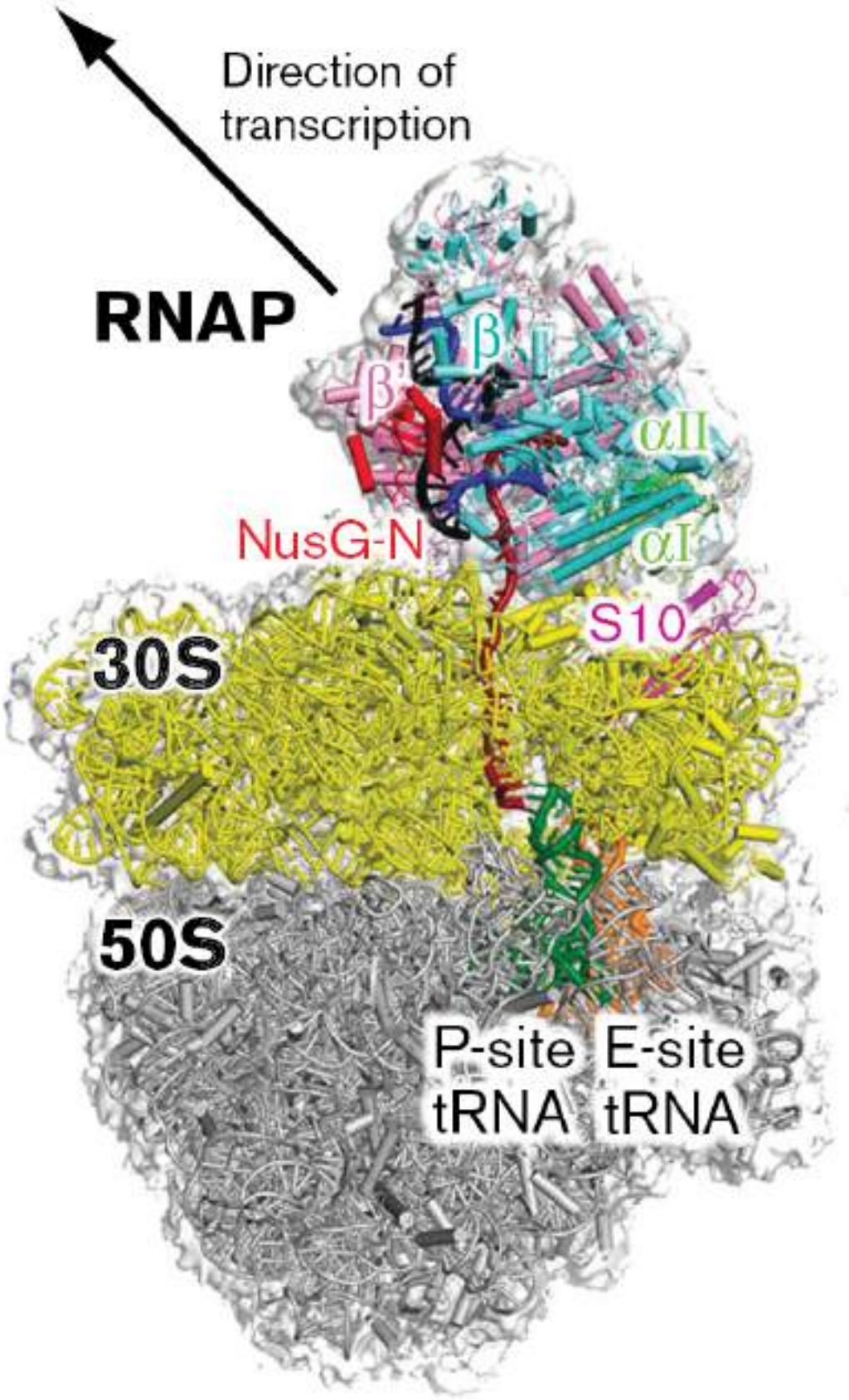
CHAPTER 8

Transcription, Translation, and Protein Processing


Direction of transcription
RNAP
β'
β
αII
αI
NusG-N
S10
30S
50S
P-site tRNA
E-site tRNA

Coupling the machines that synthesize messenger RNA and proteins. Transcription of the DNA (blue and black strands) by RNA polymerase (RNAP) generates messenger RNA (mRNA; red strand), which is translated into protein by the ribosome, composed of 30S and 50S subunits. This cryo-electron microscopy image shows the physical connection between the RNA polymerase and the ribosome that can allow coupled transcription and translation in *Escherichia coli* . As soon as the mRNA is synthesized, it is fed into the ribosome, which uses transfer RNAs (tRNAs; green and orange strands) to decode the mRNA and incorporate the correct amino acid into the growing chain. The NusG protein (shown here with just its N-terminal domain, NusG-N) helps to bridge the two enzymes and stabilize this massive complex.

WANG ET AL. 2020. *SCIENCE* . **369** :1359–1365

Chapter Sections		Learning Objectives
8.1 From DNA to RNA to Protein	»	Explain how genes are expressed through synthesis of RNA and protein and how operons can coordinate expression of adjacent genes in bacteria.

Chapter Sections		Learning Objectives
8.2 Transcription of DNA to RNA	»	Describe the molecular events that occur during the initiation, elongation, and termination of transcription in bacteria, archaea, and eukaryotes.
	»	List the classes of RNA produced during transcription, and explain how transcription can be affected by antibiotics.
8.3 Translation of RNA to Protein	»	Explain how the ribosome reads the genetic code to synthesize proteins from mRNA messages and how antibiotics can interfere with translation.
Special Topic 8 Rewriting the Genetic Code to Arrest Viral Infection		
8.4 Protein Modification, Folding, and Degradation	»	Discuss the processes of protein modification, folding, and

Chapter Sections		Learning Objectives
		degradation that occur after translation.
8.5 Secretion: Protein Traffic Control	»	Explain how proteins can be exported from the cytoplasm into the cell envelope and beyond.
eResearch Activity 8 How Do Cells of the Same Species Recognize Each Other for DNA Exchange?		

The cell accesses the vast store of data in its genome using tiny molecular machines. RNA polymerase transcribes stretches of the DNA template into temporary copies made of RNA. The ribosome translates (decodes) the RNA messages to synthesize proteins. Once made, each protein must fold properly and travel to its correct cellular or extracellular location. Proteins that have outlived their usefulness must be destroyed and their amino acids recycled.

In Chapter 8 we explore the way microbes, primarily bacteria, interpret the information held within a nucleotide sequence of DNA and convert that information into a string of amino acids—a protein. From there we look at what the cell does with those proteins once they are made. For instance, specific proteins are moved into the periplasm, while other proteins must be inserted into membranes. We also show how damaged proteins are selectively degraded. What emerges is a picture of remarkable biomolecular integration, controlled to maintain balanced growth and ensure survival.

8.1 From DNA to RNA to Protein

To survive and reproduce, every cell needs to access information encoded within DNA. Genomes of cellular microbes are made of very long stretches of DNA, usually millions of base pairs in length. These genomes can contain thousands of genes, whose function is to encode the diverse array of RNAs and proteins of the cell. When the RNA and protein products of a gene are made, we say that the gene is being expressed . Expression involves two universal processes of the cell, transcription and translation .

RNA is synthesized from a DNA template via transcription. For all three domains of life, double-stranded DNA is transcribed into single-stranded RNA. To transcribe is to copy; in this case, the copy is a structurally similar yet distinct polymer (see Fig. 7.3). For every base of DNA in a gene, there is a corresponding base of RNA, with the uracil (U) of RNA replacing the thymine (T) of DNA. Transcription will be discussed in detail in Section 8.2 .

The RNA transcripts of most genes act as coded messages (messenger RNA, or mRNA) that are subsequently translated into protein. To translate is to convert a message from one language into another; in this case, a language with an alphabet of 4 nucleotides is converted into a language with an alphabet of 20 common amino acids. Translation of the mRNA message into protein involves a genetic code, which will be discussed in Section 8.3 .

Defining a Gene

Before we describe the mechanics of transcription and translation, it will help to illustrate the alignments between the DNA sequence of a structural gene (a gene encoding a protein) and the mRNA transcript containing translation signals and the protein-coding sequences. **Figure 8.1** shows the sense ("coding"; nontemplate) and template DNA strands of a two-gene operon. The sequence of the sense

strand matches that of the mRNA transcript, but with thymines substituting for uracils. The template strand is the strand actually "read" by the transcription enzyme, RNA polymerase .

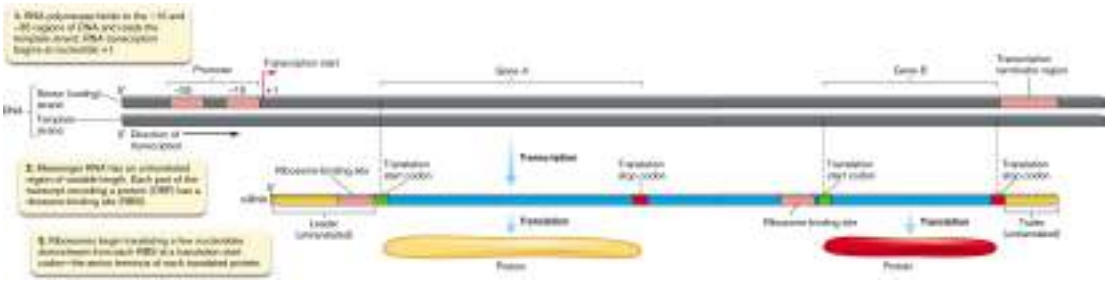


FIGURE 8.1 ■ Alignment of structural genes in a bacterial operon, the mRNA transcript, and protein products. In this figure, the term "gene" refers to the region of DNA that encodes a product. In this example, both genes encode protein. ORF = open reading frame.

In bacteria, the RNA products of transcription can be monocistronic or polycistronic, meaning that they encode the product of a single gene or multiple genes, respectively. **Figure 8.1** illustrates a polycistronic mRNA encoding two protein-coding genes. Transcription of these genes initiates at the promoter of the first gene and terminates at the end of the last gene. The combination of the regulatory regions (promoter, terminator) and all the genes that are cotranscribed in the polycistronic message is referred to as an operon . As we will see in Chapter 10 , placing multiple genes under the control of the same promoter is an efficient way to coordinate their regulation.

Transcription initiates adjacent to the promoter, and in **Figure 8.1** the +1 marks the DNA base where the synthesis of the mRNA polymer begins. A protein-coding region of the transcript is called an open reading frame (ORF) , defined as the region between a translation start codon and a translation stop codon (codon sequences will be discussed in Section 9.3). Gene *A* and gene *B* have their own translation start and stop codons, as well as ribosome-binding sites, whose function will be described later. Note

that not all of the mRNA transcript is translated into protein; an untranslated "leader" sequence precedes the gene *A* protein-coding region, and an untranslated "trailer" lies downstream of gene *B* . The leader and trailer sequences, also referred to as the 5′ and 3′ untranslated regions (UTRs), respectively, help regulate gene expression, as will be discussed in Sections 9.3 and 10.3 .

Thought Question

8.1 **Figure 8.1** illustrates an operon and its relationship to transcripts and protein products. Imagine that a mutation generates a stop codon about midway through the DNA sequence that encodes gene *A* . What would happen to the production of the gene *A* and gene *B* proteins?

To Summarize

- **Transcription and translation** are processes that synthesize RNA from a DNA template and protein from an RNA template, respectively.
- **mRNA transcripts** contain open reading frames that are translated into protein. They also contain untranslated regions upstream and downstream of the open reading frames.
- **Operons** consist of cotranscribed genes that share promoters and transcription terminators.

Glossary

expression
: Synthesis of the products encoded by a gene. This includes RNA and, for mRNAs that are translated, protein.

transcription
: The synthesis of RNA complementary to a DNA template.

translation
: The ribosomal synthesis of proteins based on triplet codons present in mRNA.

messenger RNA (mRNA)
: An RNA molecule that encodes a protein.

RNA polymerase
: Also called *DNA-dependent RNA polymerase* . An enzyme that produces an RNA complementary to a template DNA strand.

promoter
: A noncoding DNA regulatory region immediately upstream of a structural gene that is needed for transcription initiation.

operon
: A collection of genes that are in tandem on a chromosome and are transcribed into a single RNA.

open reading frame (ORF)
: A DNA sequence predicted to encode a protein.

Fig. 7.3

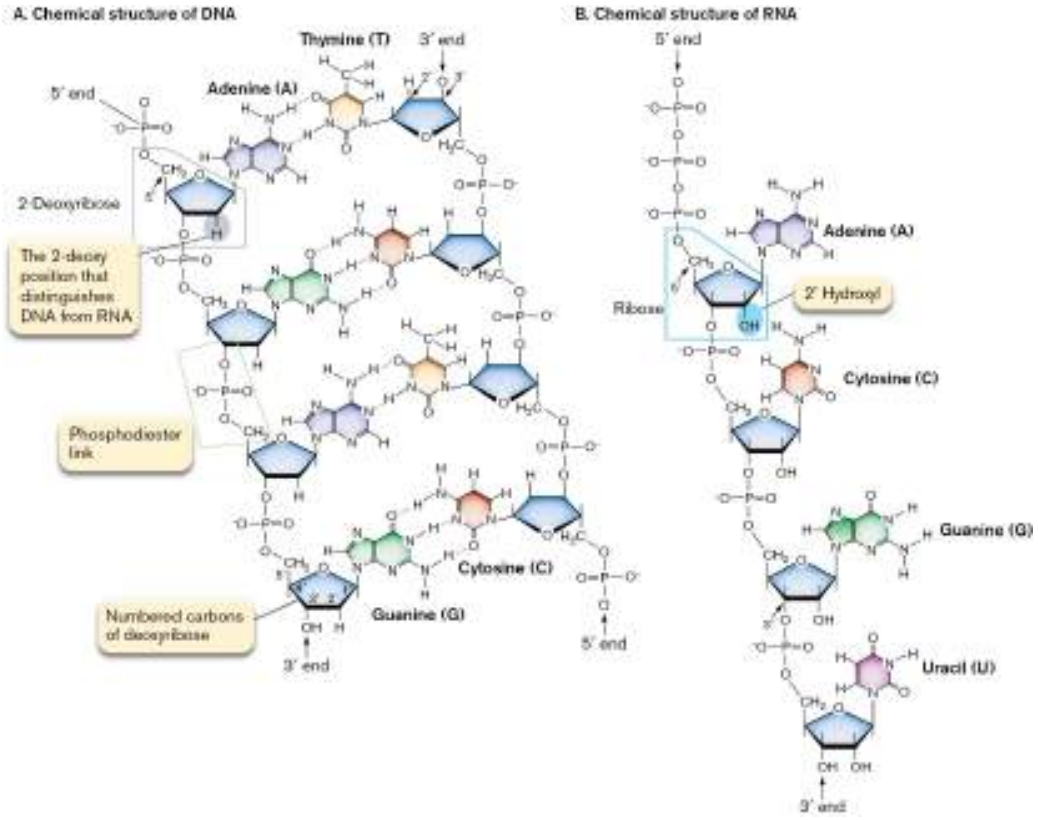


FIGURE 7.3 ■ Structures of DNA and RNA. A. In the cell, DNA bases are added only to a preexisting 3′ OH of a nucleoside monophosphate, so the 5′ ends in this figure are drawn as nucleoside monophosphates. (Dotted lines indicate hydrogen bonds between bases.) **B.** Cellular RNA molecules, however, begin with a 5′ triphosphate.

8.2 Transcription of DNA to RNA

RNA Polymerase Transcribes DNA to RNA

An enzyme complex called RNA polymerase , also known as DNA-dependent RNA polymerase , carries out transcription, making RNA copies (called transcripts) of a DNA template. The DNA template strand specifies the base sequence of the new complementary strand of RNA.

RNA polymerase in bacteria consists of a core polymerase and a sigma (σ) factor. Core polymerase contains the proteins required to elongate an RNA chain. Sigma factor is a protein needed only for initiation of RNA synthesis, not for its elongation. Together, core polymerase plus sigma factor are called the holoenzyme.

A bacterial core RNA polymerase is a complex of two alpha (α) subunits, one beta (β) subunit, and one beta-prime (β′) subunit (**Fig. 8.2**). A fifth subunit, omega (ω), is not required for transcription but plays a role in assembly and maintenance of the core RNA polymerase as well as promoter selection. The beta-prime subunit houses the Mg^{2+} -containing catalytic site for RNA synthesis, as well as sites for the rNTP (ribonucleoside triphosphate, or ribonucleotide) substrates, the DNA substrates, and the RNA products. The 3D structure of RNA polymerase shows that DNA fits into a cleft formed by the beta and beta-prime subunits (**Fig. 8.2**). The alpha subunits assemble the other two subunits (beta and beta-prime) into a functional complex. The alpha subunits also communicate through physical "touch" with various regulatory proteins that can bind DNA (for example, see **Figure 10.10**). These protein-protein interactions inform RNA polymerase what to do after the enzyme binds DNA.

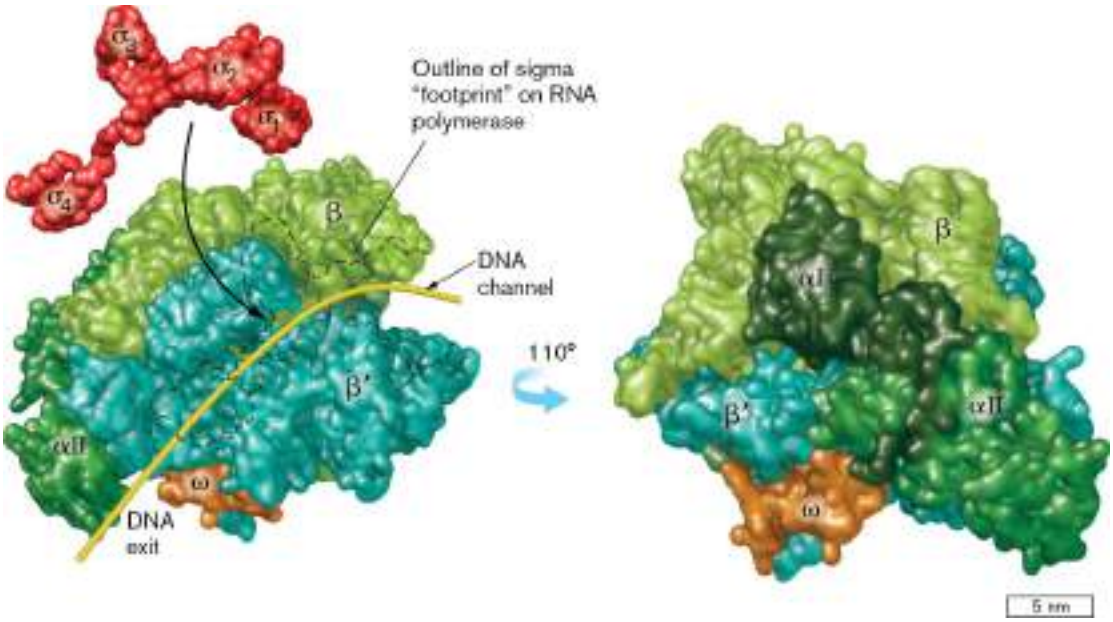


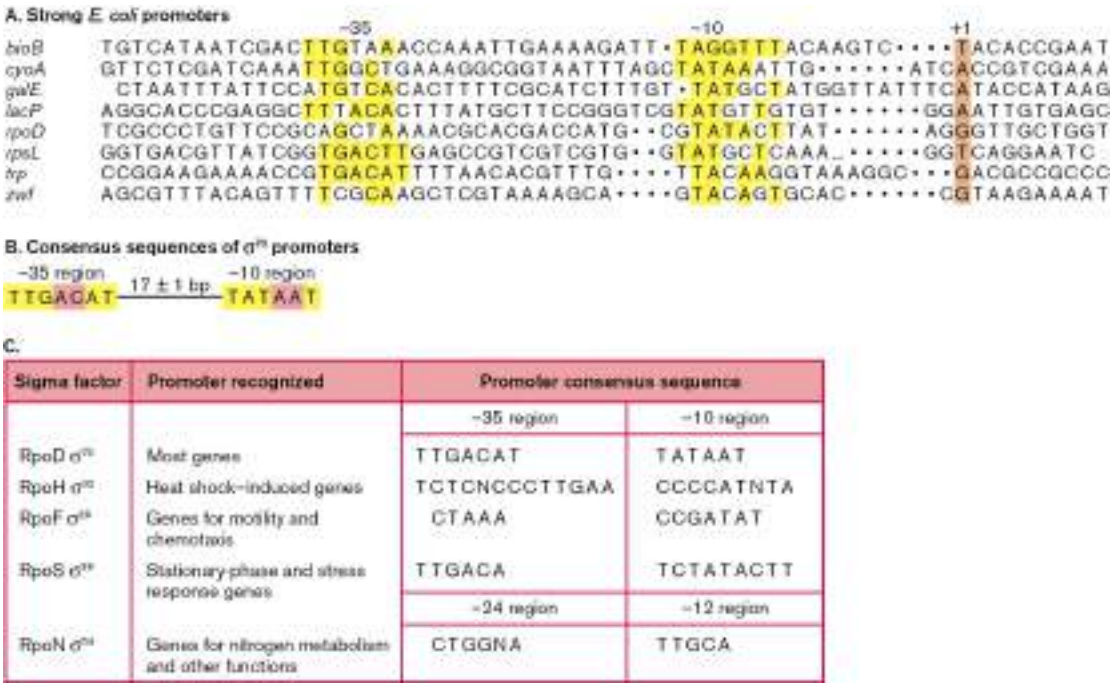
FIGURE 8.2 ■ Subunit structure of RNA polymerase. Two views of RNA polymerase. On the left, the channel for the DNA template is shown by the yellow line. Subunits (α I , α II , β, β′, and ω) are color-coded dark green, medium green, light green, cyan, and gold, respectively. Sigma factor (red), which recognizes promoters on DNA, is shown separate from core polymerase in the left-hand panel. Different functional areas of sigma factor are labeled sigma 1 through sigma 4 ($\sigma_1 - \sigma_4$). Sigma factor interacts with the alpha (α), beta (β), and beta-prime (β′) subunits. The molecule on the left is rotated 110° to give the image on the right. (PDB code: 1L9Z)

Source : Robert D. Finn et al. 2000. *EMBO J* . **19** :6833–6844.

Sigma factors. Unregulated transcription of the genome at random starting locations would be extremely wasteful and problematic to the cell. Consequently, bacteria mark the beginning of the gene or gene cluster with a specific sequence, called the promoter, and use a special protein, the sigma factor, to guide RNA polymerase to the promoter. A sigma factor first binds to core RNA polymerase through the beta and beta-prime subunits (see the dotted black outline in **Fig. 8.2**), and this holoenzyme binds to the promoter recognized by the bound sigma factor. A single bacterial species can make several different sigma factors (see **Fig. 8.3C** and Section 10.2 for some examples). Each sigma factor helps core RNA polymerase find the start of a different subset of genes. However, a single core polymerase complex can bind only one sigma factor at a time.

Promoters. Every cell has a "housekeeping" sigma factor that keeps essential genes and pathways operating. In the case of *Escherichia coli* and other rod-

shaped, Gram-negative bacteria, that factor is sigma-70, or σ^{70}, so named because it is a 70-kilodalton (kDa) protein (its gene designation is *rpoD*). Genes recognized by sigma-70 all contain similar promoter sequences that consist of two parts. The DNA base corresponding to the start of the RNA transcript is called nucleotide +1 (+1 nt). Relative to this landmark, promoter sequences are usually centered at −10 and −35 nt before the start of transcription (see **Fig. 8.3**). Other sigma factors typically recognize different consensus sequences at one or both of these positions (or at nearby locations in some cases).



C.

Sigma factor	Promoter recognized	Promoter consensus sequence	
		−35 region	−10 region
RpoD σ^{70}	Most genes	TTGACAT	TATAAT
RpoH σ^{32}	Heat shock–induced genes	TCTCNCCCTTGAA	CCCCATNTA
RpoF σ^{28}	Genes for motility and chemotaxis	CTAAA	CCGATAT
RpoS σ^{38}	Stationary-phase and stress response genes	TTGACA	TCTATACTT
		−24 region	−12 region
RpoN σ^{54}	Genes for nitrogen metabolism and other functions	CTGGNA	TTGCA

FIGURE 8.3 ■ −10 and −35 sequences of *E. coli* promoters. A. Alignment of the upstream region of a subsample of genes whose promoters are recognized by sigma-70 (σ^{70}). Dots do not represent bases; they help to align the sequences by accounting for short insertions and deletions. Yellow indicates conserved nucleotides; brown denotes transcript start sites (+1). **B.** The alignment in (A) generates a consensus sequence of sigma-70-dependent promoters (red-screened letters indicate nucleotide positions where different promoters show a high degree of variability). **C.** Some *E. coli* promoter sequences recognized by different sigma factors. "N" indicates that any of the four standard nucleotides can occupy the position.

