduce errors (Section 6.10)?
13. Describe the basic principles of gene amplification using the polymerase chain reaction (PCR). How have thermophilic and hyperthermophilic prokaryotes simplified the use of PCR (Section 6.11)?
14. Do genes for tRNAs have promoters? Do they have start codons? Explain (Sections 6.12 and 6.15).
15. The start and stop sites for mRNA synthesis (on the DNA) are different from the start and stop sites for protein synthesis (on the mRNA). Explain (Sections 6.15 and 6.19).
16. Why are amino acids so named? Write a general structure for an amino acid. What is the importance of the R group to final protein structure? Why does the amino acid cysteine have special significance for protein structure (Section 6.16)?
17. What is "wobble" and what makes it necessary in protein synthesis (Sections 6.17 and 6.18)?
18. What are aminoacyl-tRNA synthetases and what types of reactions do they carry out? How does a synthetase recognize its correct substrates (Section 6.18)?
19. The enzyme activity that forms peptide bonds on the ribosome is called peptidyl transferase. Which molecule catalyzes this reaction (Section 6.19)?

20. Define the types of protein structure: primary, secondary, tertiary, and quaternary. Which of these structures are altered by denaturation (Section 6.21)?
21. Sometimes misfolded proteins can be correctly refolded, but sometimes they cannot and are destroyed. What kinds of proteins are involved in refolding misfolded proteins? What kinds of enzymes are involved in destroying misfolded proteins (Section 6.21)?
22. How does a cell know which of its proteins are designed to function outside of the cell (Section 6.21)?

Application Questions

1. The genome of the bacterium *Neisseria gonorrhoeae* consists of one double-stranded DNA molecule that contains 2220 kilobase pairs. Calculate the length of this DNA molecule in centimeters. If 85% of this DNA molecule is made up of the open reading frames of genes encoding proteins, and the average protein is 300 amino acids long, how many protein-encoding genes does *Neisseria* have? What kind of information do you think might be present in the other 15% of the DNA?
2. Compare and contrast the activity of DNA and RNA polymerases. What is the function of each? What are the substrates of each? What is the main difference in the behavior of the two polymerases?
3. What would be the result (in terms of protein synthesis) if RNA polymerase initiated transcription one base upstream of its normal starting point? Why? What would be the result (in terms of protein synthesis) if translation began one base downstream of its normal starting point? Why?
4. In Chapter 10 we will learn about mutations, inheritable changes in the sequence of nucleotides in the genome. By inspecting Table 6.5, discuss how the genetic code has evolved to help minimize the impact of mutations.

Need more practice? Test your understanding with Quantitative Questions; access additional study tools including tutorials, animations, and videos; and then test your knowledge with chapter quizzes and practice tests at **www.microbiologyplace.com**.