The DNA promoter sequence recognized by a given sigma factor can be determined by comparison of known promoter sequences of different genes whose expression requires the same sigma. Similarities among these different promoter DNA sequences define a consensus sequence likely recognized by the sigma factor (**Fig. 8.3A** and **B**). A consensus sequence consists of the most likely base (or

bases) at each position of the predicted promoter. Although promoters are double-stranded DNA (dsDNA) sequences, convention is to present the promoter as the single-stranded DNA (ssDNA) sequence of the sense (nontemplate) strand, which has the same sequence as the RNA product.

Some positions in a consensus sequence are highly conserved, meaning that the same base is found in that position in every promoter. Other, less conserved positions can be occupied by different bases. Few promoters actually have the most common base at every position. Even highly efficient promoters usually differ from the consensus at one or two positions.

Sigma factor recognition of promoters. How do sigma factors, or any other DNA-binding proteins for that matter, recognize specific DNA sequences when the DNA is a double helix? The phosphodiester backbone is quite uniform, and the interior of paired bases appears inaccessible. However, proteins can recognize side groups of the bases that protrude from the major and minor grooves of DNA (see Fig. 7.4). Portions of sigma-70 from *E. coli* wrap around DNA, allowing certain parts of the protein to fit into DNA grooves.

Figure 8.4A shows the holoenzyme RNA polymerase complex positioned at a promoter and the points (−10 and −35 nt) where sigma factor contacts DNA. Sigma factors generally contain four highly conserved amino acid sequences, called regions (see **Fig. 8.2** , σ_1 –σ_4). Part of region 2 of the sigma-70 family recognizes −10 sequences, whereas region 4 recognizes the −35 sites. Part of region 1 helps separate the DNA strands to make a transcription "bubble" in preparation for RNA synthesis (**Fig. 8.4B**).

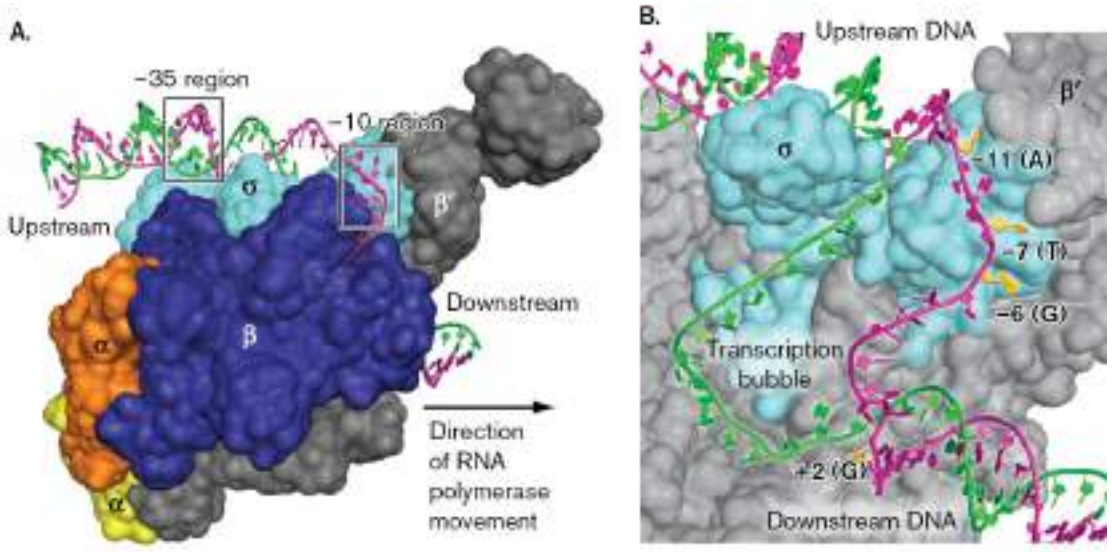


FIGURE 8.4 ■ RNA polymerase holoenzyme bound to a promoter. A. The sigma factor subunit (light blue) of the RNA polymerase holoenzyme contacts the −10 and −35 regions of the promoter. Nontemplate (coding) strand is color-coded magenta; template strand, green. **B.** Blowup of (A), with the beta subunit removed to view the transcription bubble. Some bases in the nontemplate strand are flipped outward (yellow) to interact with sigma factor or the beta subunit after the transcription bubble is formed. For several bases of the coding strand, base identity (in parentheses) and position relative to the first base transcribed (at position +1) are indicated.

Sigma factor control of complex physiological responses. A single cell will almost certainly experience numerous environmental changes during its life. Each time the environment changes, the cell is challenged to readjust its physiology, sometimes very quickly to avoid death. Large readjustments can be made by sigma factors because they can simultaneously activate a large set of genes. Gene sets controlled by different sigma factors include those dealing with nitrogen metabolism, flagellar synthesis, heat stress, starvation, sporulation, and many other physiological responses.

In response to a change in a given environment, microbes will increase the abundance of the appropriate sigma factor (described in Chapter 10). The resulting high concentration of the specialty sigma factor will dislodge and replace other sigma factors, including sigma-70, from core polymerase. In this way, the cell redirects RNA polymerase to the promoters of genes best suited for growth or survival in the new environment. For instance, if the temperature suddenly rises, the cell produces sigma-32 (**Fig. 8.3C**), which displaces sigma-70 and redirects RNA polymerase to transcribe genes involved in the heat-shock response.

Thought Questions

8.2 If each sigma factor recognizes a different promoter, how does the cell manage to transcribe genes that respond to multiple stresses, each involving a different sigma factor?

8.3 Imagine two different sigma factors with different promoter recognition sequences. What would happen to the overall gene expression profile in the cell if one sigma factor were artificially overexpressed? Could there be a detrimental effect on growth?

8.4 Why might some genes contain multiple promoters, each one specific for a different sigma factor?

The Three Stages of Transcription

Like DNA replication, transcription of DNA to RNA occurs in three stages:

1. **Initiation** , in which RNA polymerase binds to the promoter, melts open the DNA helix, and catalyzes placement of the first RNA nucleotide
2. **Elongation** , the sequential addition of ribonucleotides to the 3′ OH end of a growing RNA chain
3. **Termination** , whereby sequences trigger release of the polymerase and the completed RNA molecule

The newly released RNA polymerase can then engage another sigma factor to seek a new promoter.

Transcription initiation. RNA polymerase constantly scans DNA for promoter sequences (**Fig. 8.5** , step 1). At the promoter, RNA polymerase holoenzyme forms a loosely bound, closed complex with DNA, which remains annealed and double-stranded—that is, unmelted (step 2). To successfully transcribe a gene, this closed complex must open to form the transcription bubble. The complex opens through the unwinding of one helical turn, which causes DNA to become unpaired in this area (step 3; see also **Fig. 8.4B**). Recall from Section 7.2 that the chromosome is poised for such unwinding as a result of the negative supercoils generated during DNA replication. After promoter unwinding, RNA polymerase in the open complex becomes tightly bound to DNA.

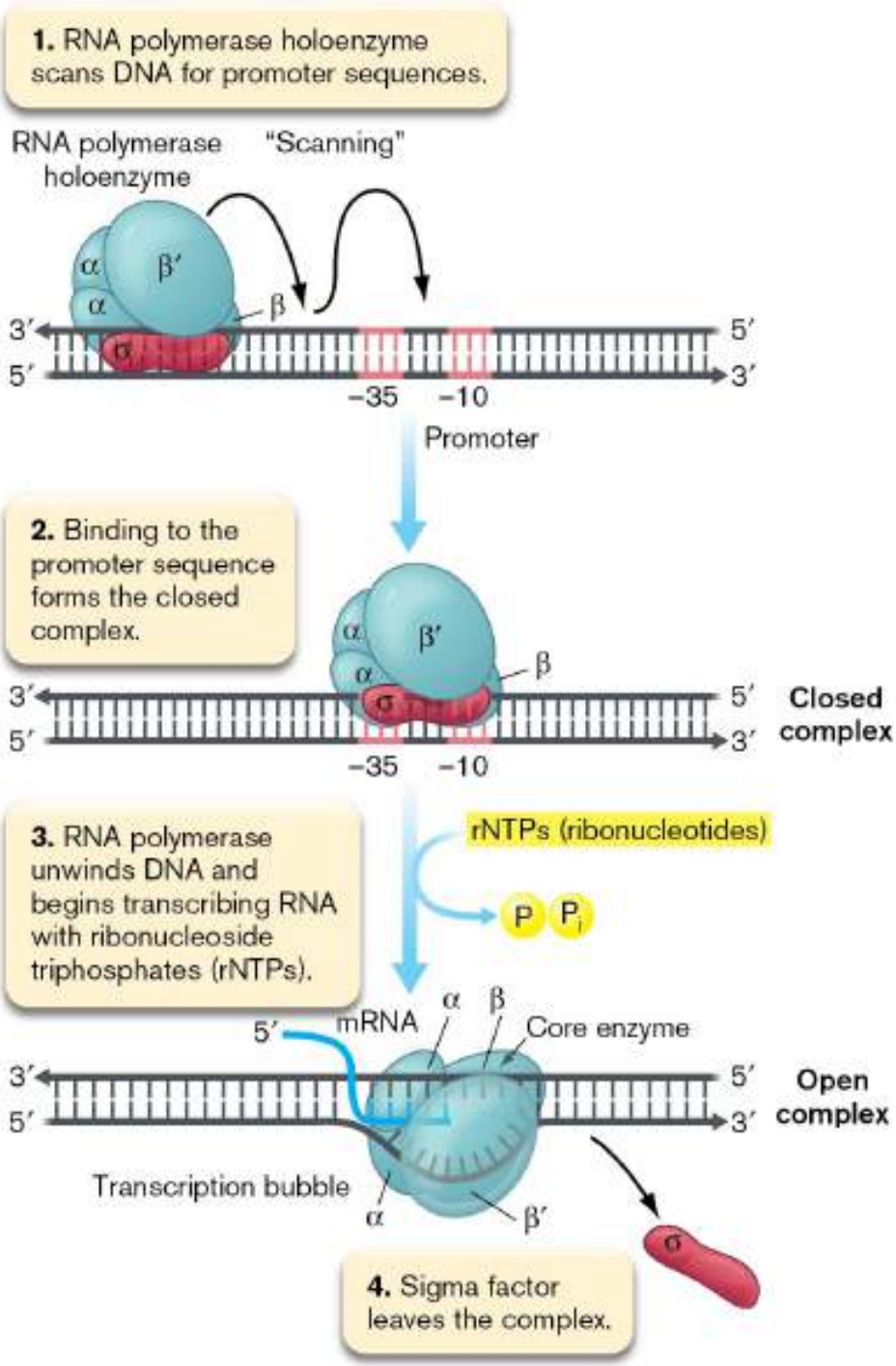
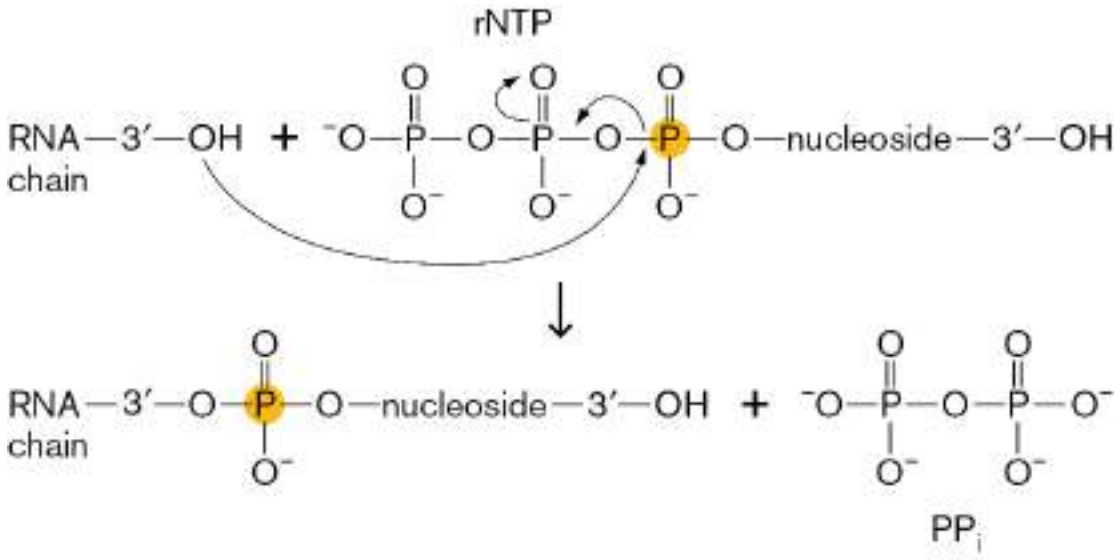


FIGURE 8.5 ■ The initiation of transcription. Sigma factor helps RNA polymerase find promoters but is discarded after the first few RNA bases are polymerized. (Omega is not shown.)

The open-complex form of RNA polymerase begins transcription. The first ribonucleoside triphosphate (rNTP) of the new RNA chain is usually a purine (A or

G). The purine base-pairs to the position designated +1 on the DNA template, which marks the start of the gene. As the enzyme complex moves along the template, subsequent rNTPs diffuse through a channel in the polymerase and into position at the DNA template. After the first base is in place, each subsequent rNTP transfers a ribonucleoside monophosphate (rNMP) to the growing chain while releasing a pyrophosphate (PP_i):



The "energy released" (actually the free energy change) after cleavage of the rNTP groups is used to form the phosphodiester link to the growing polynucleotide chain. (Free energy change, ΔG, is presented in Chapter 13 .) This is the same reaction that occurs during DNA elongation (see Chapter 7), but here it involves rNTPs instead of dNTPs (deoxyribonucleoside triphosphates).

Transcription elongation. The transition from initiation to elongation occurs when the promoter-bound sigma factor is released by the core polymerase; this step is referred to as promoter escape. Regions 3 and 4 of the sigma factor (**Fig. 8.2**) occupy the channel from which synthesized transcript exits the polymerase and escapes the promoter. Once reaching about nine bases in length, the RNA transcript is able to dislodge the sigma factor from the core polymerase and facilitate promoter escape (**Fig. 8.5** , step 4). The newly liberated sigma factor can recycle onto an unbound core RNA polymerase to direct another round of promoter binding. Meanwhile, the original RNA polymerase continues to move along the template, synthesizing RNA at approximately 45 bases per second. As the DNA helix unwinds, a 17-bp transcription bubble forms and propagates with the RNA polymerase complex. DNA unwinding generates positive DNA supercoils ahead of the advancing bubble. The supercoils are removed by DNA topoisomerase enzymes (see Section 7.2).

Transcription termination. How does RNA polymerase know when to stop? Again, the secret is in the sequence. All bacterial genes use one of two types of known transcription termination signals: either Rho-dependent or Rho-independent.

Rho-dependent termination relies on a protein called Rho and an ill-defined sequence at the untranslated 3′ end of the gene that appears to be a strong pause site. Rho factor binds to an exposed region of RNA after the ORF segment at C-rich sequences that lack obvious secondary structure. This is the transcription terminator pause site. Rho monomers assemble as a hexamer around the RNA (**Fig. 8.6A**). Then, like a person pulling a raft to shore by a rope tied to a tree, Rho pulls itself to the paused RNA polymerase by threading downstream RNA through the ring via an intrinsic ATPase activity. Once Rho touches the polymerase, an RNA-DNA helicase activity also performed by Rho appears to unwind the RNA-DNA heteroduplex, which releases the completed RNA molecule and frees the RNA polymerase.

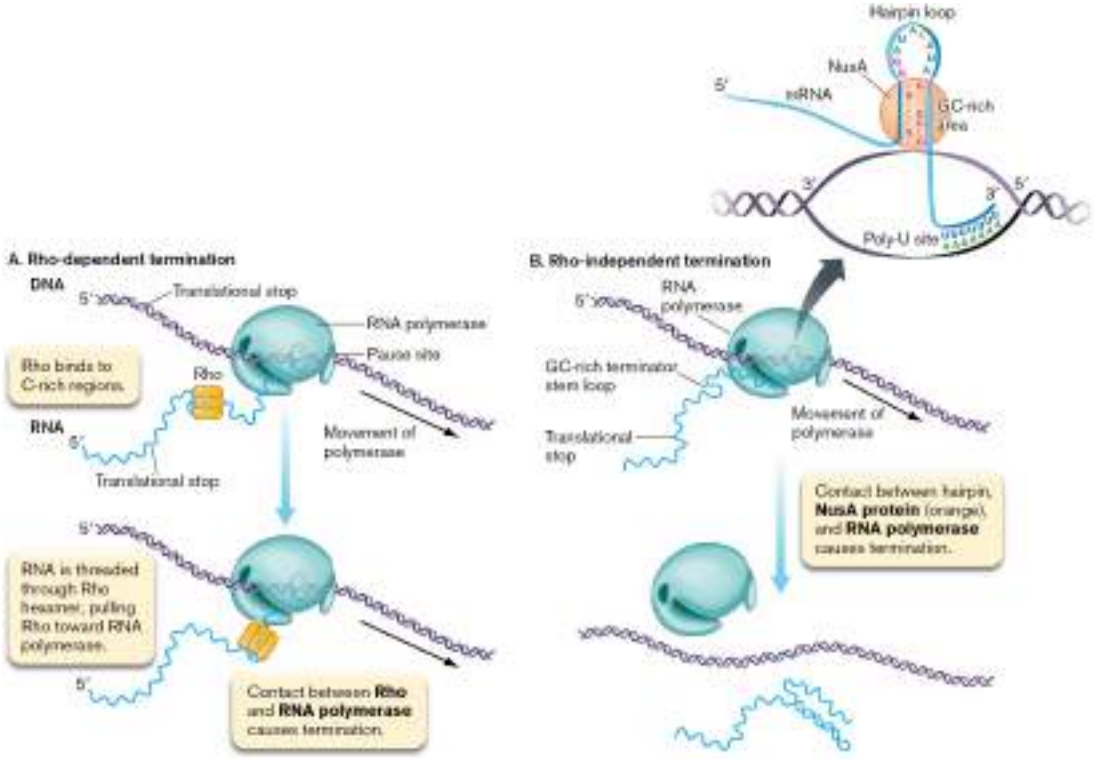


FIGURE 8.6 ■ The termination of transcription. A. Rho-dependent termination requires Rho factor but not NusA. **B.** Rho-independent termination requires NusA but not Rho factor.

DAVID SCHARF/SCIENCE SOURCE

The second type of termination, called Rho-independent termination or intrinsic termination, occurs in the absence of Rho. Rho-independent termination requires a

GC-rich region of RNA roughly 20 bp upstream from the 3′ terminus, as well as a poly-U site at the terminus: four to eight consecutive residues of uridine, the nucleoside containing the base uracil (**Fig. 8.6B**). The GC-rich sequence contains complementary bases and forms a stem loop structure that contacts RNA polymerase. Contact halts nucleotide addition, causing RNA polymerase to pause. A protein called NusA stimulates transcription pause at these sites. While the polymerase is paused, the DNA-RNA heteroduplex is weakened because the poly-U−poly-A base pairs at the 3′ terminus contain only two hydrogen bonds per pair, so they are easier to melt. Melting the hybrid molecule releases the transcript and halts transcription. The pause in polymerase movement, and thus transcription, is important to prevent tighter base-pairing downstream of the UA region.

Antibiotics Reveal RNA Synthesis Machines

To be useful in medicine, all antibiotics must ideally meet two fundamental criteria: They must kill or inhibit the growth of a pathogen, and they must not harm the host. Antibiotics used against bacteria, for instance, must selectively attack features of bacterial targets not shared with eukaryotes. Many antibiotics possess highly specific modes of action, binding to and altering the activity of one specific "machine" of one type of cell.

One example is the antibiotic rifamycin B (**Fig. 8.7A**), produced by the actinomycete *Amycolatopsis rifamycinica* (**Fig. 8.7B**). Rifamycin B and its derivatives such as rifampicin (**Fig. 8.7A**) selectively target bacterial RNA polymerase and bind to the polymerase's beta subunit near the Mg^{2+} active site, thus blocking the RNA exit channel (**Fig. 8.7C**). RNA polymerase can carry out two or three polymerization steps, but then it stops because the nascent RNA cannot exit. The rifamycins do not prevent RNA polymerase from binding to promoters, and because an mRNA molecule already passing through the exit channel masks the rifamycin-binding site, the rifamycins cannot bind to RNA polymerases already transcribing DNA. The semisynthetic derivative of rifamycin B, rifampicin (also called rifampin), is used for treating tuberculosis (caused by *Mycobacterium tuberculosis*), leprosy (caused by *M. leprae*), and bacterial meningitis (caused by *Neisseria meningitidis)* .

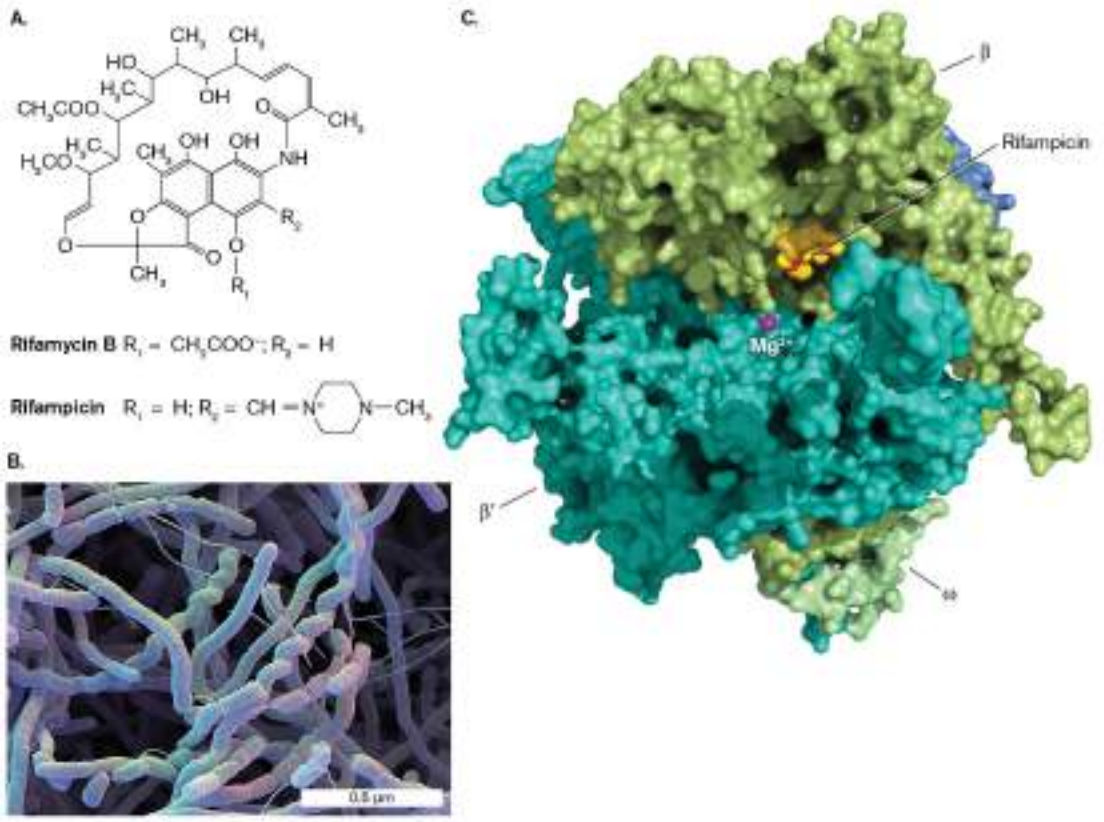


FIGURE 8.7 ■ Structure and mode of action of rifamycins. A. The basic rifamycin structure. The R groups indicated alter the pharmacology of the basic structure. **B.** Electron micrograph of *Amycolatopsis* . **C.** Rifampicin binds to the beta subunit and blocks the channel where elongating mRNA exits the RNA polymerase of *Thermus aquaticus* . (PDB code: 1I6V)

Source: Part C modified from E. A. Campbell et al. 2001. *Cell* **104** :910–912.
DAVID SCHARF/SCIENCE SOURCE

Thought Question

8.5 If rifamycins target bacterial RNA polymerase, why don't they also kill their producer, the bacterium *Amycolatopsis rifamycinica?*

Actinomycin D (**Fig. 8.8A**) is another antibiotic produced by an actinomycete. Its phenoxazone ring is a planar structure that stacks, or intercalates, between GC base pairs within DNA. Its side chains then extend in opposite directions along the minor groove (**Fig. 8.8B**). Because it mimics a DNA base, actinomycin D blocks

the elongation phase of transcription; but because it binds any DNA, it is not selective for bacteria. It can be used to treat human cancers because cancer cells replicate rapidly, but it has severe side effects because it inhibits DNA synthesis in normal cells too. (Antibiotics are discussed further in Chapter 27 .)

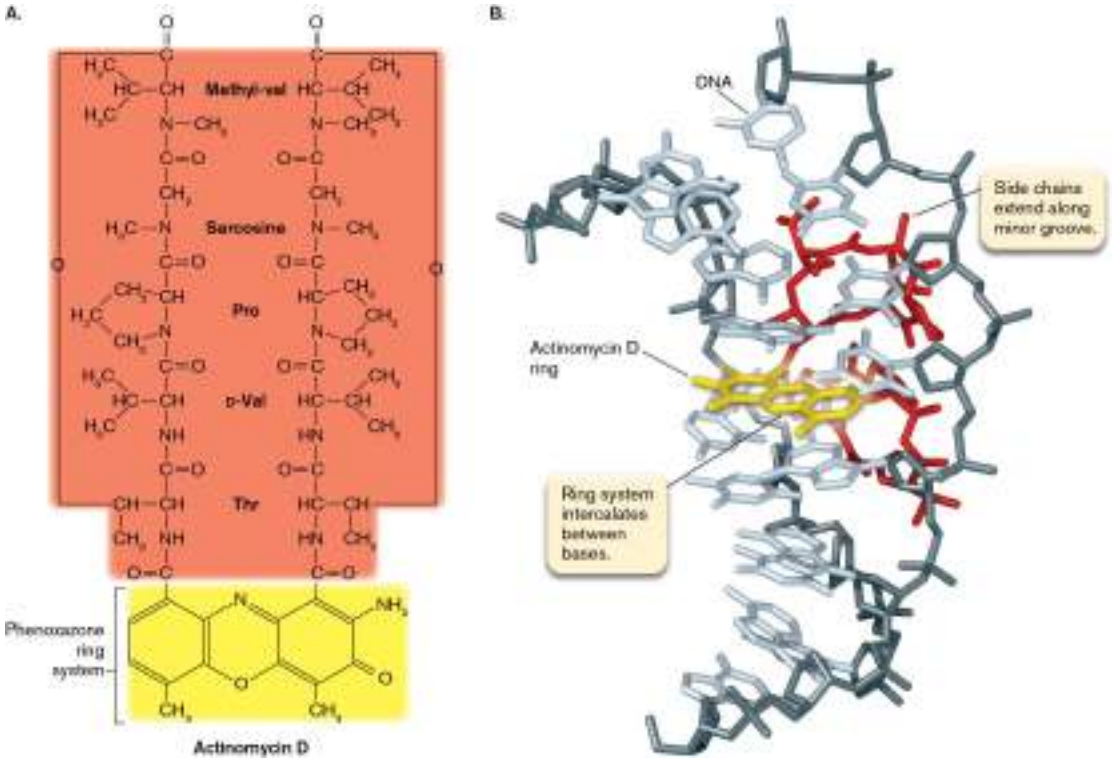


FIGURE 8.8 ■ Structure and mode of action of actinomycin D. This antibiotic inserts its ring structure (yellow) **(A)** between parallel DNA bases and wraps its side chains (red) along the minor groove **(B)** . (PDB code: 1DSC)

Antibiotics helped scientists determine the details of RNA and protein synthesis. First, purified RNA polymerase and ribosomes were used to demonstrate transcription and translation in cell-free systems (in test tubes). Then, protein subunits of RNA polymerase extracted from antibiotic-resistant and antibiotic-sensitive strains of *E. coli* were mixed together to reconstruct chimeric molecules. To determine which RNA polymerase subunit was targeted by a rifamycin molecule, for instance, RNA polymerases from rifamycin-sensitive and rifamycin-resistant strains were purified from cell extracts. The component parts of the polymerases were separated, and a chimeric RNA polymerase was reassembled (like a 3D jigsaw puzzle) by mixing subunits from antibiotic-sensitive and antibiotic-resistant polymerases. The reconstituted polymerase preparation was tested to see whether it could make RNA in the presence of rifamycin. Resistance was found to occur only when the subunit conveying resistance was added; in this case, the beta subunit.

Investigators used this information in subsequent in vitro assays and X-ray crystallography to learn more about the enzymatic mechanism of the beta subunit.

Different Classes of RNA Have Different Functions

There are six classes of RNA in bacteria, each with a different function (**Table 8.1**). As indicated in Figure 8.1 , messenger RNA (mRNA) serves the vital function of providing the template for protein synthesis. Molecules of mRNA average 1,000–1,500 bases in length but can be much longer or shorter, depending on the size of the protein they encode. The other five classes of RNA lack open reading frames and are not translated into protein. One class of untranslated RNA, called ribosomal RNA (rRNA) , forms the scaffolding on which ribosomes are built. As will be discussed later, rRNA also forms the catalytic center of the ribosome. Another class of untranslated RNA, transfer RNA (tRNA) , ferries amino acids to the ribosome. A unique property of rRNA and tRNA is the presence of unusual modified bases not found in other types of RNA.

TABLE 8.1 **Classes of RNA in Bacteria** [a]

RNA class	Function	Number of types	Average size	Approximate half-life	Unusual bases
mRNA (messenger RNA)	Encodes protein	Thousands	1,500 nt	3–5 minutes	No
rRNA (ribosomal RNA)	Part of the ribosome; synthesizes protein	3	5S: 120 nt; 16S: 1,542 nt; 23S: 2,905 nt	Hours	Yes
tRNA (transfer RNA)	Shuttles amino acids	41 (86 genes)	80 nt	Hours	Yes
sRNA [b] (small	Controls transcripti	20–30	<100 nt	Variable	No

TABLE 8.1	Classes of RNA in Bacteria [a]				
RNA class	**Function**	**Number of types**	**Average size**	**Approximate half-life**	**Unusual bases**
RNAs, or regulatory RNAs)	on, translation, or RNA stability				
tmRNA (properties of transfer and messenger RNA)	Frees ribosomes stuck on damaged mRNA	Roughly one per species	300–400 nt	3–5 minutes	No
Catalytic RNA	Carries out enzymatic reactions (e.g., RNase P)	?	Varies	3–5 minutes	No

A third important class of untranslated RNA is called small RNA (sRNA) . Molecules of sRNA do not encode proteins but are used to regulate the stability or translation of specific mRNAs into proteins. As exemplified in **Figure 8.1** , all mRNA molecules contain untranslated leader sequences that precede the actual coding region. Complementary sequences within some of these RNA leader regions snap back on themselves to form double-stranded stems and stem loop structures that can obstruct ribosome access and limit translation. Regulatory sRNAs can base-pair with these regions in mRNA and either disrupt or stabilize intrastrand stem structures.

A fourth class of untranslated RNA is designated as transfer messenger RNA (tmRNA) in reference to the properties it shares with both tRNA and mRNA. The function of tmRNA is to release ribosomes stalled on truncated mRNA. Without the aid of tmRNA, ribosomes would stall on truncated mRNA messages missing their normal translation stop codon at the 3′ end. At these stalled ribosomes, the "mRNA" section of tmRNA replaces the truncated mRNA, and the ribosome resumes translation. Translation of the tmRNA section adds a 10-amino-acid tag to the defective, truncated protein, marking it for degradation. Then a translation stop

codon at the end of the tmRNA tag region enables the ribosome to disengage and release the defective protein. The released ribosome is then free to initiate translation of another mRNA message, while the defective protein is degraded.

Catalytic RNAs , also referred to as ribozymes , constitute the fifth class of untranslated RNA. While catalytic RNA molecules are usually found associated with proteins, the enzymatic (catalytic) activity of the complex actually resides in the RNA portion rather than in the protein. All of these functional RNAs may represent remnants of the ancient "RNA world," where the earliest ancestral cells were built of RNA parts whose functions were later assumed by proteins. The RNA world model is discussed in Chapter 17 .

RNA Stability

In contrast to the DNA template, most prokaryotic mRNA transcripts are short-lived, owing to their degradation by intracellular RNases. RNA stability is measured in terms of half-life, which is the length of time the cell needs to degrade half the molecules of a given RNA species. The modified bases found in rRNAs and tRNAs render them relatively resistant to RNase digestion. These more stable RNA molecules have half-lives of the order hours. In contrast, the average half-life for mRNA is 1–3 minutes but can be as short as 15 seconds. Some mRNAs are far more stable, with half-lives of 10 minutes or longer.

Why would a cell allow the waste inherent in this short-lived use for mRNA? Microbes periodically face extremely rapid changes in their environment, including changes in temperature, salinity, or nutrients. To survive, cells must be prepared to react quickly and halt synthesis of superfluous or even detrimental proteins. An effective way to do this is to deprive the ribosomes of the mRNA templates for translation. The RNases rapidly destroy existing mRNA, while regulators that specifically block the transcription of target genes ensure that no new mRNA is made from those genes. Accordingly, rapid mRNA degradation makes it necessary to continually transcribe genes as long as their protein products are needed.

The short half-lives of mRNA are found not only in fast-growing bacteria like *E. coli* but also in slow-growing bacteria. The marine cyanobacterium *Prochlorococcus* doubles about once per day, but Claudia Steglich (**Fig. 8.9**) and colleagues at the Massachusetts Institute of Technology and the University of Freiburg showed that even in this slow-growing organism, mRNA is rapidly turned over, with a median mRNA half-life of 2.4 minutes. Thus, even slow growers rapidly recycle their mRNA, indicating that they must be primed for quick responses to environmental change.



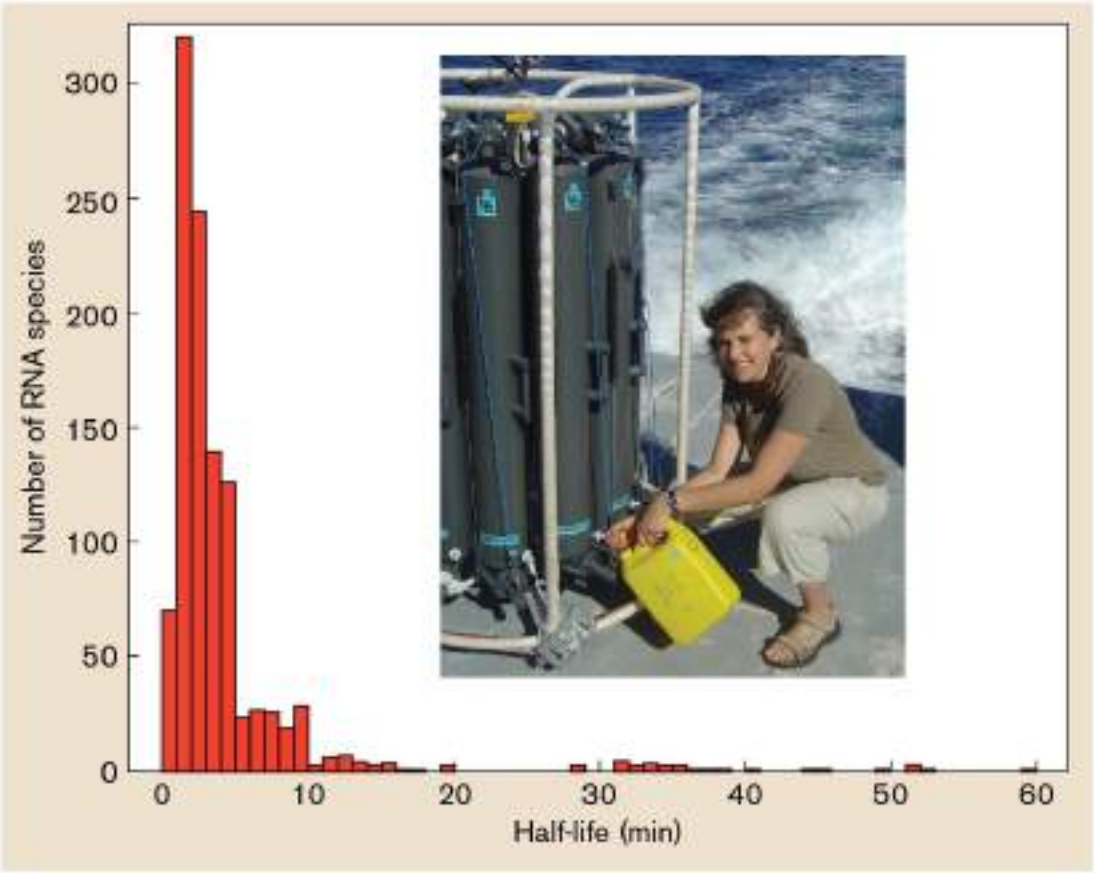


FIGURE 8.9 ■ Claudia Steglich (inset) collecting water samples for the study of *Prochlorococcus* . The half-lives of most *Prochlorococcus* transcripts are short, despite the long doubling time of this bacterium.

Source : C. Steglich et al. 2010. *Genome Biol* . **11** :R54, fig. 1b.
COURTESY OF CLAUDIA STEGLICH

Transcription in Archaea and Eukaryotes

Across all three domains—Bacteria, Archaea, and Eukarya—transcription of DNA into RNA proceeds in a similar manner, although there are important differences. Transcription is performed in all three domains by multisubunit DNA-dependent RNA polymerases. The catalytic cores of these polymerases (for example, the beta and beta-prime subunits in bacteria) are evolutionarily related and may have their origin in the RNA world (see Chapter 17) as an RNA-dependent RNA polymerase. Homologs of the alpha homodimer of bacteria are also found in the RNA polymerases of Archaea and Eukarya, and all three domains share the elongation factor NusG. Given the conservation of these key components, it is not surprising that during the elongation phase, the mechanism of RNA polymerization in all

three domains is similar. Where Archaea and Eukarya differ significantly from Bacteria is at the initiation and termination stages of transcription.

Archaea have a single RNA polymerase, while eukaryotes have three RNA polymerases. Eukaryotic RNA polymerases I and III synthesize stable, untranslated RNAs (rRNA and tRNA, respectively), while RNA polymerase II synthesizes mRNA. Structurally, the RNA polymerase of archaea is more similar to RNA polymerase II of eukaryotes than to the bacterial polymerase (**Fig. 8.10** ; also see Section 19.1). Whereas the bacterial polymerase core is composed of only 5 subunits, the archaeal and eukaryotic versions contain 12 or more.

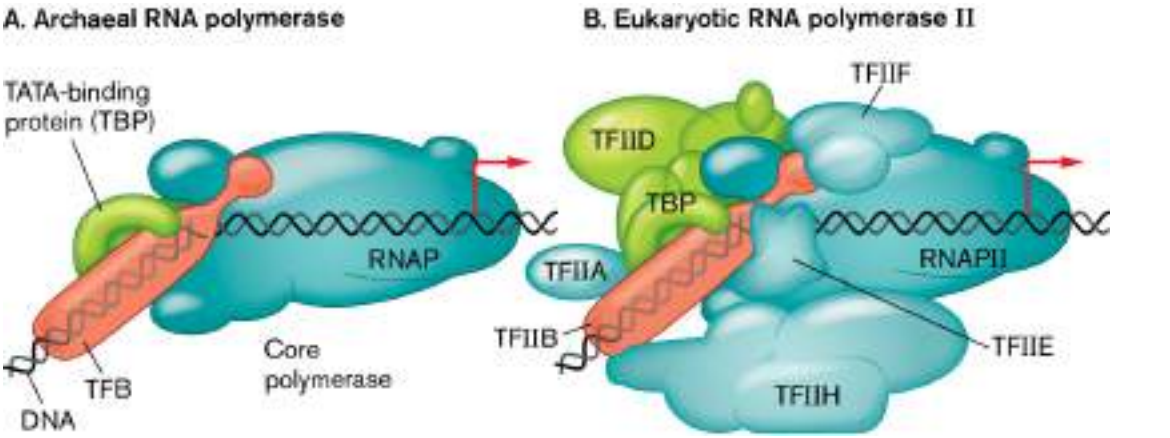


FIGURE 8.10 ■ RNA polymerases from Archaea and Eukarya exhibit homology. The core RNA polymerase (RNAP) subunits of the archaeon *Sulfolobus* **(A)** show homology to those of the eukaryotic RNA polymerase II (RNAP II) **(B)** . The TATA-binding protein (TBP; green) and transcription factor B (TFB; pink) of *Sulfolobus* also show homology to eukaryotic counterparts. The bent red arrows indicate transcription start sites.

Rather than sigma factors, eukaryotes and archaea use other proteins to recruit the RNA polymerase to the promoter; hence, the sequence motifs within the promoter are also different. The TATA-binding protein (TBP) recognizes a motif in the promoter called the TATA box. TBP bends the DNA and recruits transcription factor B (TFB in archaea, TFIIB in eukaryotes) to the TFB recognition element of the promoter. The RNA polymerase binds TBP and TFB, and this complex is sufficient to generate the transcription bubble that initiates transcription. A third transcription factor, called TFE in archaea and TFIIE in eukaryotes, also plays a role in initiation when conditions are suboptimal. Eukaryotes employ additional initiator proteins, but these three are sufficient for initiation at strong promoters. The initiator proteins remain at the promoter (TBP, TFB/TFIIB) or are removed from the RNA polymerase shortly after chain elongation begins (TFE/TFIIE).

Note: Because the transcription machinery of archaea is different from that of bacteria, archaea are usually resistant to antibacterial antibiotics that target those processes.

Transcription is terminated in archaea by a mechanism that is similar to Rho-independent termination in bacteria: a string of U's at the 3' UTR produces a weak RNA-DNA hybrid molecule that easily melts to release a completed mRNA. Termination in archaea differs from termination in bacteria because it does not seem to require adjacent RNA stem loop structures. And unlike bacteria, archaea have many genes with multiple terminators, which can result in mRNAs with different 3' UTR lengths depending on the terminator used. Termination in eukaryotes is less well understood, but it appears to involve specific terminator protein complexes for different types of RNA.

As in bacteria, many genes in archaea are found in operons. The mRNA transcripts from these operons will contain open reading frames for the translation of two or more genes. In contrast, most eukaryotic mRNAs are monocistronic. Like bacteria, archaea show much less posttranscriptional processing of mRNA than eukaryotes do. Archaeal mRNAs do not contain a 7-methlyguanosine cap on the 5' ends of the mRNA and are not polyadenylated on their 3' ends. While some archaeal genes contain introns, which are spliced out of the mRNA prior to translation, introns are not nearly as common as they are in eukaryotic genes. Thus, archaeal transcription shares features with both bacteria and eukaryotes.

To Summarize

- **RNA polymerase holoenzyme** , consisting of core RNA polymerase and a sigma factor, initiates transcription of a DNA template strand.
- **Sigma factors** help core RNA polymerase locate consensus promoter sequences near the beginning of a gene. The sequences identified and bound by *E. coli* sigma-70 are located approximately −10 and −35 bp upstream of the transcription start site.
- **Different sigma factors** help initiate transcription of different sets of genes.
- **RNA polymerase binds promoter DNA** , forming the closed complex.
- **The DNA strands separate** (melt) and form a "bubble" of DNA around the polymerase called the **open complex** .
- **Rho-dependent and Rho-independent** mechanisms mediate transcription termination.
- **Antibiotics that inhibit bacterial transcription include rifamycin B** (which binds to RNA polymerase to inhibit transcription initiation) and

actinomycin D (which binds DNA to nonselectively inhibit transcription elongation).

- **Messenger RNA** (mRNA) molecules encode proteins.
- **Noncoding RNAs** include **ribosomal RNAs** (rRNAs), **transfer RNAs** (tRNAs), and **small RNAs** (sRNAs). Some noncoding RNAs can regulate gene expression (sRNA), have catalytic activity (catalytic RNA, or ribozymes), or function as a combination of tRNA and mRNA (tmRNA).
- **The transcription machinery of archaea** is related more to eukaryotes than to bacteria, but the **archaeal organization of genes into operons** and general **lack of posttranscriptional processing** is more similar to bacteria.

Glossary

RNA polymerase
: Also called *DNA-dependent RNA polymerase* . An enzyme that produces an RNA complementary to a template DNA strand.

DNA-dependent RNA polymerase
: See **RNA polymerase** .

transcript
: An RNA copy of a DNA template.

template strand
: A DNA strand (or an RNA strand in some viruses) that is used as a template for the synthesis of mRNA.

sigma factor
: A protein needed to bind RNA polymerase for the initiation of transcription in bacteria.

consensus sequence
: A sequence of nucleotides or amino acids with a common function at many nucleic acid or protein positions. Consists of the base pair or amino acid most frequently found at each position in the sequence.

Rho factor
: A bacterial protein involved in terminating transcription.

ribosomal RNA (rRNA)
: An RNA molecule that includes the scaffolding and catalytic components of ribosomes.

transfer RNA (tRNA)
: An RNA that carries an amino acid to the ribosome. The anticodon on the tRNA base-pairs with the codon on the mRNA.

small RNA (sRNA)

A non-protein-coding regulatory RNA molecule that modulates translation or mRNA stability.

transfer messenger RNA (tmRNA)

A molecule resembling both tRNA and mRNA that rescues ribosomes stalled on damaged mRNAs lacking a stop codon.

catalytic RNA

Also called *ribozyme* . An RNA molecule that is capable of catalyzing reactions.

ribozyme

See **catalytic RNA** .

Figure 10.10

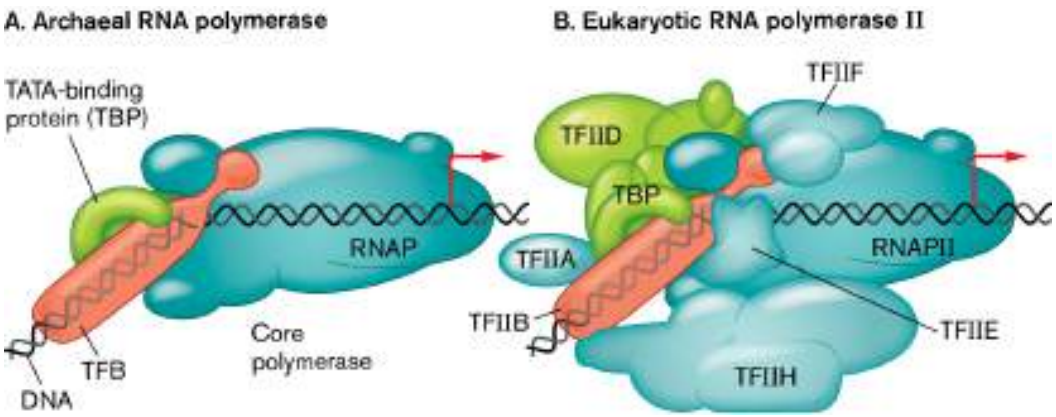


FIGURE 8.10 ■ RNA polymerases from Archaea and Eukarya exhibit homology. The core RNA polymerase (RNAP) subunits of the archaeon *Sulfolobus* **(A)** show homology to those of the eukaryotic RNA polymerase II (RNAP II) **(B)** . The TATA-binding protein (TBP; green) and transcription factor B (TFB; pink) of *Sulfolobus* also show homology to eukaryotic counterparts. The bent red arrows indicate transcription start sites.

Figure 8.1

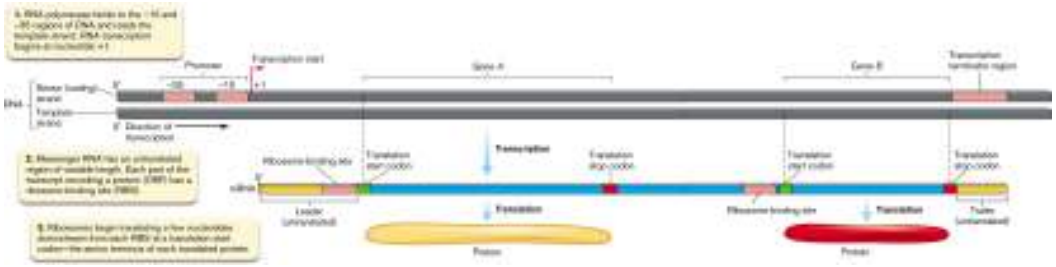


FIGURE 8.1 ■ Alignment of structural genes in a bacterial operon, the mRNA transcript, and protein products. In this figure, the term "gene" refers to the region of DNA that encodes a product. In this example, both genes encode protein. ORF = open reading frame.

Fig. 7.4

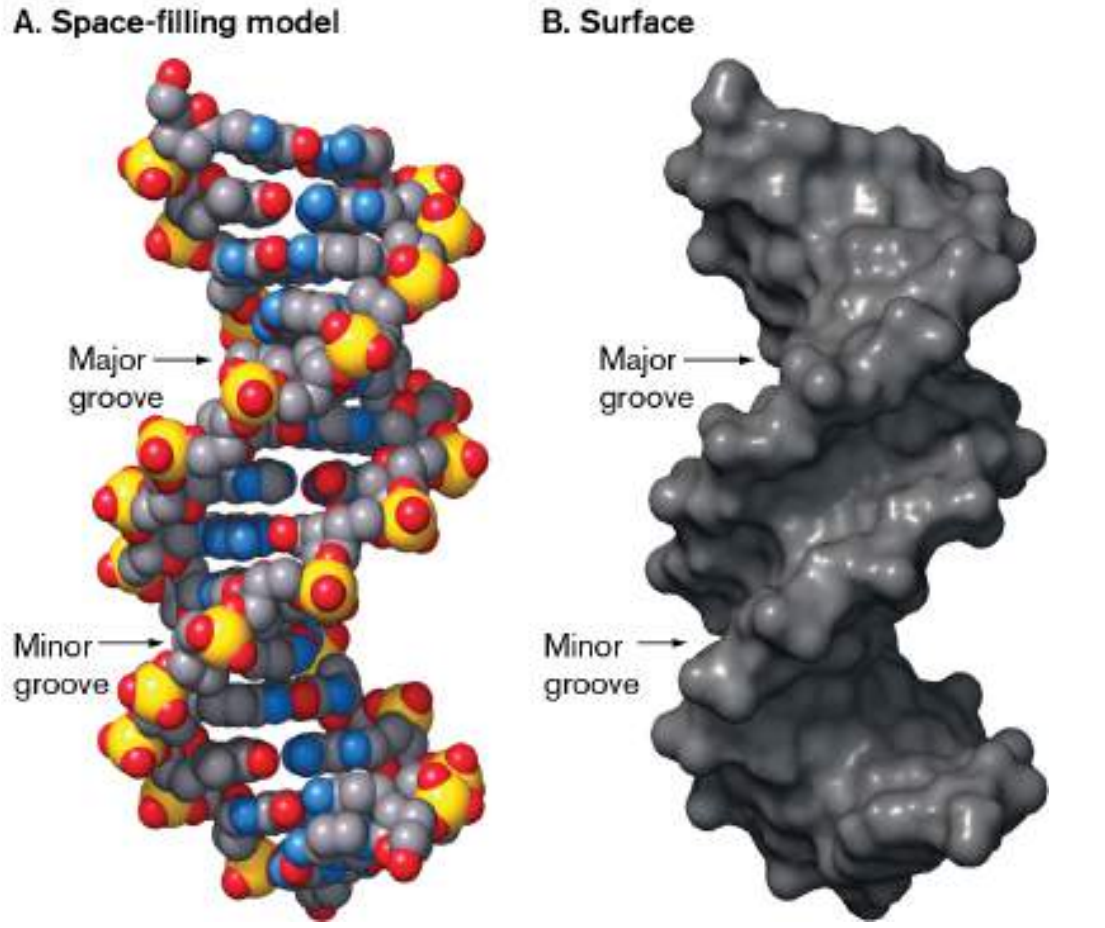


FIGURE 7.4 ■ Models of DNA. A. Space-filling model of DNA. **B.** DNA surface, modeled using nuclear magnetic resonance. (PDB code: 1K8J)

Endnotes

1. Note a: Six major classes of RNA exist in all bacteria. They differ in function, quantity, average size, and half-life, as well as whether they contain modified bases. Return to reference a
2. Note b: Small RNAs include antisense RNA and micro-RNA. Return to reference b

8.3 Translation of RNA to Protein

The Genetic Code

Once a gene has been transcribed into mRNA, the next stage is translation, the decoding of the RNA message by the ribosome to synthesize protein. Recall that there are four bases of RNA but 20 common amino acids. How does a ribosome read an alphabet with four letters to make words with an alphabet that uses 20 letters? Researchers in the 1960s suspected that the ribosome could read multiple adjacent bases simultaneously, such that different combinations of bases could represent each of the amino acids. The stretch of bases that would encode an amino acid was termed a codon .

How many bases long would a codon need to be to cover all 20 amino acids? A codon that is one base long is clearly insufficient, as it could specify only four different amino acids. A two-base codon likewise falls short, providing only 4 × 4, or 16, combinations of bases to cover all 20 amino acids. Through a careful analysis of mutants of *Escherichia coli* , Francis Crick and colleagues discovered that codons are triplets of bases. This meant the genetic code had a total of 4 × 4 × 4, or 64, codons.

Once the triplet nature of the genetic code was discovered, the next critical step was to identify which amino acid each of the 64 codons represents. RNA was the suspected, but unproven, codon-containing substrate for translation. Marshall Nirenberg and his postdoctoral associate Heinrich Matthaei at the National Institutes of Health (**Fig. 8.11A**) designed a cell-free system (a cell lysate of *E. coli*) to test the RNA hypothesis. They used RNA molecules, synthesized by Maxine Singer (**Fig. 8.11B**), that contained simple, known, repeated sequences, such as poly-A (polyadenylic acid), poly-U (polyuridylic acid), poly-AAU, and poly-ACAC.

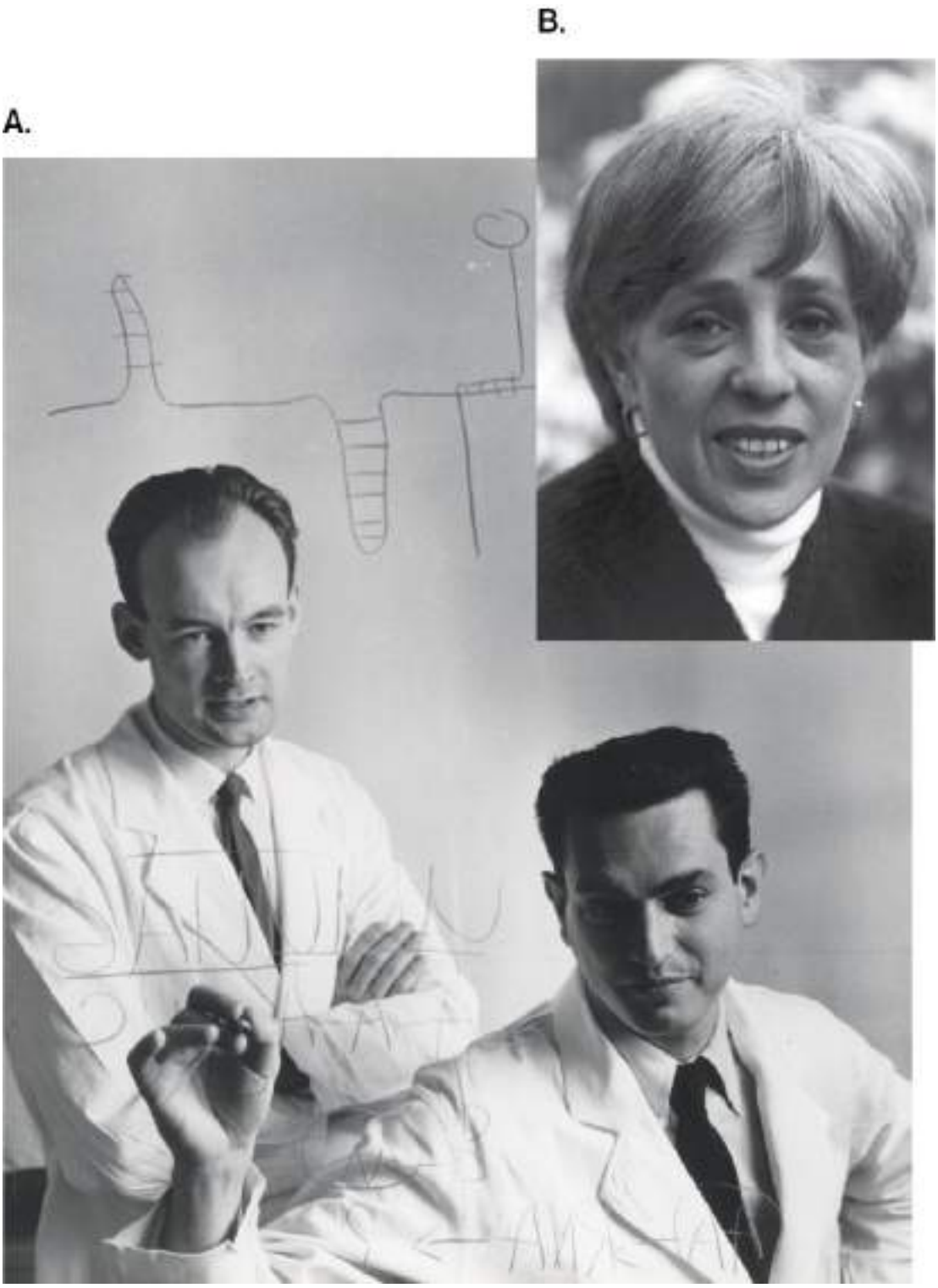


FIGURE 8.11 ■ Many scientists have contributed to our understanding of the genetic code. A. Heinrich Matthaei (left) and Marshall Nirenberg (right) were the first to crack the

genetic code. An early hand-drawn model of what would eventually be known as translation is behind them. **B.** Maxine Singer, a key contributor to the genetic code experiments, also helped develop guidelines for recombinant DNA research.

COURTESY OF NATIONAL LIBRARY OF MEDICINE/NATIONAL INSTITUTES OF HEALTH
NATIONAL CANCER INSTITUTE

These RNA molecules were tested to see whether the repetitive sequence might direct incorporation of a specific amino acid into a protein. Each synthetic polynucleotide was tested in the presence of a radiolabeled amino acid. If the radioactive amino acid was incorporated into a polypeptide, the polypeptide would also be radioactive. On the morning of May 27, 1960, the results of the experiment showed that the poly-U RNA specified the assembly of radioactive polyphenylalanine. This result meant that the triplet UUU encoded phenylalanine. It was the first breakthrough in cracking the genetic code.

Through painstaking effort, Nirenberg, Har Gobind Khorana, and colleagues deciphered the remainder of the genetic code. For their work, Nirenberg and Khorana, together with Robert W. Holley for his work on tRNA, shared the 1968 Nobel Prize in Physiology or Medicine. Remarkably, and with few exceptions, the code operates universally across species.

In the standard genetic code (**Fig. 8.12**), we can see that most amino acids have multiple codon synonyms. Glycine (Gly), for example, is translated from four different codons, and leucine (Leu) from six, but methionine (Met) and tryptophan (Trp) each have only one codon. Because more than one codon can encode the same amino acid, the code is said to be degenerate, or redundant. Notice that in most cases, synonymous codons differ only in the last base. How the cell handles this degeneracy (redundancy) in the code will be explained later.

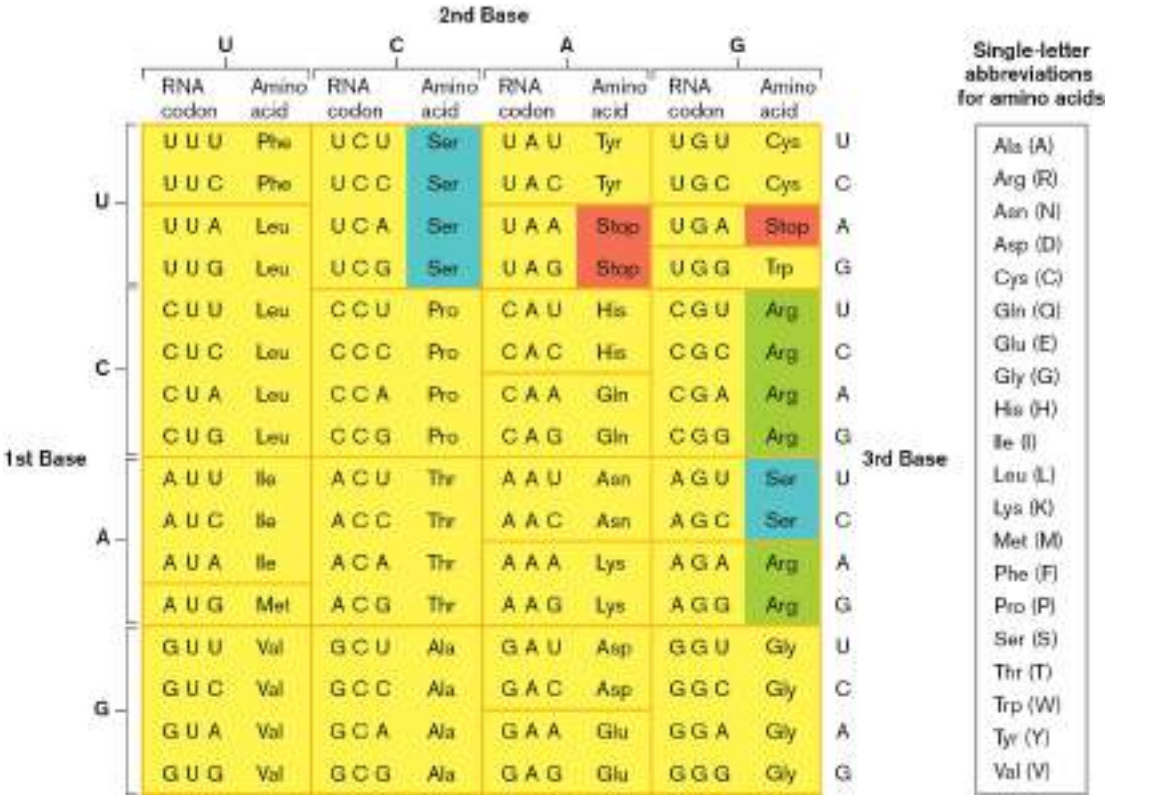


FIGURE 8.12 ■ The standard genetic code. Codons within a single box encode the same amino acid. Blue- and green-highlighted amino acids are encoded by codons in two boxes. Stop codons are highlighted red. Often, single-letter abbreviations for amino acids are used to convey protein sequences (see legend).

Only 61 out of a possible 64 codons specify amino acids. The three triplets that do not encode amino acids (UAA, UAG, and UGA) are equally important, as they tell the ribosome when to stop reading a gene sentence. Called stop codons , they trigger a series of events (to be described later) that dismantle ribosomes from mRNA and release a completed protein.

Thought Question

8.6 How might the redundancy of the genetic code be used to establish evolutionary relationships between different species? *Hints:* 1. Genomes of different species have different overall GC content. 2. Within a given genome, one can find segments of DNA sequence with a GC content distinctly different from that found in the rest of the genome.

Anticodons and tRNA Molecules

Translation is the process of reading (decoding) the string of codons in mRNA to make a string of amino acids. Decoding mRNA for translation is biochemically different from decoding DNA for transcription. Specificity during transcription relies on the ability of incoming RNA nucleotides to form complementary base pairs with the bases along the DNA template strand. In contrast, during translation the mRNA template is not directly contacted by the amino acids it encodes. Rather, amino acids are attached to small adapter RNAs, called tRNAs, which have RNA sequences called anticodons that match and bind to specific codons on the mRNA being translated.

Approximately 80 bases in length (**Fig. 8.13A**), a tRNA laid flat looks like a cloverleaf with three loops (**Fig. 8.13B**). In nature, however, the molecule folds into the "boomerang-like" 3D structure shown in **Figure 8.13C** . The bottom loop of all tRNA molecules (as depicted in **Fig. 8.13B** and **C**) harbors the anticodon triplet that base-pairs with codons in mRNA (**Fig. 8.14**). As a result, this loop is called the anticodon loop. Notice that codon-anticodon pairings are aligned in an antiparallel manner. Most tRNA molecules begin with a 5′ G, and all end with a 3′ CCA, to which an amino acid attaches via an ester link to the ribose of the terminal adenosine (see **Figs. 8.13** and **8.14**). Because the 3′ end of the tRNA accepts the amino acid, it is called the acceptor end.

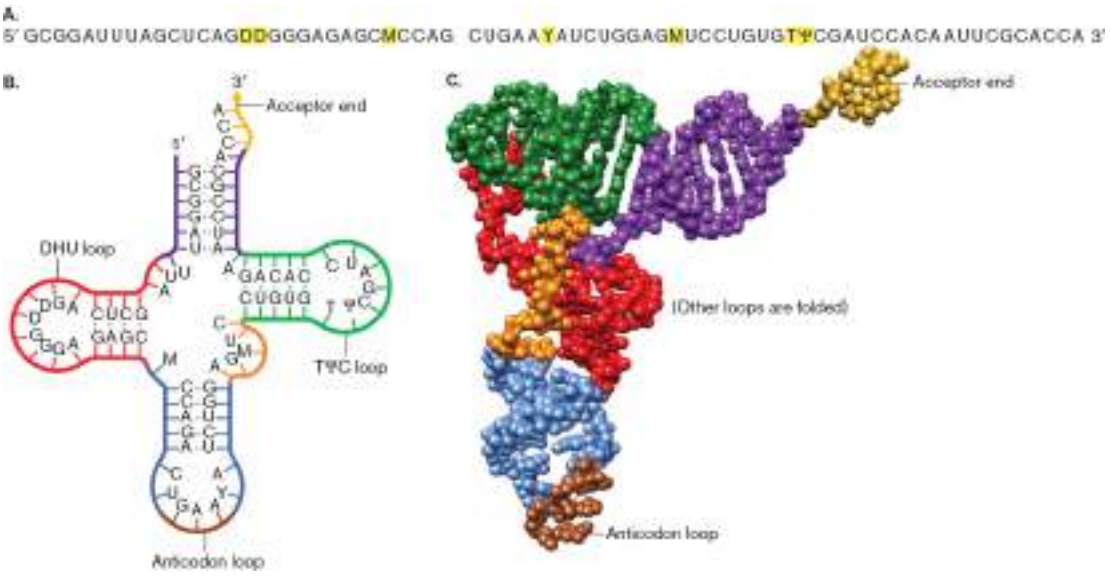


FIGURE 8.13 ■ Transfer RNA. A. Primary sequence of a tRNA molecule. The letters D, M, Y, T, and Ψ stand for modified bases found in tRNA. **B.** Cloverleaf structure. DHU (or D) is dihydrouracil, which occurs only in the DHU loop; TΨC consists of thymine, pseudouridine, and cytosine bases that occur as a triplet in the TΨC loop. The DHU and TΨC loops are named for the modified nucleotides that are characteristically found there. **C.** Three-dimensional structure. The anticodon loop binds to the codon, while the acceptor end binds to the amino acid. (PDB code: 1GIX)

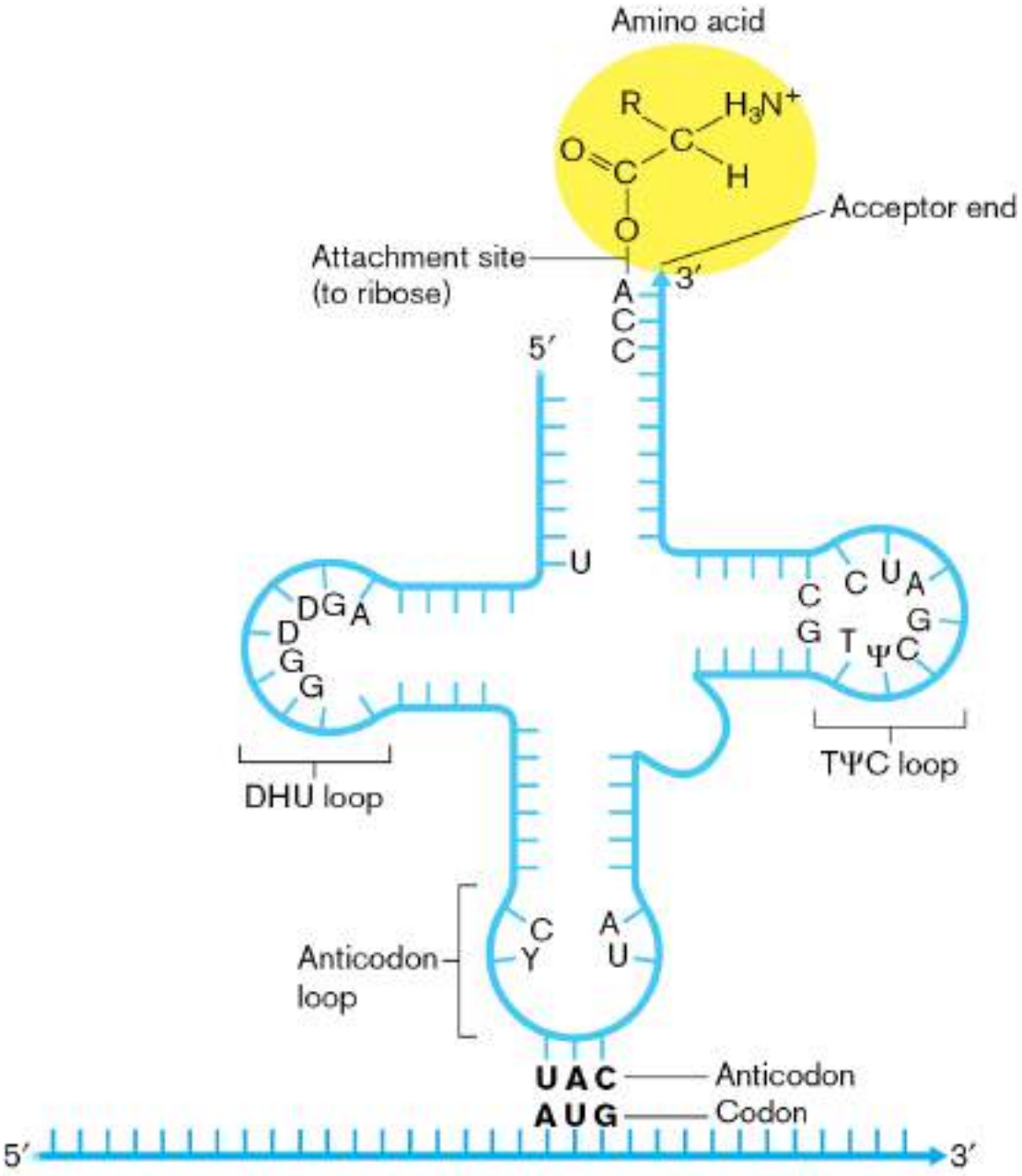


FIGURE 8.14 ■ Codon-anticodon pairing. The tRNA anticodon consists of three nucleotides at the base of the anticodon loop. The anticodon hydrogen-bonds with the mRNA codon in an antiparallel fashion. This tRNA is "charged" with an amino acid covalently attached to the 3′ end via an ester link to the ribose of the terminal adenosine.

As noted previously, in the genetic code many amino acids have codon synonyms. This redundancy is found primarily in the third position of the codon (for example, UU U and UU C both encode phenylalanine). A single tRNA for phenylalanine can recognize both codons because of "wobble" in the first position of the anticodon, which corresponds to the third position of the codon (remember, as with DNA, base-pairing between RNA strands is antiparallel). The wobble is due in part to the curvature of the anticodon loop and to the use of an unusual base (inosine) at this position in some tRNA molecules. The wobble structure allows one anticodon to pair with several codons differing only in the third position. This wobble property has allowed many microbes to economize and cover all 61 amino acid–encoding codons with fewer than 61 tRNAs in their genomes.

Aminoacyl-tRNA Synthetases Attach Amino Acids to tRNA

When a tRNA anticodon (for instance, GCG) pairs with its complementary codon (CGC in this case) during translation, the ribosome has no way of checking that the tRNA is attached (that is, charged) to the "correct" amino acid (here, arginine). Consequently, each tRNA must be charged with the proper amino acid before it encounters the ribosome. How are amino acids correctly matched to tRNA molecules and affixed to their 3′ ends?

Matching and attaching the correct amino acid to the correct tRNA is the function of enzymes called aminoacyl-tRNA synthetases . Every aminoacyl-tRNA synthetase has a specific binding site for its cognate (matched) amino acid. Each enzyme also has a site that recognizes the correct tRNA and an active site that joins the carboxyl group of the amino acid to the 3′ OH (class II synthetases) or 2′ OH (class I synthetases) of the tRNA by forming an ester (**Fig. 8.15**). The first step in the charging of tRNA is to activate the amino acid, forming an aminoacyl-AMP molecule via a reaction with ATP. Second, the amino acid is transferred to the hydroxyl residue of the terminal

adenine of the tRNA, resulting in the release of AMP. Finally, the tRNA synthetase disengages and releases the charged tRNA. Amino acids initially attached to the 2′ OH are then moved to the 3′ OH position.

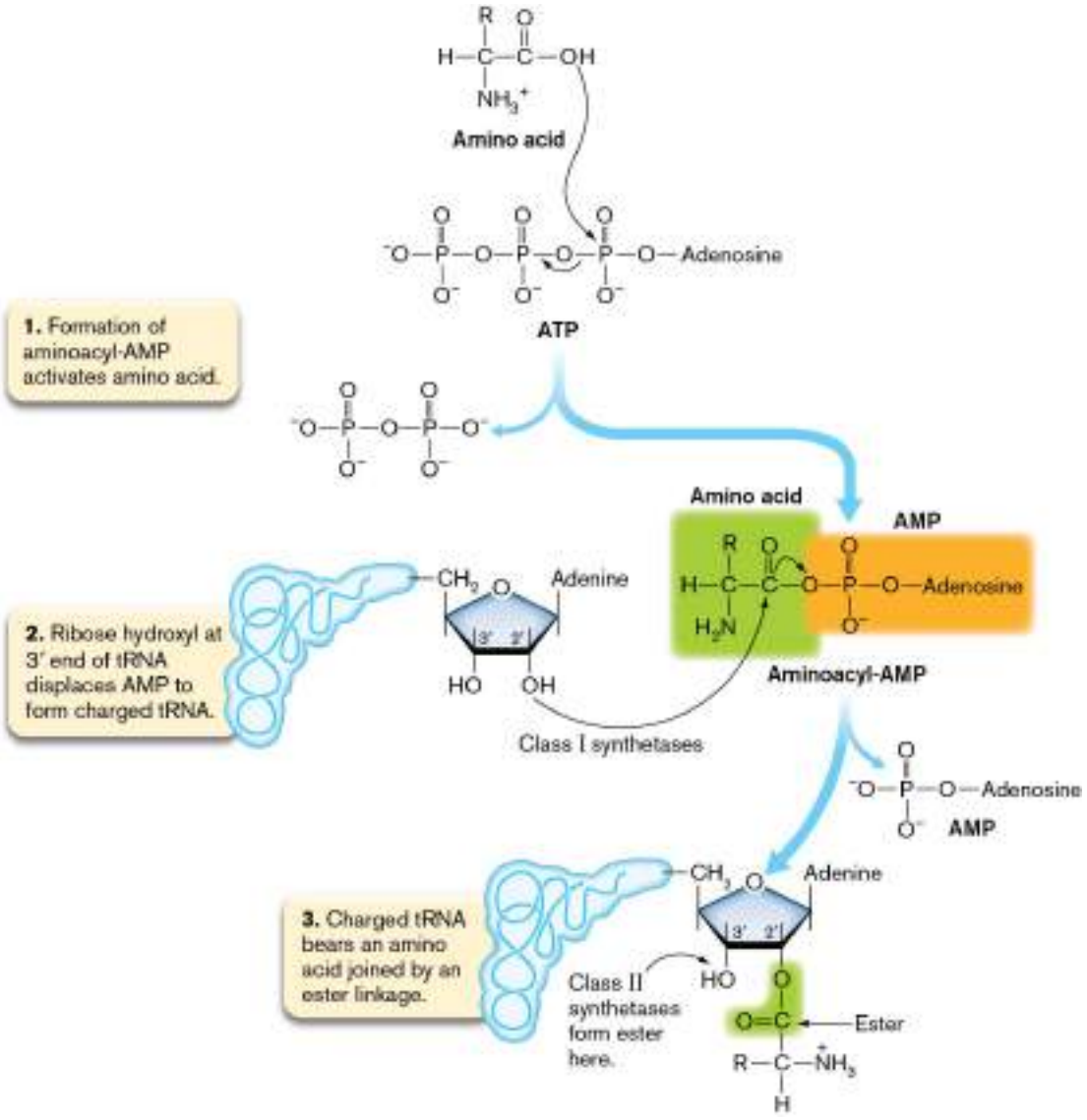


FIGURE 8.15 ■ Charging of tRNA molecules by aminoacyl-tRNA synthetases. At the end of this process, each amino acid is attached to the 3′ end of a specific tRNA molecule. Curved arrows indicate nucleophilic attack by electrons.

Note: Most bacteria have 20 aminoacyl-tRNA synthetases—one for each amino acid. Some bacteria, however, have only 18, choosing instead to modify already-charged glutamyl-tRNA and aspartyl-tRNA by adding an amine to make glutamine and asparagine derivatives.

Transfer RNAs that have different RNA sequences but carry the same amino acid are referred to as isoacceptors. Each aminoacyl-tRNA synthetase must recognize its own set of tRNA isoacceptors but not bind to any other tRNA. Specificity is based on recognizing unique features of the tRNA located within the three loops. These unique features often involve unusual modifications to the bases, which accounts for the strange letter codes seen in **Figure 8.13A** (for example, D and Ψ). Structures of some of the odd bases are shown in **Figure 8.16** . Wybutosine (yW), for example, has three rings instead of the two found in a normal purine. In the cloverleaf structure of tRNA, two of the loops are named after modifications that are invariantly present in those loops. One is called the TΨC loop because in every tRNA this loop has the nucleotide triplet thymidine–pseudouridine (Ψ)–cytosine. The other loop always contains dihydrouridine and is called the DHU loop (see **Figs. 8.13** and **8.14**).

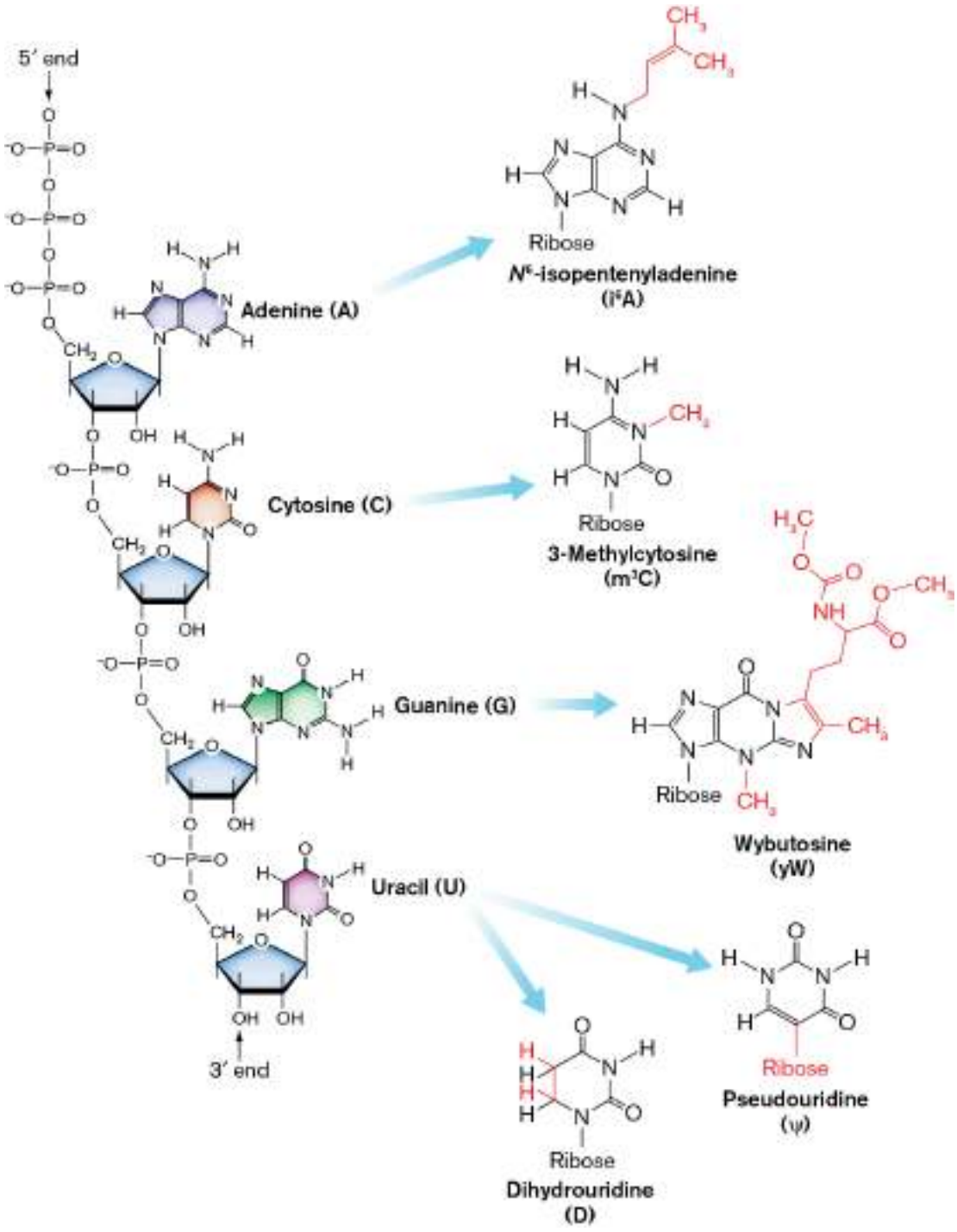


FIGURE 8.16 ■ Modified bases in tRNA and rRNA molecules. Modifications to the canonical RNA bases are highlighted in red.

How do these unusual bases end up in tRNA? During transcription of the tRNA genes, normal, unmodified bases are incorporated into the transcript. Some of these are modified later by specific enzymes to make inosine and other odd bases. The remarkable stability of tRNA molecules is explained, in part, by these unusual bases, because they are poor substrates for RNases.

The Ribosome, a Translation Machine

Ribosomes catalyze the linkage of amino acids during translation, using mRNA as the code and charged tRNAs as the source of amino acids. A ribosome is a massive complex of protein and ribosomal RNA (rRNA). Whereas amino acid and nucleotide monomers average 110 and 325 daltons (Da), respectively, a ribosome is over 1,000,000 Da. The rRNA component forms the core structure of the enzyme, including the binding sites for tRNA, as well as the channels for mRNA and the elongating polypeptide. Remarkably, it is also the rRNA component, not the protein, that catalyzes the amino acid linkages (peptide bonds), through an enzyme called peptidyltransferase . Peptidyltransferase is a ribozyme , which is an RNA molecule that carries out catalytic activity. Proteins surrounding this active center offer structural assistance to ensure that the RNA is folded properly and to interact with tRNA substrates.

Ribosomes are composed of two complex subunits, each of which includes rRNA and protein components. In bacteria and archaea, the subunits are named 30S and 50S for their "size" in Svedberg units. Svedberg units represent the rate at which a molecule sediments under the centrifugal force of a centrifuge (discussed in eAppendix 3). Within the living cell, the two subunits exist separately but come together on mRNA to form the functional 70S ribosome (**Fig. 8.17**).

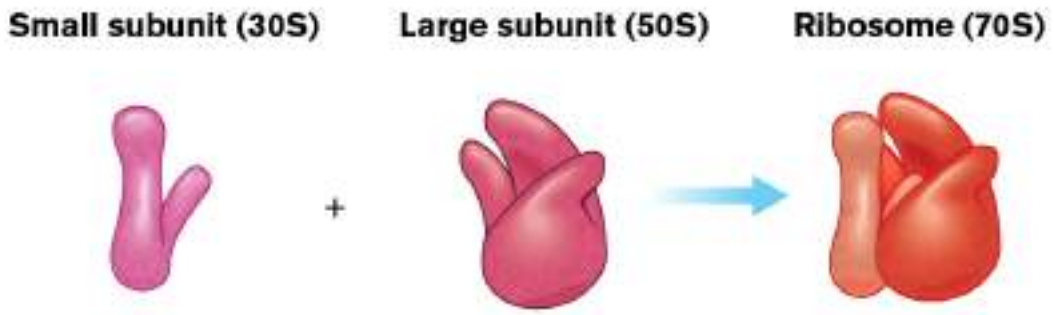


FIGURE 8.17 ■ Bacterial ribosome structure. As this schematic illustrates, a section of the 30S subunit fits into the valley of the 50S subunit when forming the 70S ribosome.

Note: Because they represent not a mass but a rate of sedimentation in a centrifuge, Svedberg units are not directly additive. That's why 30S and 50S subunits combine to form a 70S, not 80S, ribosome.

The 30S subunit (also called the small subunit, or SSU) of *E. coli* contains 21 ribosomal proteins assembled around one 16S rRNA molecule. The 16S rRNA (a SSU rRNA) molecule forms the channel for mRNA and the binding sites for tRNA, and it matches codons with anticodons during translation. The 50S subunit (also called the large subunit, or LSU) consists of 33 proteins formed around two rRNA molecules (5S and 23S). **Figure 8.18** presents the 3D spatial arrangement of rRNA and protein in the 50S subunit. Note that the majority of the ribosome is RNA. The 23S rRNA (a LSU rRNA) mediates the peptidyltransferase activity of the ribosome (**Fig. 8.18** blowup, loop V).

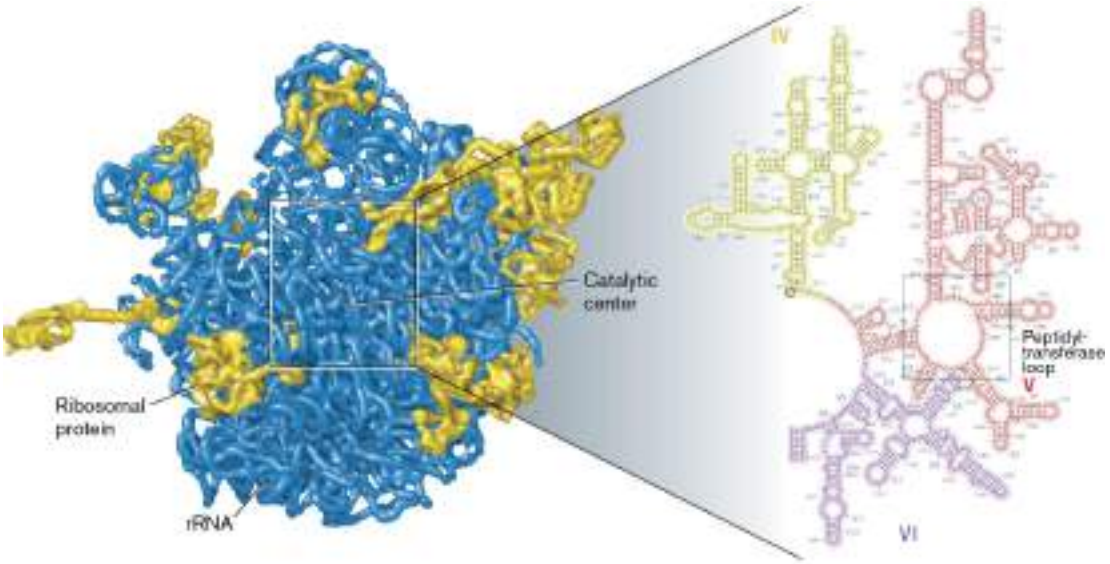


FIGURE 8.18 ■ RNA-protein interfaces in the large (50S) ribosomal subunit. Blue = rRNA; gold = proteins. (PDB code: 1GIY) **Blowup:** Partial secondary structure of 23S rRNA, which includes the peptidyltransferase catalytic site. Note the many double-stranded hairpin structures that fold into domains (Roman numerals). Nucleotides are numbered in black; stem structures in blue.

How does such a complex molecular machine get built? Assembly begins with the transcription of the genes that encode rRNA. These genes in the DNA are sometimes called ribosomal DNA (rDNA). The 16S, 23S, and 5S rRNAs are initially transcribed as one RNA molecule and posttranscriptionally processed into separate rRNA molecules. During rRNA transcription, ribosomal proteins begin to assemble within the developing secondary rRNA structures. Thus, the ribosome is built by precise, timed molecular RNA-RNA, RNA-protein, and protein-protein interactions.

The ribosome is an ancient enzyme that is shared by all three domains of life. In size and structure, archaeal and bacterial ribosomes are similar, although some ribosomal proteins in archaea are not found in bacteria, and vice versa. Eukaryotic ribosomes are

similar in overall structure, but larger and more complex than those of bacteria and archaea. The eukaryotic SSU and LSU are 40S and 60S in size, respectively, and the holoenzyme is 80S. The eukaryotic homologs of the bacterial 16S and 23S rRNA genes are longer, and consequently have larger sizes: 18S and 28S, respectively (for yeast the latter is 25S). Eukaryote LSUs contain a 5S rRNA, like bacteria, but also contain a 5.8S rRNA which has no counterpart in bacteria.

Despite differences in the length of rRNA and the sets of protein subunits involved, the ribosomes of bacteria, archaea, and eukaryotes have very similar mechanisms of translating mRNA into protein. The sequences of ribosomal RNAs from all microbes are very similar, in large part because rRNA plays an important catalytic role in the activities of ribosomes. But there are differences in rRNA sequences that increase in relation to the evolutionary distance between species. In fact, the close but not perfect similarity in the rRNA sequences allowed Carl Woese to argue successfully for the existence of the archaea as a domain distinct from bacteria and eukaryotes (see Chapter 17).

How Do Ribosomes Find Where to Start Translation?

The ribosome reads the mRNA as a string of consecutive codons, without overlap and without spacers between the codons. Because there are three bases per codon, each mRNA has three potential reading frames. For instance, in the sequence AUGCCAAA, one reading frame would include the codons AUG and CCA, the second reading frame would include UGC and CAA, and the third, GCC and AAA. How, then, is the right reading frame found by the ribosome, and how does the ribosome know where to begin translation along the mRNA?

Start codons mark where translation starts and set the correct reading frame. AUG is the most frequent start codon (90%), while GUG (8.9%), UUG (1%), and CUG (0.1%) are also used to lesser extents. Although these start codons are used to begin all proteins, they are not restricted to that role; they also encode amino acids

found in the middle of coding sequences. Consequently, there must be something else about mRNA that signifies whether an AUG codon, for example, marks the start of a protein.

The key is a section of untranslated RNA sequence positioned upstream of the protein-coding segment of mRNA (see Fig. 8.1). This leader RNA contains a purine-rich consensus sequence (5′-AGGAGGU-3′) located four to eight bases upstream of the start codon. This upstream sequence is called the ribosome-binding site , or Shine-Dalgarno sequence , after Lynn Dalgarno and her student John Shine at Australian National University, who discovered it in 1974.

In *E. coli* , the Shine-Dalgarno site is complementary to a sequence of 16S rRNA (5′-ACCUCCU-3′) found in the 30S ribosomal subunit. Binding of mRNA to this site positions the start codon (such as AUG or GUG) precisely in the ribosome P site, ready to pair with its cognate tRNA and initiate translation. In multigene operons, each gene has its own Shine-Dalgarno sequence upstream of its start codon; while the genes share a common mRNA transcript, their open reading frames on the transcript are translated independently (see Fig. 8.1).

In archaea and in some lineages of bacteria, the use of Shine-Dalgarno sequences to initiate translation is much less common. In some cases, the mRNAs completely lack a 5′ untranslated region: the mRNA sequence begins immediately with the start codon. Translation of genes without Shine-Dalgarno sequences is initiated by one of several possible mechanisms that are still under investigation.

Note: Eukaryotic ribosomes have a different mechanism for finding the start codon. The 5′ end of eukaryotic mRNA molecules is modified posttranscriptionally with a 7-methylguanosine 5′ cap. Ribosomes scan downstream of this 5′ cap for an AUG located within a short specific sequence context called the Kozak sequence, named after its discoverer, Marilyn Kozak.

The Three Stages of Protein Synthesis

Once the ribosome has been properly positioned on the message, polypeptide synthesis can begin. The ribosome moves along the mRNA in a 5′-to-3′ direction, translating each codon of the message into an amino acid that it adds to the growing polypeptide chain. Like transcription, polypeptide synthesis has three stages: initiation, which brings the two ribosomal subunits together, placing the first amino acid in position; elongation, which sequentially adds amino acids as directed by the mRNA transcript; and termination. Several steps in this process can be inhibited by certain antibiotics (described later in this section).

As the ribosome moves forward, each incoming tRNA molecule successively occupies one of three binding sites (**Fig. 8.19**). The first position is the aminoacyl-tRNA acceptor site (A site) , where the incoming aminoacyl-tRNA enters the complex and its anticodon binds the codon of the mRNA. The second position is the peptidyl-tRNA site (P site) , which binds the tRNA that is attached to the growing polypeptide. The third position is the exit site (E site) , where the tRNA will be jettisoned from the ribosome after giving up its amino acid to the polypeptide chain. Next we discuss how these binding sites are used in the initiation, elongation, and termination of translation.

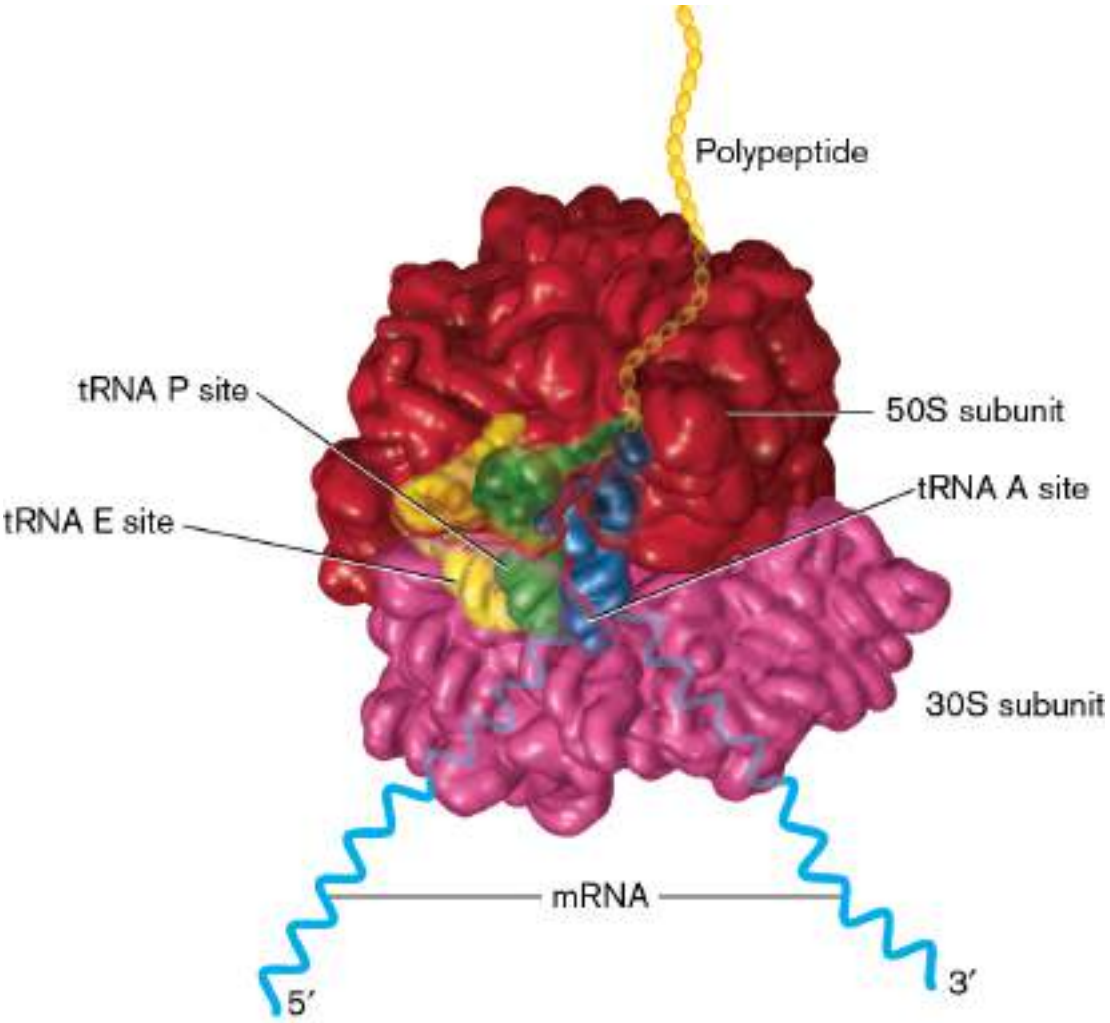


FIGURE 8.19 ■ Binding of tRNA. X-ray-crystallographic model of *Thermus thermophilus* ribosome with associated tRNAs. (The 50S subunit is red, 30S is pink, and tRNAs in the A, P, and E sites are blue, green, and yellow, respectively.) The 30S subunit of the ribosome travels along the mRNA (light blue) in a 5′-to-3′ direction, and the growing polypeptide (yellow) exits from a channel formed in the 50S subunit. (PDB codes: 1GIX and 1GIY)

Initiation of translation. In bacteria such as *E. coli* , initiation of protein synthesis requires three small proteins called initiation factors (IF1, IF2, and IF3).

Initially, IF3 binds to the 30S subunit to separate the 30S and 50S subunits. Separating the subunits allows other initiation factors and mRNA access to the 30S subunit (**Fig. 8.20** , step 1). IF1 and mRNA join the 30S subunit. IF1 blocks the A site, and the mRNA sequence for the ribosome-binding site aligns with its complementary sequence on 16S rRNA (step 2).

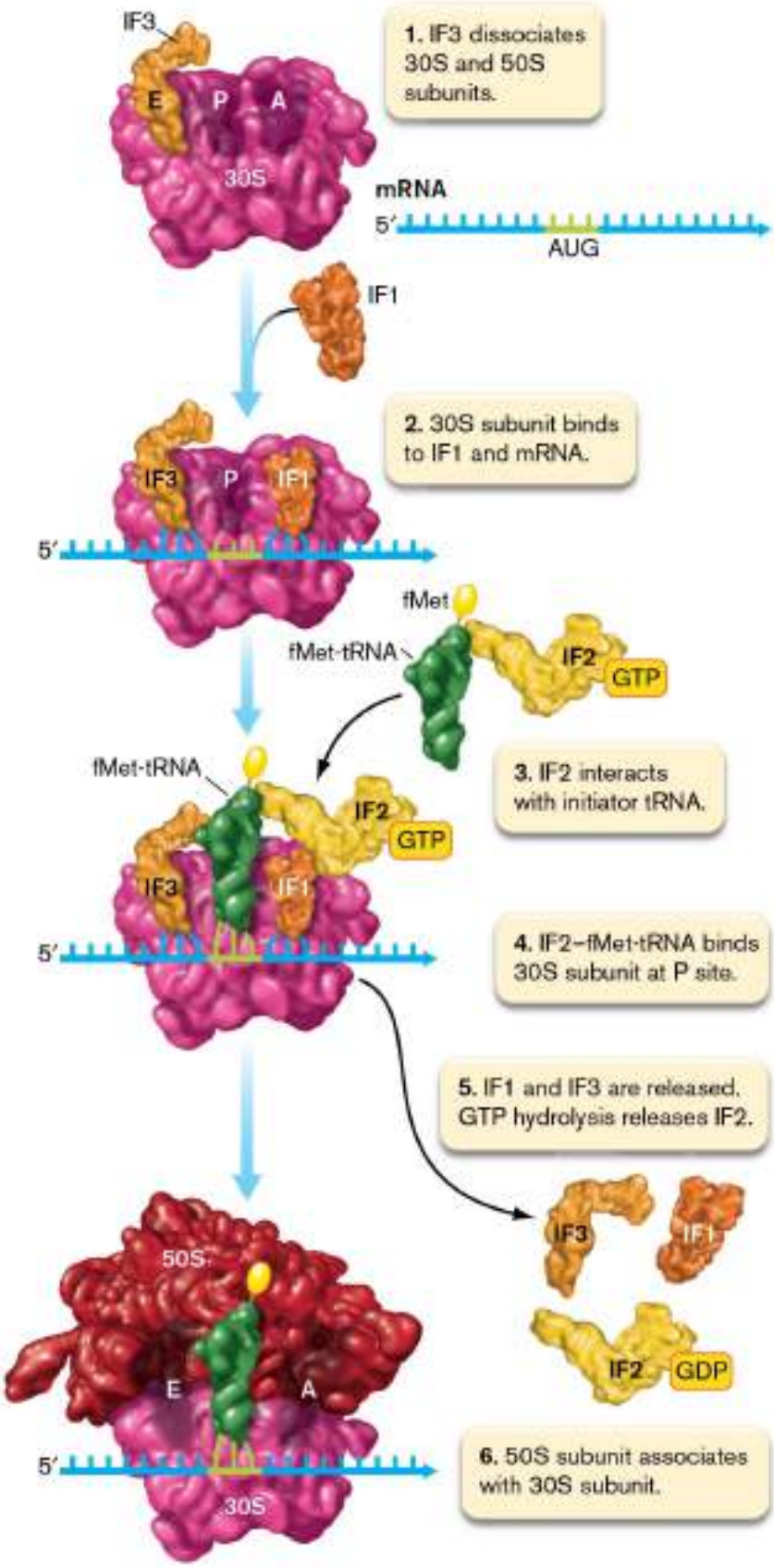

IF3
E
P
A
30S
1. IF3 dissociates 30S and 50S subunits.
mRNA
5′
AUG
IF1
2. 30S subunit binds to IF1 and mRNA.
IF3
P
IF1
5′
fMet
fMet-tRNA
IF2
GTP
fMet-tRNA
IF2
GTP
IF3
IF1
3. IF2 interacts with initiator tRNA.
5′
4. IF2–fMet-tRNA binds 30S subunit at P site.
5. IF1 and IF3 are released. GTP hydrolysis releases IF2.
IF3
IF1
IF2
GDP
50S
E
A
5′
30S
6. 50S subunit associates with 30S subunit.

FIGURE 8.20 ■ Translation initiation. The end result of translation initiation is assembly of the 50S-30S-mRNA complex with the initiator fMet-tRNA set in the P site. See text for details.

See above for *Protein Synthesis* animation

Next, IF2 complexed to guanosine triphosphate (GTP) escorts the initiator *N*-formylmethionyl-tRNA (fMet-tRNA) to the start codon located at what will be the P site (**Fig. 8.20** , steps 3 and 4). *N*-formylmethionyl-tRNA binds to all start codons and is the only aminoacyl-tRNA to bind directly to the P site. Once the initiator tRNA is in place, GTP hydrolysis releases IF2, along with IF1 and IF3 (step 5).

The 50S subunit then docks to the 30S subunit (step 6). The ribosome is now ready for elongation. Note that initiation is much more complex in archaea, involving as many as six different IF proteins.

Elongation of the polypeptide. In elongation, three basic steps are repeated: (1) A charged tRNA enters the A (acceptor) site; (2) the amino acid brought to the A site is covalently attached to the tRNA-bound amino acid in the P site, adding to the polypeptide chain; (3) the ribosome moves forward by one codon in a 5′-to-3′ direction (**Fig. 8.21**). First an elongation factor (EF-Tu) associates with GTP to form a complex (EF-Tu-GTP) that binds to most charged aminoacyl-tRNAs (except for the initiator tRNA). Sequences within the 50S rRNA recognize features of the EF-Tu-GTP–aminoacyl-tRNA complex, guiding it into the A site (**Fig. 8.21** , step 1). Correct selection of a tRNA is guided in large part by codon-anticodon pairing, but conformational changes sensed by various ribosomal proteins customize the fit. If the fit is not perfect, the tRNA is rejected.

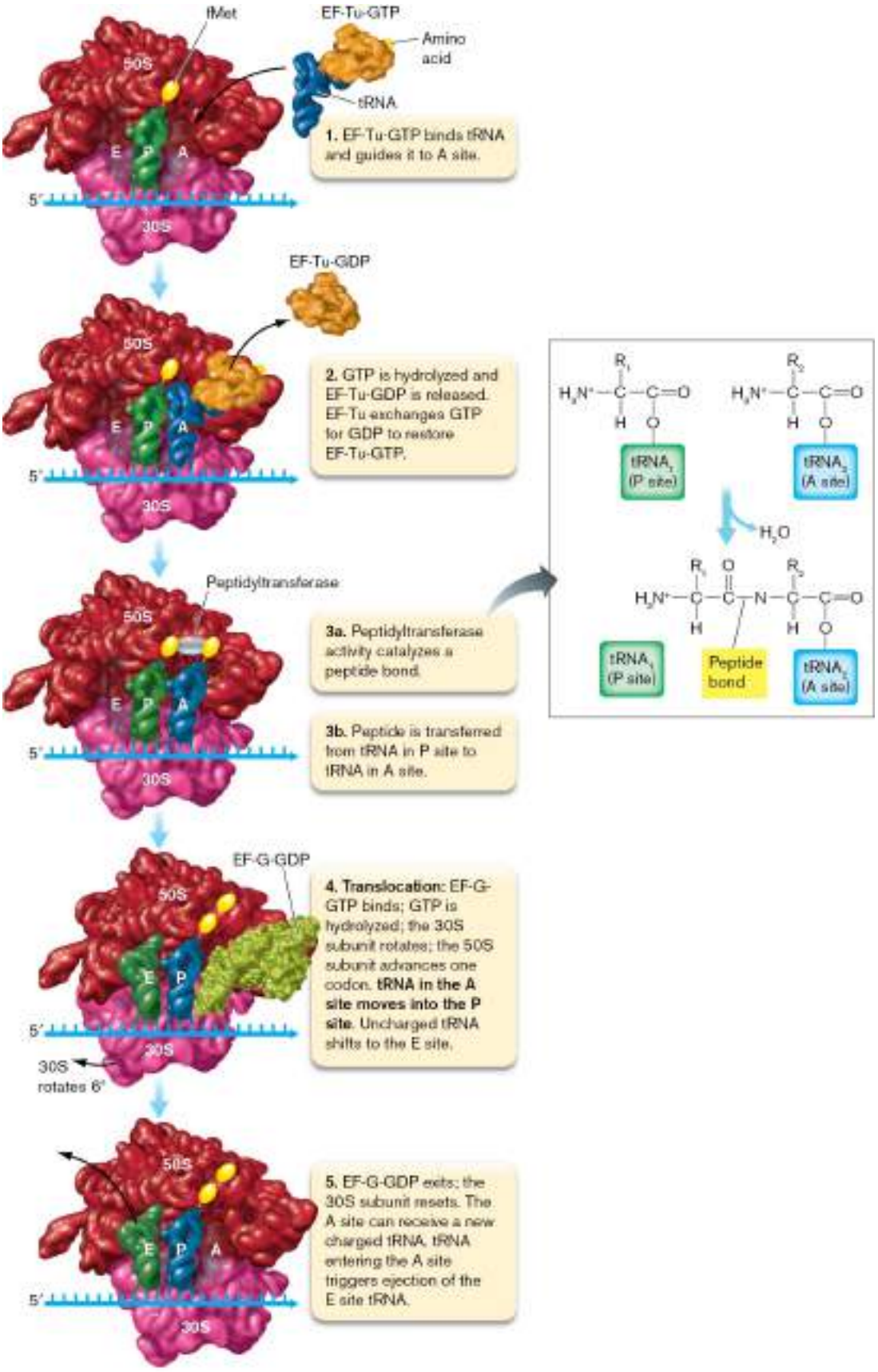

fMet
EF-Tu-GTP
Amino acid
tRNA
50S
E
P
A
5′
3′
30S
1. EF-Tu-GTP binds tRNA and guides it to A site.
EF-Tu-GDP
2. GTP is hydrolyzed and EF-Tu-GDP is released. EF-Tu exchanges GTP for GDP to restore EF-Tu-GTP.
Peptidyltransferase
3a. Peptidyltransferase activity catalyzes a peptide bond.
3b. Peptide is transferred from tRNA in P site to tRNA in A site.
EF-G-GDP
4. Translocation: EF-G-GTP binds; GTP is hydrolyzed; the 30S subunit rotates; the 50S subunit advances one codon. tRNA in the A site moves into the P site. Uncharged tRNA shifts to the E site.
30S rotates 6°
5. EF-G-GDP exits; the 30S subunit resets. The A site can receive a new charged tRNA. tRNA entering the A site triggers ejection of the E site tRNA.
tRNA1 (P site)
tRNA2 (A site)
H2O
Peptide bond

FIGURE 8.21 ■ Elongation of the peptide. The elongation cycle moves the ribosome forward by one codon, lengthening the peptide chain by attaching it to the amino acid of the incoming tRNA, and ejecting the tRNA from which the peptide chain was transferred. The inset shows the peptidyltransferase reaction that generates the peptide bond.

See above for *Protein Synthesis* animation

Once an aminoacyl-tRNA is in the A site, a peptide bond is formed between the amino acid moored there and the terminal amino acid linked to tRNA in the P site. Simultaneously, a GTP is hydrolyzed, and the resulting EF-Tu-GDP (EF-Tu complexed with guanosine diphosphate) is expelled. Peptide bond formation effectively transfers the polypeptide from the tRNA in the P site to the tRNA in the A site (**Fig. 8.21** , steps 2 and 3).

For protein synthesis to continue, the ribosome must advance by one codon, which moves the peptidyl-tRNA from the A site into the P site, leaving the A site vacant. The process, called translocation , involves another elongation factor, EF-G, associated with GTP. EF-G-GTP binds to the ribosome, GTP is hydrolyzed, and the 30S subunit rotates clockwise to ratchet the 50S subunit ahead on the message by one codon (**Fig. 8.21** , step 4). Translocation is completed by the exit of EF-G-GDP and the rotation of the 30S subunit back to its initial conformation (step 5). The translocation maneuver opens up the A site, moves the peptidyl-tRNA into the P site, and slides uncharged tRNA into the E (exit) site. The next aminoacyl-tRNA that enters the A site stimulates a conformational change in the ribosome that signals through to the E site and ejects the uncharged tRNA.

Notice that EF-Tu and EF-G recycle sequentially on and off the ribosome by binding to the same area of the ribosome. EF-G-GTP and the EF-Tu-GTP–aminoacyl-tRNA complex are structurally similar, an example of molecular mimicry. Because these factors bind to the

same site, they must cycle on and off the ribosome sequentially. Although this process seems incredibly complex, the ribosomes of *E. coli* manage to link together 16 amino acids per second.

Termination of translation. Eventually, the ribosome arrives at the end of the coding region, but not the end of the mRNA. As noted earlier, the end of the coding region is marked by one of three stop codons. Translocation after the last peptide bond forms brings the mRNA stop codon into the A site (**Fig. 8.22** , step 1). No tRNA binds, but one of two release factors (RF1 or RF2) will enter (also step 1). Binding of release factor leads to ejection of tRNA in the E site and activates the peptidyltransferase, thereby cutting the bond that tethers the completed polypeptide to tRNA in the P site (step 2).

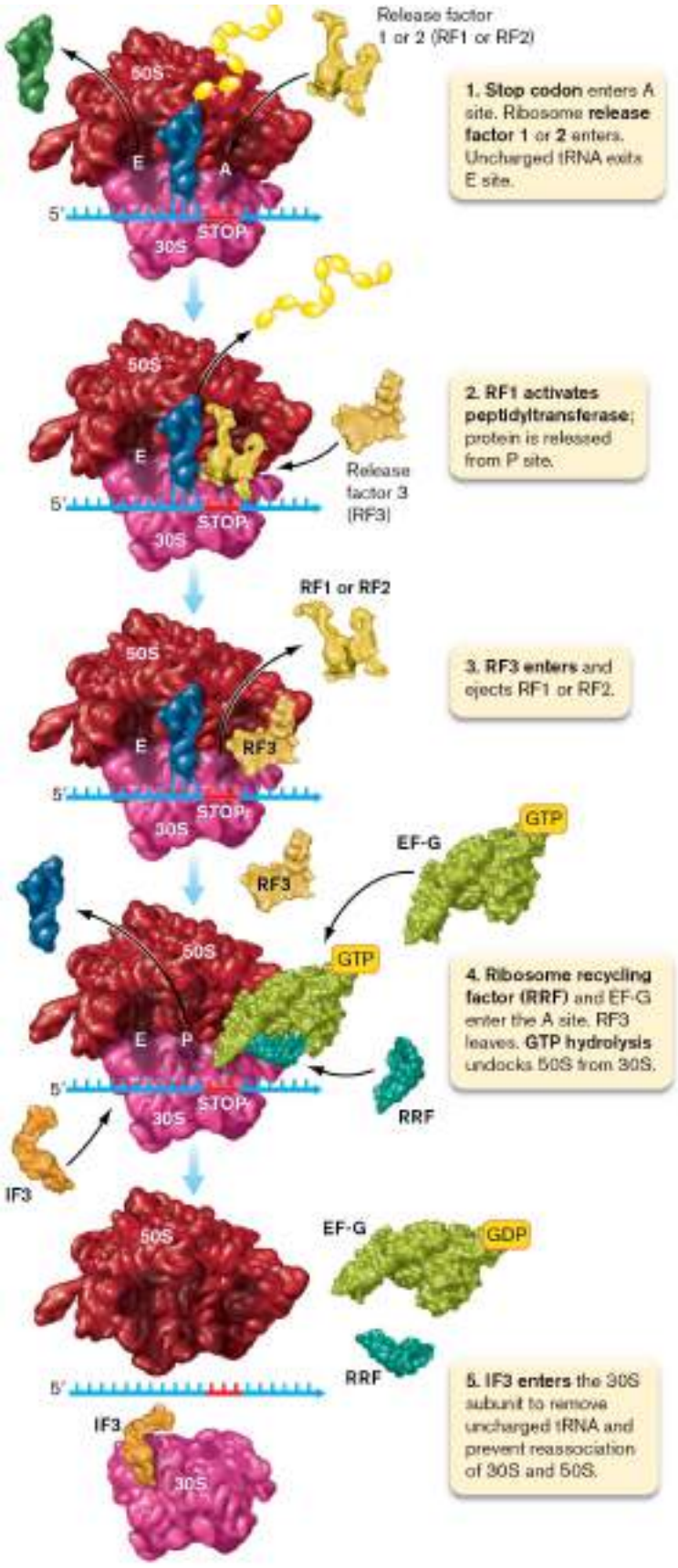

Release factor
1 or 2 (RF1 or RF2)
50S
E
A
5'
STOP
30S
1. Stop codon enters A site. Ribosome release factor 1 or 2 enters. Uncharged tRNA exits E site.
2. RF1 activates peptidyltransferase; protein is released from P site.
Release factor 3 (RF3)
RF1 or RF2
3. RF3 enters and ejects RF1 or RF2.
RF3
EF-G
GTP
4. Ribosome recycling factor (RRF) and EF-G enter the A site. RF3 leaves. GTP hydrolysis undocks 50S from 30S.
P
RRF
IF3
GDP
5. IF3 enters the 30S subunit to remove uncharged tRNA and prevent reassociation of 30S and 50S.

FIGURE 8.22 ■ Termination of translation. The completed protein is released, and the ribosomal subunits are recycled.

See above for *Protein Synthesis* animation

With the protein released, the ribosome disassembles. RF3 causes RF1 or RF2 to depart the ribosome (**Fig. 8.22** , step 3). Then, ribosome recycling factor (RRF), along with EF-G, binds at the A site, and accompanying hydrolysis of GTP undocks the two ribosomal subunits (step 4). IF3 then reenters the 30S subunit to eject the remaining uncharged tRNA and mRNA (step 5), thereby preventing the 30S and 50S subunits from redocking. The liberated ribosomal subunits are now free to diffuse through the cell, ready to bind yet another mRNA and begin the translation sequence anew.

Sometimes, premature transcription termination or cleavage of the mRNA leaves an mRNA without a translation stop codon. Lacking the signal to release the mRNA, the ribosome becomes temporarily "stuck" on these truncated mRNAs. These stuck ribosomes are released by the action of tmRNAs, which also mark for degradation the defective proteins synthesized from the shortened mRNAs.

Thought Questions

8.7 How might one gene code for two proteins with different amino acid sequences?

8.8 Why involve RNA in protein synthesis? Why not translate directly from DNA?

8.9 Codon 45 of a 90-codon gene was changed into a translation stop codon, producing a shortened (truncated) protein. What kind of mutant could produce a full-length protein from the gene without removing the stop codon? *Hint:* What molecule recognizes a codon?

Antibiotics That Affect Translation

Most of our antibiotics were discovered as natural products of fungi or of bacteria such as actinomycetes (presented in Chapter 18). Streptomycin (**Fig. 8.23A**), a well-known member of the aminoglycoside family of antibiotics, is produced by the actinomycete *Streptomyces* . The drug targets bacterial small ribosomal subunits by binding to a region of 16S rRNA in the 30S subunit that forms part of the decoding A site and to protein S12, a protein critical for maintaining the specificity of codon-anticodon binding. Streptomycin bound to 16S rRNA makes decoding of mRNA at the A site "sloppy" by permitting illicit codon-anticodon matchups that result in a mistranslated protein sequence.

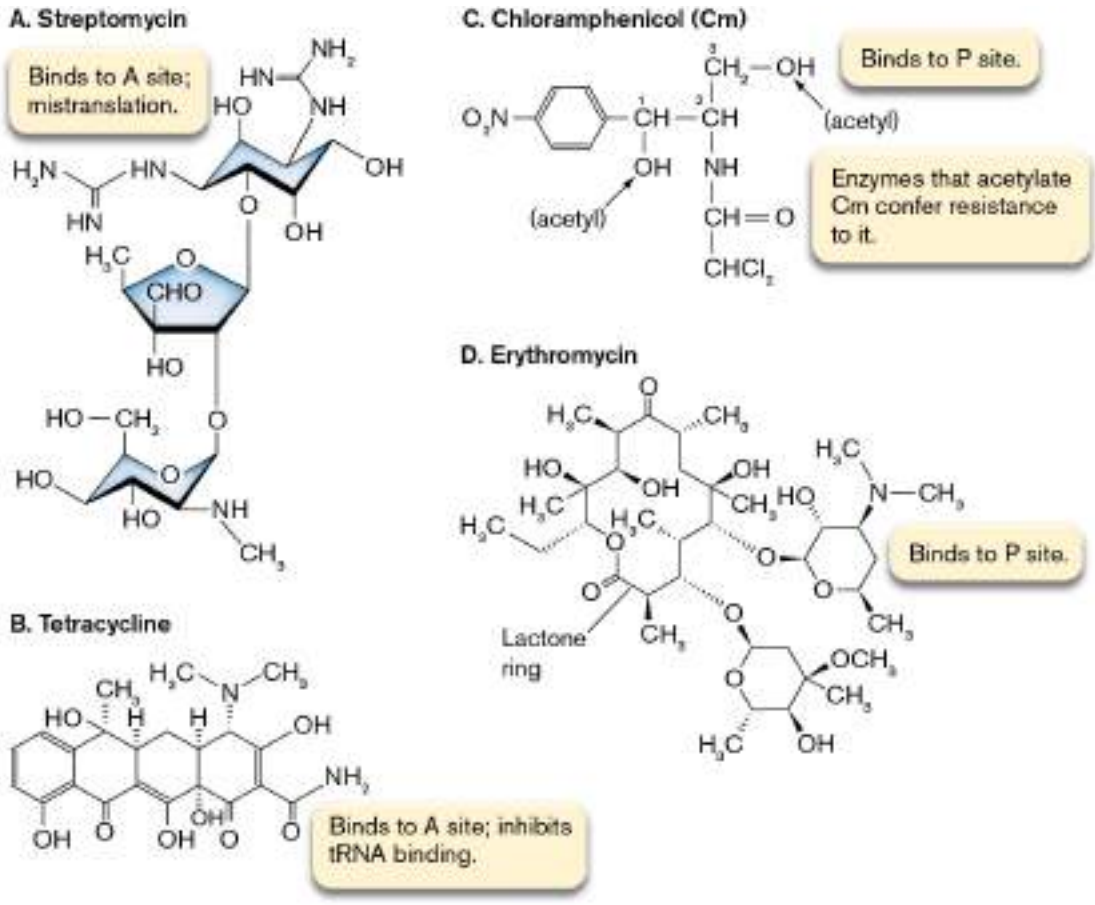


FIGURE 8.23 ■ Antibiotics that inhibit protein synthesis in bacteria. Streptomycin **(A)** and tetracycline **(B)** bind to the A site. Streptomycin causes mistranslation; tetracycline inhibits tRNA binding. Chloramphenicol **(C)** and erythromycin **(D)** bind to the peptidyl-tRNA site, thereby inhibiting peptide bond formation.

Bacteria can become resistant to streptomycin via spontaneous mutations in the S12 gene (*rpsL*) or 16S rRNA. The altered ribosomal protein or rRNA remains functional but does not bind streptomycin. Some bacteria gain resistance by acquiring an aminoglycoside phosphotransferase that modifies streptomycin so that it cannot bind to its target. Additional mechanisms are discussed in Chapter 27 . Other therapeutically important aminoglycosides include gentamicin, tobramycin, and amikacin.

Tetracyclines are polyketides, a class of antibiotics whose biosynthesis is presented in Chapter 15 . Tetracycline and derivatives such as doxycycline (**Fig. 8.23B**) also target the 30S subunit, where they bind to 16S rRNA near the A site. But instead of causing mistranslation, the tetracyclines prevent aminoacyl-tRNA from binding to the A site. Resistance to tetracycline can be conferred by an efflux transport system that effectively removes the antibiotic from the bacterial cell. Other resistance mechanisms are described in Chapter 27 .

Another species of *Streptomyces* produces chloramphenicol (**Fig. 8.23C**), which attacks the 50S subunit. It binds to 23S rRNA at the peptidyl-tRNA site (that is, the P site) and inhibits peptide bond formation. Resistance to this drug comes from an ability to synthesize the enzyme chloramphenicol acetyltransferase, which modifies chloramphenicol in a way that destroys its activity.

Erythromycin, made by *Streptomyces erythraeus* , is one of a large group of related antibiotics called macrolides, whose hallmark is a large lactone ring of 12−22 carbon atoms (**Fig. 8.23D**). Macrolides attack the 50S subunit by binding to 23S rRNA in the nascent peptide exit tunnel near the P site. Binding alters peptidyltransferase structure and interferes with peptide bond

formation. Resistance to macrolides usually involves an efflux pump or methylation of the relevant area of 23S rRNA. A recently discovered resistance mechanism uses a protein to knock macrolides off of the ribosome.

Other translation-targeting antibiotics interfere with mRNA binding to the ribosome (kasugamycin), prevent translocation by targeting EF-G (fusidic acid), or use structural similarity to tRNA (molecular mimicry) to trick peptidyltransferase into action without having a bona fide tRNA in the A site (puromycin). We chronicle the discovery and use of antibiotics more completely in Chapter 27 .

Thought Question

8.10 While working as a member of a pharmaceutical company's drug discovery team, you find that a soil microbe snatched from the jungles of South America produces an antibiotic that will kill even the most deadly, drug-resistant form of *Enterococcus faecalis* , which causes bacterial endocarditis. Your experiments indicate that the compound stops protein synthesis. How could you more precisely determine the antibiotic's mode of action? *Hint:* Can you use mutants resistant to the antibiotic?

Polysomes and Coupled Transcription-Translation

Polysomes. Once a ribosome begins translating mRNA and moves beyond the ribosome-binding site, another ribosome can immediately jump onto that site. The result is an mRNA molecule with multiple ribosomes moving along its length at once. The multiribosome structure is known as a polysome (**Fig. 8.24**). Ribosomes in a polysome are closely packed along the mRNA, which helps protect the message from degradation by RNases and enables

the speedy production of multiple copies of the protein from a single mRNA molecule.

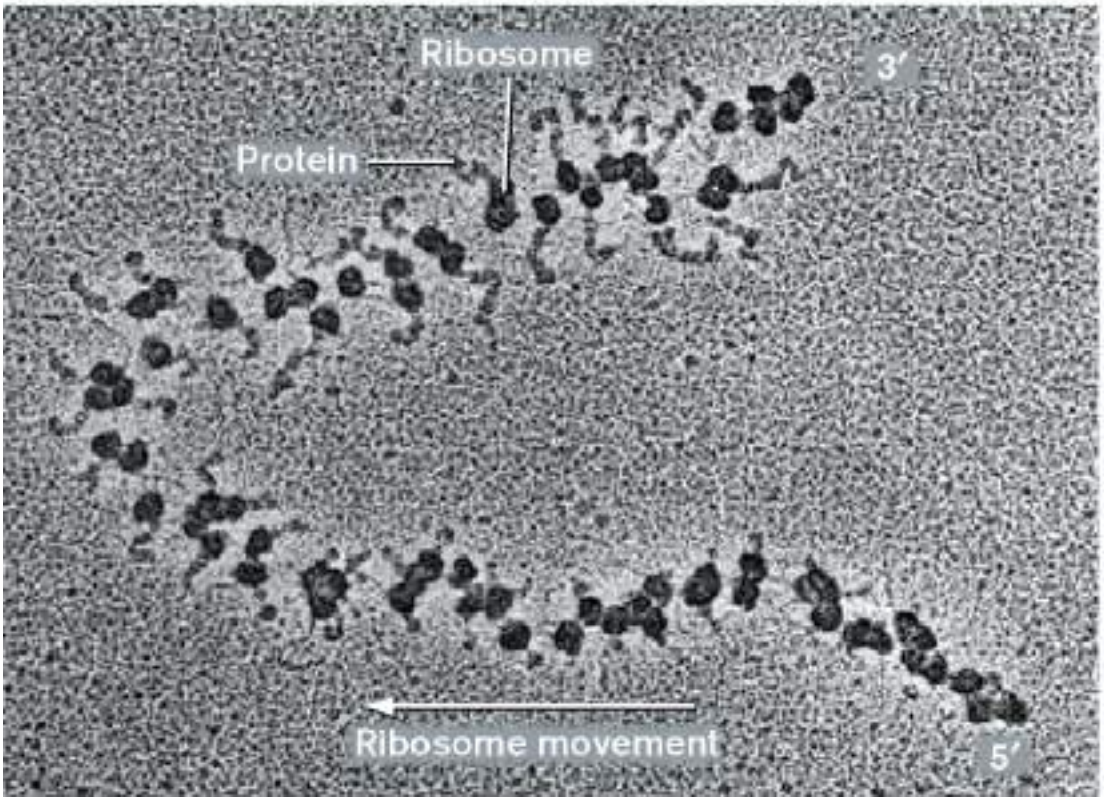


FIGURE 8.24 ■ A polysome from a eukaryotic cell. Several ribosomes may translate a single mRNA molecule at the same time. The beginning (5′ end) of the mRNA is at lower right; the 3′ end is at upper right. Note that the synthesized protein molecule grows longer and longer as the ribosome approaches the 3′ end of the mRNA, where the protein molecule is most clearly seen. Polysomes also occur in prokaryotes.

O. L. MILLER, JR. 1982. *EMBO J.* **1** :59

Coupled transcription and translation. Because bacteria and archaea lack nuclear membranes, ribosomal subunits floating through the cytoplasm have an opportunity to bind to the 5′ end of mRNA and begin making protein even before RNA polymerase has finished making the full-length mRNA molecule. The simultaneous

building of both mRNA and proteins is called coupled transcription and translation (**Fig. 8.25**). During coupled transcription and translation, the lead ribosome can "catch up" to the RNA polymerase and establish a physical connection between the two enzymes with the help of the NusG bridging protein. These coupled processes can also be coordinated with protein secretion across the membrane (see Figure 3.27 for details). The coupling of transcription and translation in bacteria makes it possible for the cell to use translation as a means of regulating transcription. One such regulatory process, attenuation, is explained further in Chapter 10 .

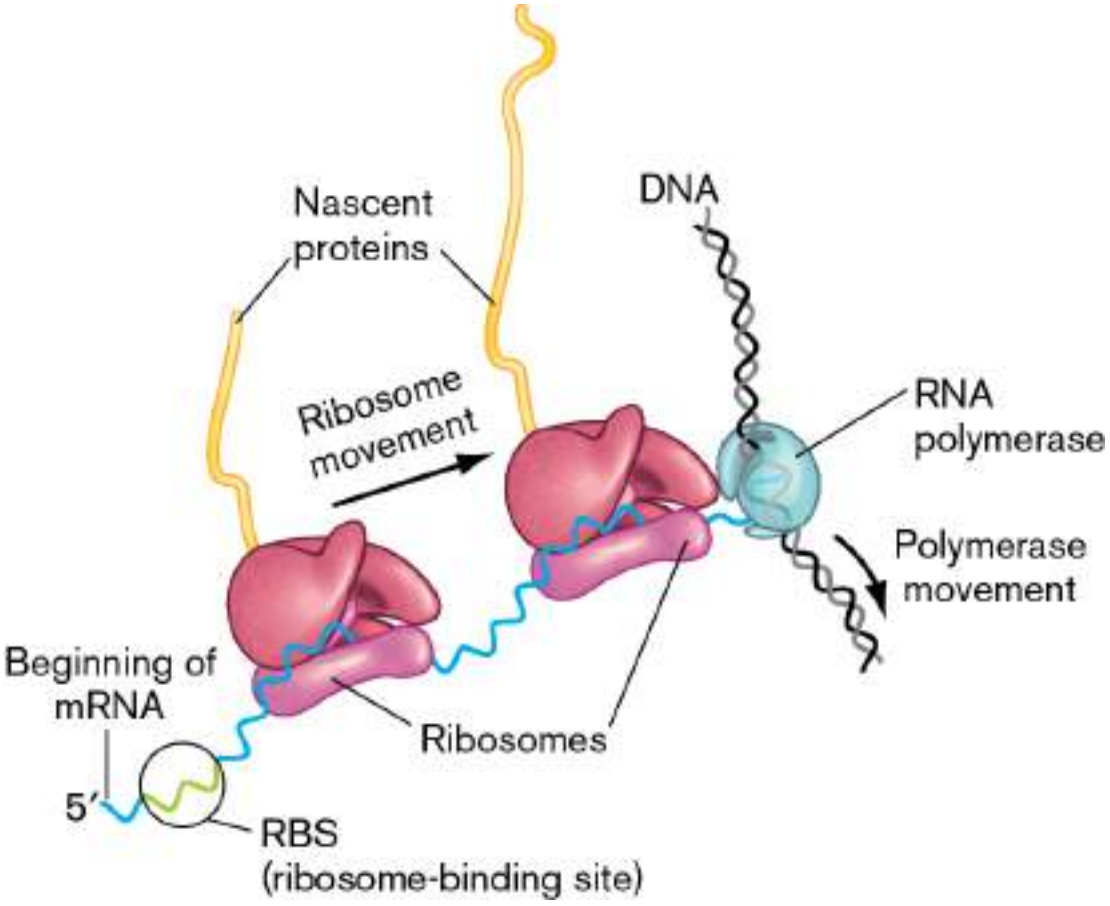


FIGURE 8.25 ■ Coupled transcription and translation in bacteria. During coupled transcription and translation in prokaryotes, ribosomes attach at mRNA ribosome-binding sites and start synthesizing protein before transcription of the gene is complete. Note that the lead ribosome can catch up to and

contact the RNA polymerase through a bridging protein (NusG, not shown).

Note: Recent evidence in the Gram-positive bacterium *Bacillus subtilis* suggests that its RNA polymerase outpaces the ribosome, so the two enzyme complexes do not form physical connections as they do in *E. coli* . This has consequences for gene regulation: Unlike *E. coli* , *Bacillus* generally lacks transcription attenuators under ribosome control (see Section 10.3).

Once transcription is complete, the mature mRNA can diffuse out of the nucleoid. In rod-shaped cells, most ribosomes are located at the poles, being generally excluded from the nucleoid (**Fig. 8.26**), and the poles are where most translation occurs over the lifetime of the mRNA. Thus, although transcription-translation coupling happens when transcripts are first made in the nucleoid (**Fig. 8.27**), most translation occurs independent of transcription in nucleoid-free regions.

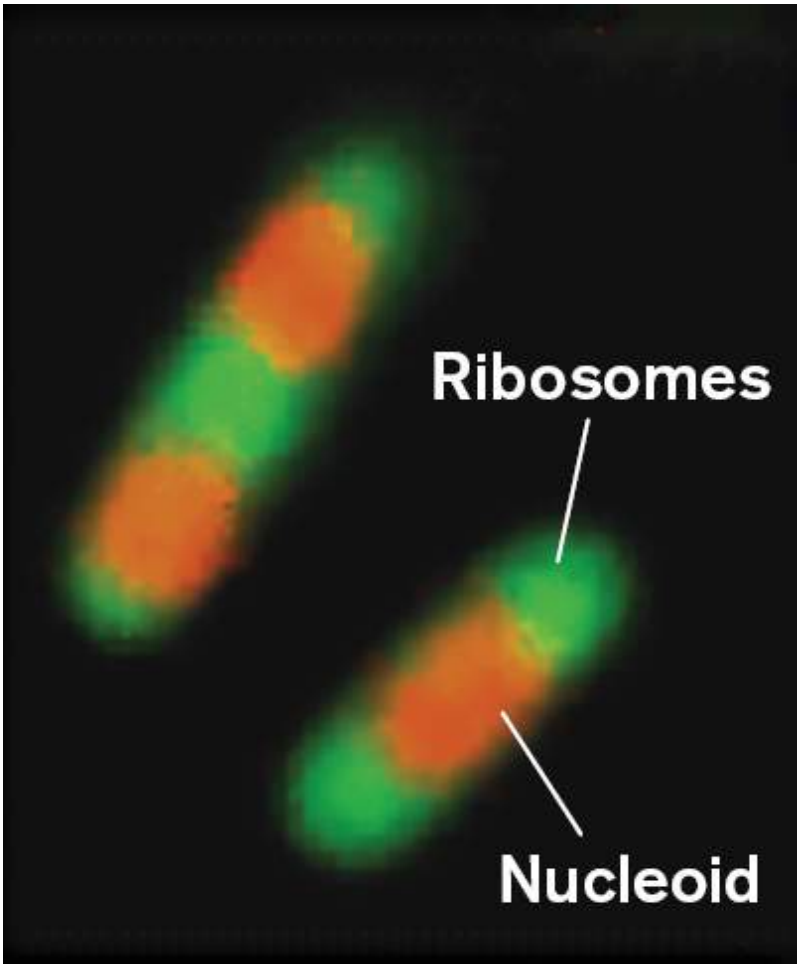


FIGURE 8.26 ■ Most translation in *E. coli* is not coupled to transcription. Most ribosomes exist in nucleoid-free areas of the cell, primarily at the cell poles. DNA was visualized using a red fluorescent dye. Ribosomes contained a protein (S2) fused to yellow fluorescent protein, which here appears green.

S. BAKSHI ET AL. 2012. *MOL. MICROBIOL* . **85** :21

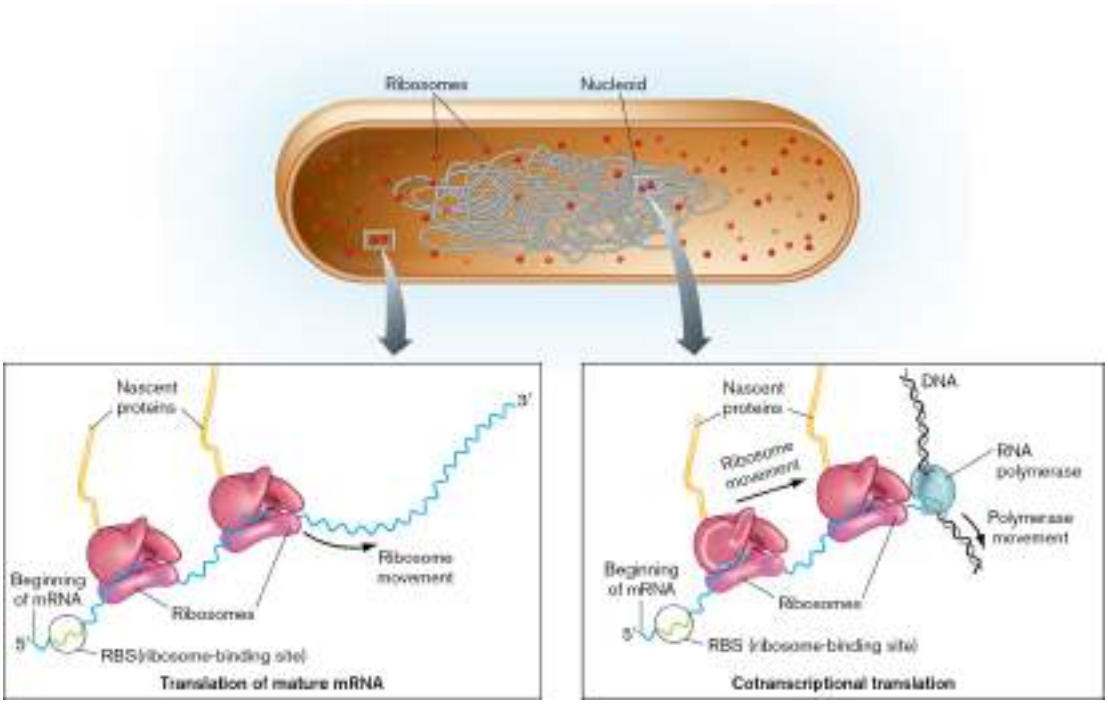


FIGURE 8.27 ■ Spatial localization of transcription and translation in *E. coli* . Coupled transcription and translation occurs in or near the nucleoid (right blowup), whereas translation of mature, fully transcribed mRNA occurs at the cell poles (left blowup).

Transcription in eukaryotic microbes is not coupled to translation. In contrast to bacteria and archaea, eukaryotic microbes use distinct cell compartments to carry out most of their transcription and translation. They transcribe genes in the nucleus, where internal, noncoding parts of the mRNA (introns) are removed by a process known as RNA splicing. The processed transcripts are then exported to the cytoplasm and translated. Most eukaryotic DNA viruses do the same.

Some Microbes Have Modified Genetic Codes

The standard genetic code (see **Fig. 8.12**) contains 61 sense codons and 3 nonsense (stop) codons. The code is ancient and was

likely established before the three domains of life emerged from the last universal common ancestor. A question worth asking is: Once the code is set, can it be changed? For example, if the codon UAU is reassigned from tyrosine to cysteine, then whenever the UAU codon appears in a transcript, tyrosine will be replaced with cysteine. Such a change could alter the sequence of hundreds or thousands of proteins at the same time, which could be lethal for an organism. Not surprisingly, then, codon reassignments are not common in nature, and when found, they are usually limited to one or a few codon reassignments per organism.

How are codons reassigned? The usual way is to change the anticodon sequence in a tRNA so that it carries the same amino acid but now recognizes a different codon. In some cases, the codon was recognized as a stop codon, but after the change, the ribosome can "read through" the translation stop signal. For SR1, a lineage of bacteria associated with human periodontal disease, UGA has changed from a stop codon to one that encodes glycine. In the ciliate *Condylostoma magnum* , all three stop codons of the standard code (UGA, UAA, and UAG) encode amino acids. How does translation stop in this organism? The stop codons actually serve dual functions in this eukaryotic microbe: They encode amino acids when located at internal positions within the mRNA transcript, and they serve as stop codons when found at the 3′ end of the transcript. Just how the ribosome determines correctly which way to read these codons on the basis of location in the transcript is still unknown. In cases such as these, in which stop codons have been reassigned to sense codons, the adaptive benefit is not known, and it may be that they are simply tolerated evolutionary "accidents" that do no real help or harm to the organism (see example in the Chapter 18 eResearch Activity).

In contrast, two other codon reassignments have more obvious benefit because they expand the genetic code and introduce novel amino acids—pyrrolysine and selenocysteine—into proteins. Pyrrolysine and selenocysteine have unique side chains that are related to lysine and cysteine, respectively (**Fig. 8.28**), and when incorporated into the polypeptide chain they confer new biochemical

properties on the protein. Cells that utilize pyrrolysine and selenocysteine require special enzymes for their synthesis and ligation to their own dedicated tRNAs.

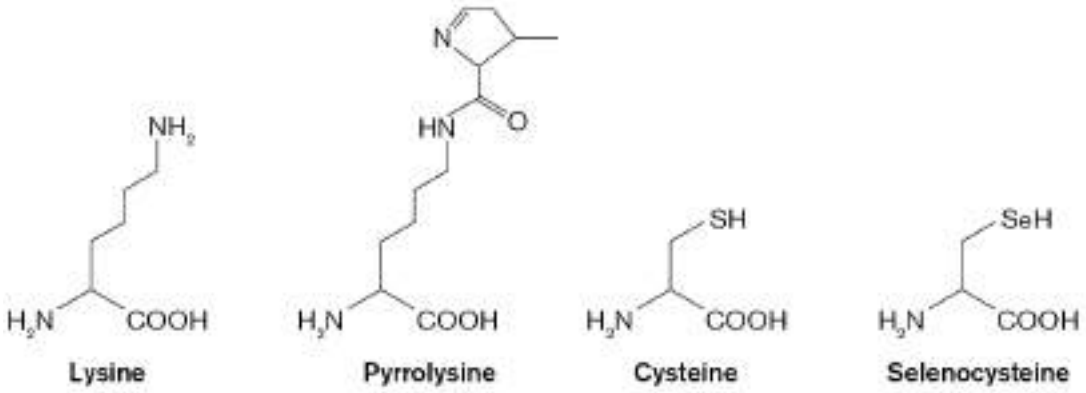


FIGURE 8.28 ■ The 21st and 22nd amino acids for translation. The structures of pyrrolysine and selenocysteine, compared to lysine and cysteine.

Pyrrolysine has been found in the proteins of some bacteria and archaea, but thus far not in eukaryotes. The marine bacterium *Acetohalobium arabaticum* is the only known organism that can expand its genetic code in response to environmental cues. *A. arabaticum* can grow on several carbon sources, including pyruvate and the methyl-containing molecule trimethylamine. When growing on pyruvate, it uses the canonical 20 amino acids to make proteins. However, growth on trimethylamine requires a pyrrolysine-containing enzyme. In response to this growth condition, *A. arabaticum* synthesizes the machinery to make pyrrolysine and ligate it onto its tRNA for incorporation into protein.

Selenocysteine has been found in proteins from all three domains of life. There are over 50 bacterial protein families that contain selenocysteine, many of which are enzymes involved in oxidation-reduction (redox) chemistry and protection from oxidative stress. In "selenoproteins" involved in redox chemistry, the selenocysteines have usually replaced cysteines at

critical positions in the protein, and this change has been demonstrated to improve catalytic efficiency of the enzymes.

What if some codons are not reassigned, but rather are eliminated from the organism entirely? Could such a modified organism resist viruses that still use all 64 codons? This question is addressed in the synthetic biology study described in Special Topic 8 .

SPECIAL TOPIC 8 Rewriting the Genetic Code to Arrest Viral Infection

Viruses exploit the gene expression machinery of their hosts to reproduce during infection (see Chapters 6 and 11). In all known cases of viral infection, this process involves tricking the ribosome into translating viral mRNA into protein. Viruses also use the host's tRNAs at the sense codons of the viral mRNA and the host's release factors at the nonsense codons of the viral mRNA. What if the host could stop providing all these resources to the virus because the host had been engineered to no longer need them?

This question was answered by Jason Chin and colleagues at the Medical Research Council Laboratory in Cambridge, UK. These researchers reasoned that if they could remove certain tRNAs and one release factor from the host, incoming phages that rely on these for translation would not be able to make their full-length proteins needed to complete replication and lysis of the host cell.

In order to delete those tRNAs and the release factor, the researchers exploited the redundant nature of the genetic code to rewrite the host genome. As indicated in Figure 8.12 , synonymous codons can encode the same amino acid. If all copies of one synonym in the genome are replaced with a

different synonym, then the tRNA with the cognate anticodon for that first synonym will no longer be used for translation. Importantly, this change in DNA has no effect on the protein synthesized: The correct amino acid is still incorporated at this codon's position in the reading frame. Likewise, if all copies of the TAG-coded stop codon that binds release factor 1 are replaced by TGA- or TAA-coded stop codons that bind release factor 2, then release factor 1 is no longer necessary for translation termination.

The researchers followed this strategy to rewrite the genome of *Escherichia coli* and used synthetic biology techniques for the endeavor. Every TCG- and TCA-coded codon was replaced with a synonymous codon encoding serine (**Fig. ST 8.1**). The *serT* and *serU* genes encoding the tRNAs that recognize the TCG- or TCA-coded codons were deleted from this recoded genome. Likewise, the TAG-coded stop codons were replaced by TGA or TAA, and the *prfA* gene encoding release factor 1 was deleted. This recoded version of *E. coli* , named Syn61Δ3, used only 61 of the 64 possible codons yet was still able to faithfully translate its mRNAs and grow in culture.

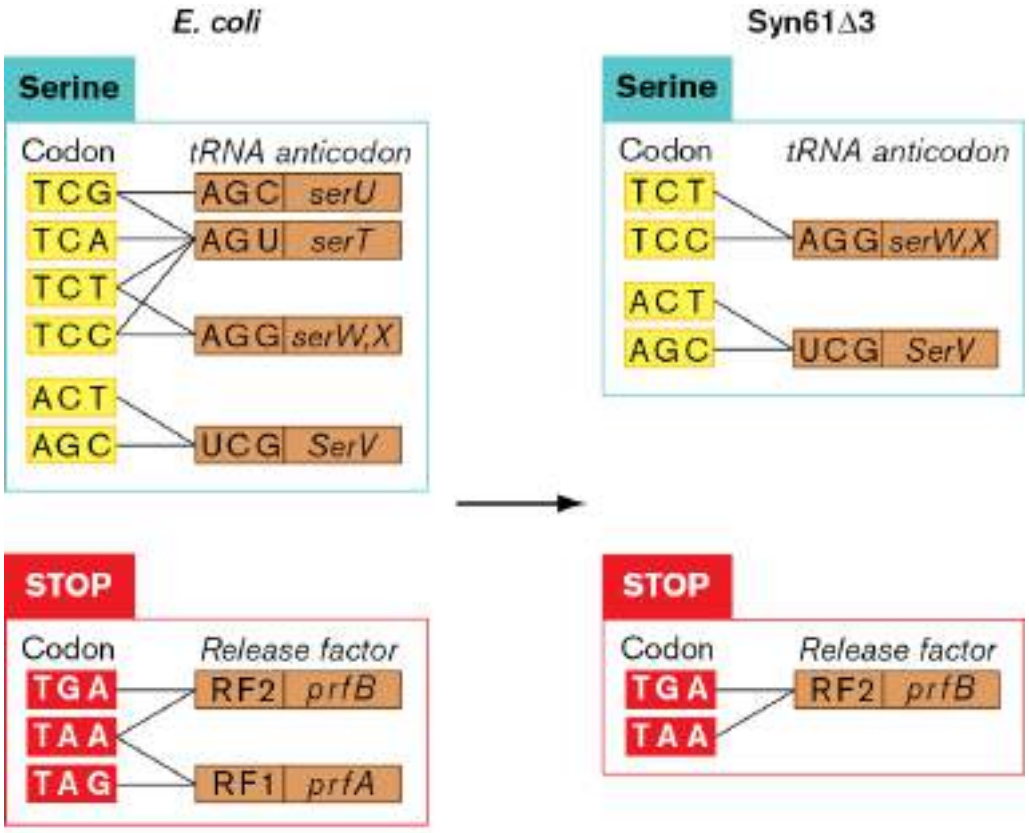


FIGURE ST 8.1 ■ Recoding the *E. coli* genome and removing the superfluous tRNAs and release factor. Two serine codons and one stop codon were recoded to synonymous serine or stop codons, and the tRNAs or release factor 1 that recognize these removed codons were deleted from the genome, resulting in strain Syn61Δ3. Note that some codons are recognized by multiple tRNAs or release factors. Also note that for ease of comparison with the codon sequences (shown 5′ to 3′), the anticodons are shown in the 3′-to-5′ direction (see Figure 8.14 for how base pairing between codons and anticodons operates).

The next step was to challenge the Syn61Δ3 cells with phages that infect and lyse wild-type *E. coli* . The researchers hypothesized that phage with TCG-, TCA-, and/or TAG-coded codons in their genome would be unable to synthesize the proteins essential for reproduction, including those that form

the capsid shell. The ribosome would stall at these codons, unable to add a serine to the growing polypeptide or, at the absent stop codon, release from the mRNA at the appropriate location (**Fig. ST 8.2**). The Syn61Δ3 cells were challenged with a cocktail of 5 viruses that lyse *E. coli* : λ, P1 *vir* (a "virulent" mutant that can lyse but not lysogenize the host), T4, T6, and T7. Earlier construction stages of the mutant, called ev2 and ΔRF1, grew to dense cultures after 4 hours of incubation but were completely lysed when the phage cocktail was added (**Fig. ST 8.3**). In contrast, the Syn61Δ3 strain, with all three codons missing, grew to dense cultures whether viruses were present or not. This latter outcome supported the hypothesis that the host is completely resistant to virus infection when it does not synthesize several tRNAs and a release factor.

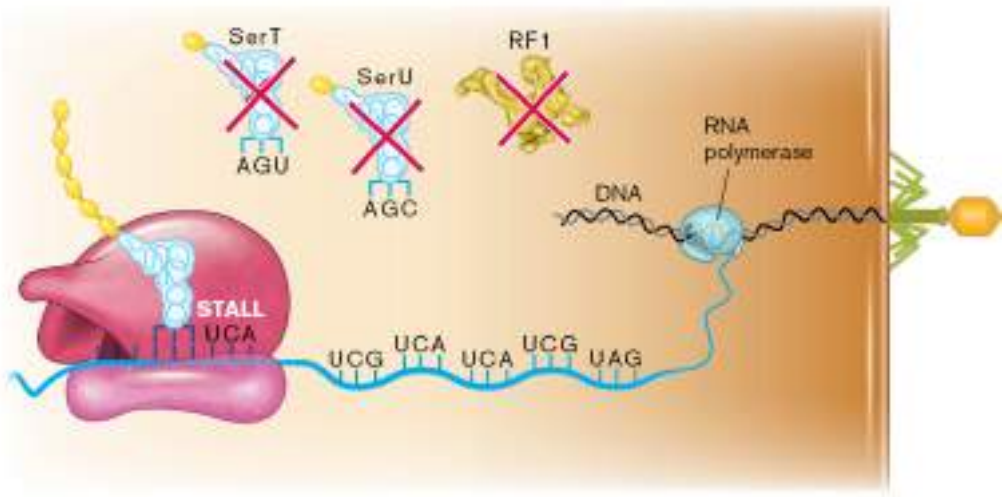


FIGURE ST 8.2 ■ Translation progression and termination stall for phage mRNAs in Syn61 Δ 3. The two serine tRNAs and release factor 1 are missing in the Syn61Δ3 host, so the codons requiring these factors cannot be utilized by the translating ribosome.

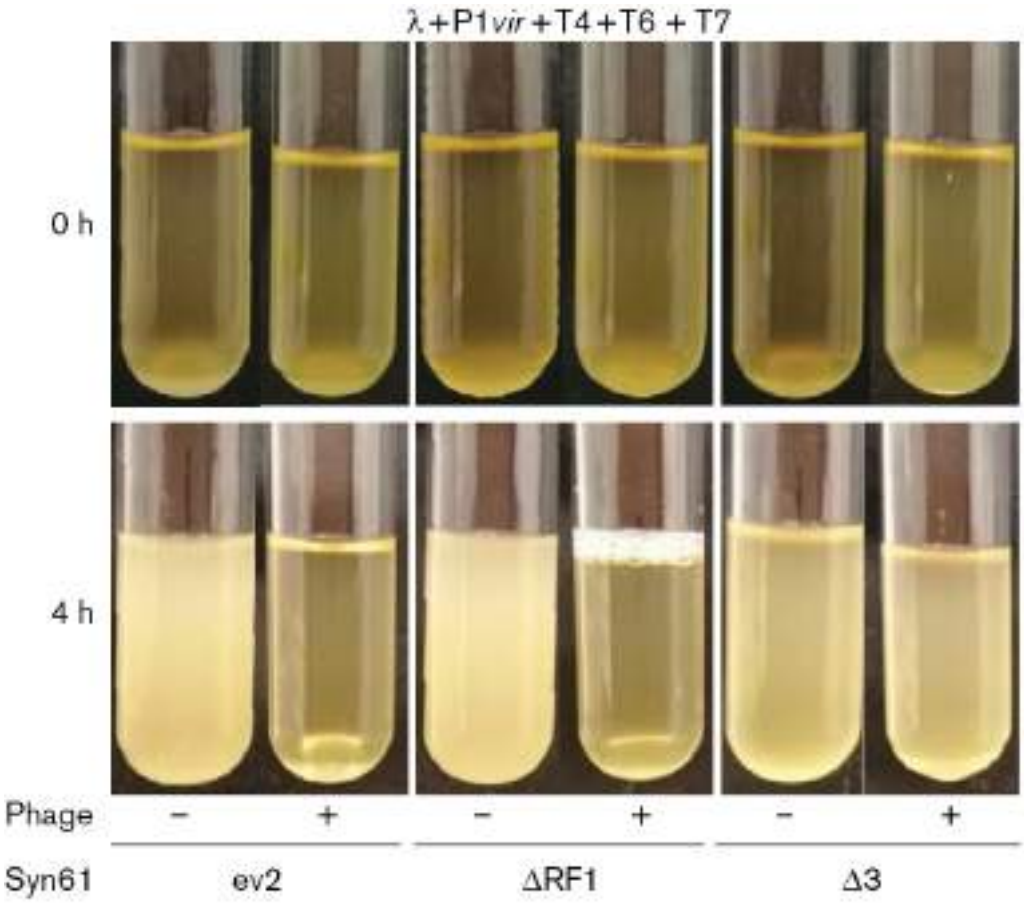


FIGURE ST 8.3 ■ Syn61 Δ 3 is immune to infection by a phage cocktail. Dilute cultures of strain Syn61Δ3 and two earlier (incomplete) stages of Syn61Δ3 synthesis, ev2 and ΔRF1, were challenged by the addition of a cocktail of 5 phages capable of infecting and lysing wild-type *E. coli.* Culture density was observed at 0 and 4 hours after phage addition. Control cultures of the strains did not receive the phages and grew to high density by 4 hours. Outcomes for the cultures treated with phage depended on strain genotype.

Robertson, Wesley E., Louise F. H. Funke, Daniel de la Torre, Julius Fredens, Thomas S. Elliott, et al. 2021. Sense codon reassignment enables viral resistance and encoded polymer synthesis. *Science* **372** :1057–1062.

Why was this resistance so complete, given that viruses, like all life forms, can evolve through mutation to achieve a more fit state? The likely explanation is that they would have to, by mutation, replace most (if not all) of the TCG-, TCA-, and TAG-coded codons in their genome with codons that could be read by the host. The virus genomes simply contained too many of these codons to replace all at once via random mutation.

The researchers followed these experiments with phage by reexpanding the genetic code in a special way. Rather than encode for serine again, the TCG- and TCA-coded codons were used to encode novel, noncanonical amino acids. These artificial amino acids, not used by natural organisms, could be charged to tRNAs whose anticodons recognize TCG- or TCA-coded codons and could become incorporated into protein via translation. This proof-of-concept exercise provides a method for engineers to produce proteins with brand-new properties not seen in nature and in the future could have important industrial or medical applications.

RESEARCH QUESTION

Could you engineer a virus that could infect Syn61Δ3 where the virus still used the TCG-, TCA-, and TAG-coded codons in its genome? *Hint:* Think of what genes the host lacks that could be expressed by the virus, but also consider which codons should be used in those virus copies of the gene(s) that need translation.

Thought Questions

8.11 Why do you think evolution by natural selection favors changes in codons, rather than in anticodons?

8.12 A major way that bacteria acquire new functions is through the acquisition and expression of genes from other microbes via horizontal gene transfer (see Chapters 9 and 17). How would this mechanism of innovation be affected if the recipient bacterium changed its genetic code?

8.13 The incorporation of pyrrolysine and selenocysteine into the genetic code involved stop-to-sense changes, rather than sense-to-sense. Why do you think this was the case?

8.14 While some organisms have alternative genetic codes, they still use codons composed of three bases. Why is it unlikely that organisms will evolve to use a codon composed of four or more bases?

To Summarize

- **Triplet nucleotide codons** in mRNA encode specific amino acids. **Transfer RNA molecules** interpret the genetic code and bring specific amino acids to the A (acceptor) site in the ribosome.
- **Specific codons** mark the beginning and end of a gene. The Shine-Dalgarno sequence in mRNA, located before the start codon, helps the ribosome find the correct reading frame in the mRNA.
- **Initiation of protein synthesis** in bacteria requires three initiation factors that bring the ribosomal subunits together on an mRNA molecule.
- **Peptide bond formation** by the ribosome is carried out by ribosomal RNA, not protein. The polypeptide elongates by one

amino acid when the ribosome ratchets one codon length along the mRNA.
- **Translation terminates** when a stop codon is reached. A release factor enters the A site and triggers peptide bond formation, thus freeing the completed protein from tRNA in the P site.
- **Ribosome recycling factor** and EF-G bind to the A site to dissociate the two ribosomal subunits from the mRNA.
- **Antibiotics that affect translation** can cause ribosomes to misread mRNA (streptomycin), inhibit aminoacyl-tRNA binding to the A site (tetracycline), interfere with peptidyltransferase (chloramphenicol), trigger peptide bond formation prematurely (puromycin), interfere with peptide bond formation (erythromycin), or prevent translocation (fusidic acid).
- **Transcription and translation** are coupled in bacteria and archaea. Translation of mature mRNA can occur after transcription is complete, but new mRNA is usually being translated while it is still being elongated.
- **Changes to the standard genetic code, though uncommon** , can turn stop codons into sense codons and can introduce novel amino acids such as pyrrolysine and selenocysteine into proteins.

Glossary

codon
: A set of three nucleotides that encodes a particular amino acid.

stop codon
: One of three codons (UAA, UAG, UGA) that do not encode an amino acid, and thus trigger the end of translation.

anticodon

A set of three nucleotides in the middle loop of a tRNA that base-pairs with a codon in mRNA.

aminoacyl-tRNA synthetase

An enzyme that condenses a specific amino acid with the 3′ OH group of the correct tRNA, thereby charging the tRNA.

ribosome

A large enzyme, composed of RNA and protein subunits, that translates mRNA into protein.

peptidyltransferase

A ribozyme that catalyzes the formation of peptide bonds.

ribozyme

See **catalytic RNA** .

start codon

A codon (usually AUG, encoding methionine) that signals the first amino acid of a protein.

ribosome-binding site

Also called *Shine-Dalgarno sequence* . In bacteria, a stretch of nucleotides upstream of the start codon in an mRNA that hybridizes to the 16S rRNA of the ribosome, correctly positioning the mRNA for translation.

Shine-Dalgarno sequence

See **ribosome-binding site** .

acceptor site (A site)

The region of a ribosome that binds an incoming charged tRNA.

peptidyl-tRNA site (P site)

The region of a ribosome that contains the growing protein attached to a tRNA.

exit site (E site)

The region of a ribosome that holds the uncharged, exiting tRNA.

translocation

The energy-dependent movement of a ribosome to the next triplet codon along an mRNA.

release factor

A molecule that enters a ribosome A site containing an mRNA stop codon and initiates protein cleavage from the tRNA.

polysome

A cell structure consisting of multiple ribosomes performing translation on the same mRNA molecule.

Fig. 8.1

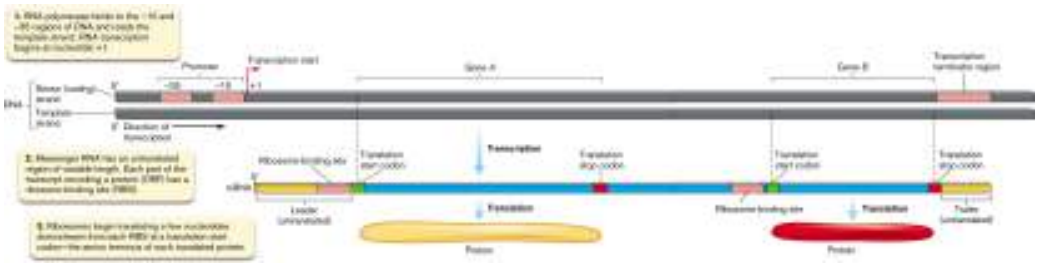


FIGURE 8.1 ■ Alignment of structural genes in a bacterial operon, the mRNA transcript, and protein products. In this figure, the term "gene" refers to the region of DNA that encodes a product. In this example, both genes encode protein. ORF = open reading frame.

Fig. 8.1

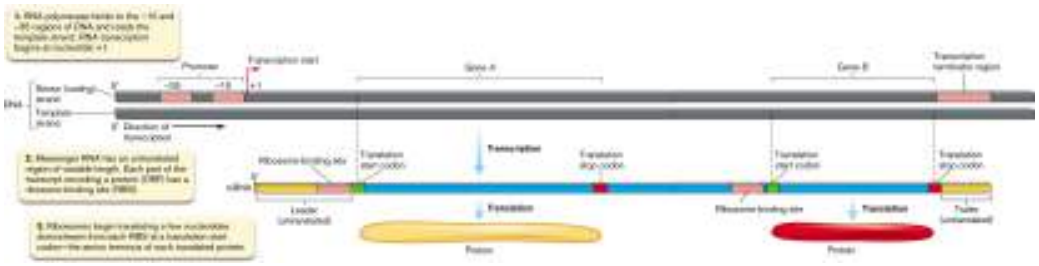


FIGURE 8.1 ■ Alignment of structural genes in a bacterial operon, the mRNA transcript, and protein products. In this figure, the term "gene" refers to the region of DNA that encodes a product. In this example, both genes encode protein. ORF = open reading frame.

Figure 8.12

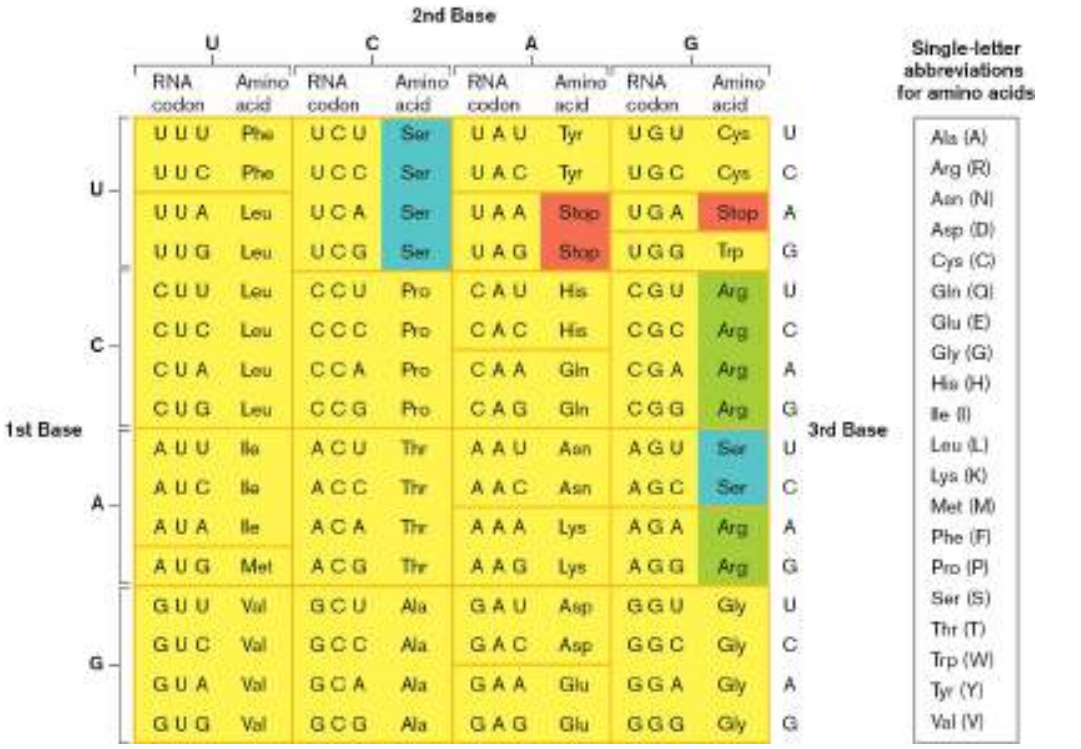


FIGURE 8.12 ■ The standard genetic code. Codons within a single box encode the same amino acid. Blue- and green-highlighted amino acids are encoded by codons in two boxes. Stop codons are highlighted red. Often, single-letter abbreviations for amino acids are used to convey protein sequences (see legend).

Figure 3.27

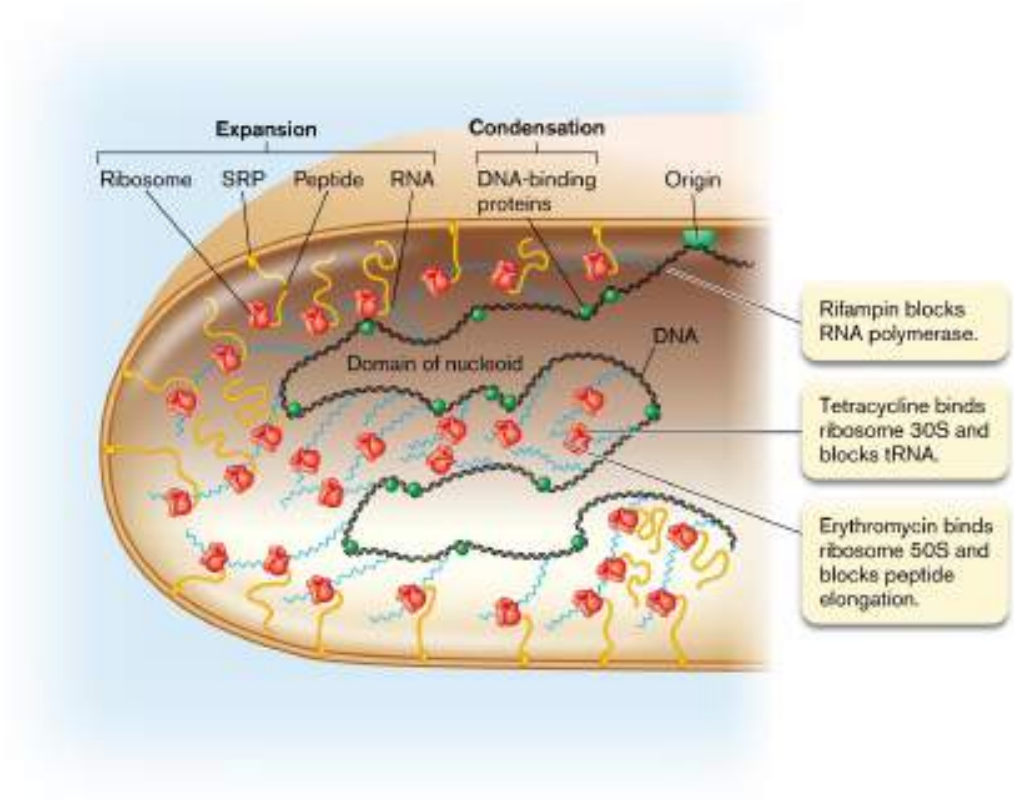


FIGURE 3.27 ■ DNA transcription and RNA translation to peptides. The nucleoid forms chromosome loops called domains, which loop out from the origin of attachment to the cell envelope. Bacterial transcription of DNA to RNA is coordinated with translation of RNA to make proteins. Growing peptide chains destined for the membrane bind the signal recognition particle (SRP) for membrane insertion.

8.4 Protein Modification, Folding, and Degradation

Once a protein is made, is it functional? For many proteins, translation is not the last step in producing a functional molecule. Often a protein must be modified after translation, either to achieve an appropriate 3D structure or to regulate its activity. Primary, secondary, and tertiary structures of proteins can be modified after the primary protein sequence has been assembled by the ribosomes. And what happens when a protein is damaged or is no longer needed? A healthy cell "cleans house" by degrading damaged or unneeded proteins. The precious amino acids are then recycled into making new proteins.

Protein Processing after Translation

Completed proteins released from the bacterial ribosome contain *N*-formylmethionine (fMet) at the N terminus, while archaeal and eukaryotic proteins have methionine (Met) in this position. In bacteria, archaea, and eukaryotes, this N-terminal amino acid is often removed or modified to produce the final polypeptide chain. For all three domains, methionine aminopeptidase removes the entire amino acid, while in bacteria another enzyme, methionine deformylase, provides the option to remove just the *N*-formyl group, leaving methionine. *N*-formylmethionine is important during the course of an infection because fMet peptides are produced only by bacteria and mitochondria, not by archaea or by the cytoplasmic ribosomes of eukaryotes. Our white blood cells can detect low concentrations of fMet peptides (about 10^{-12} M) as a warning sign of invading bacteria or of necrotic (dying) host cells releasing mitochondria.

A protein's function depends on its three-dimensional shape and its chemical properties. As we have seen, the standard genetic code

provides 20 options for side-chain chemistry at each amino acid position in the protein polymer. Expanding the code during translation to include pyrrolysine or selenocysteine provides only a modest increase in options for new protein chemistry. Substantially more options become available if proteins are modified after translation. In fact, all three domains of life utilize protein modification by the covalent attachment of molecules, including sugars, lipids, and inorganic substrates, to specific amino acids after a protein is synthesized. Over 900 different types of posttranslational protein modifications have been reported thus far from in vivo and in vitro studies. These modifications serve various roles: stabilizing proteins, localizing proteins to specific regions of the cell, and regulating the activity of the proteins. For the latter, a key property of many posttranslational modifications is that they are reversible; thus, protein function can be regulated by reactions that add or remove the modifications.

As we will see in Chapter 10 , the addition of phosphoryl or methyl groups (**Fig. 8.29**) can change the activity of signal transduction proteins in bacteria and archaea, such as the ones used in chemotactic motility. Adenylylation, the covalent attachment of adenosine 5′-monophosphate, can regulate the activity of enzymes such as glutamine synthetase. Attachment of acetyl groups via acetylation can serve multiple functions, including protein stabilization and the regulation of protein activity. Lipidation is the covalent attachment of lipids to proteins. Lipidation provides a hydrophobic tail that anchors these lipoproteins to the cytoplasmic membrane or to the outer membrane in Gram-negative organisms.

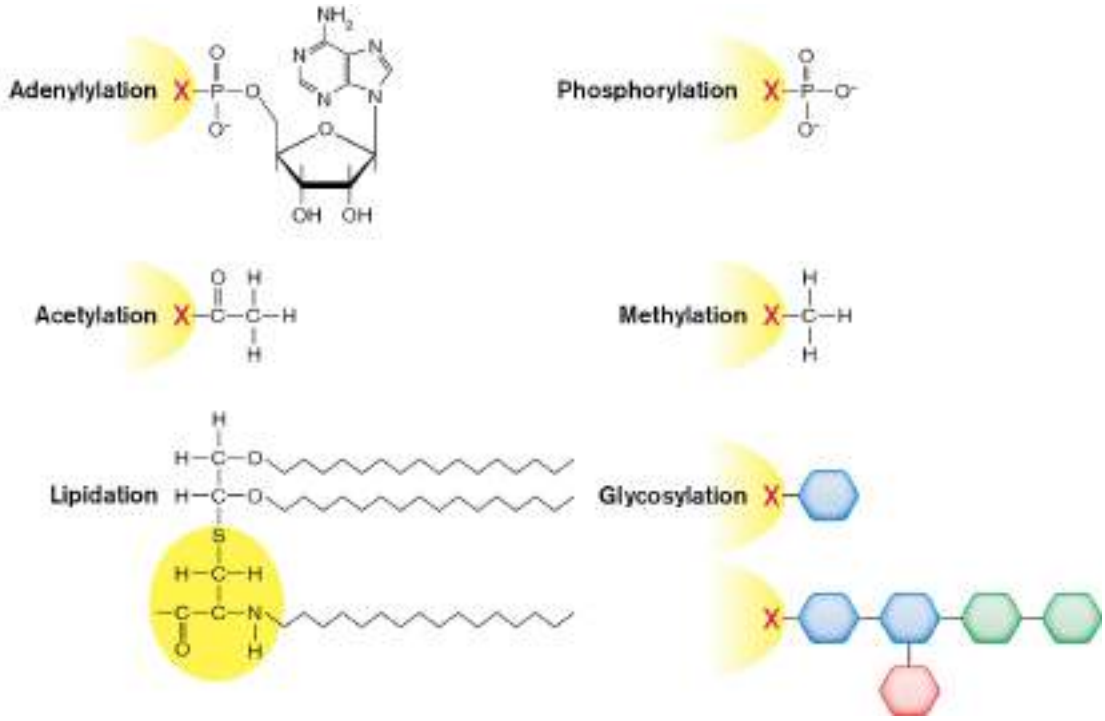


FIGURE 8.29 ■ Examples of posttranslational modifications to proteins. Molecules are typically covalently linked to an amino acid side chain (X). The lipidation example shown involves the modification of a protein's N-terminal cysteine (red) at both the side chain and the amino group. Glycosylations can involve single sugars (represented by hexagons) or polysaccharide chains, often of varied sugar composition and degree of branching within the chain.

Glycosylation is the covalent addition of monosaccharides or polysaccharides to generate glycoproteins (**Fig. 8.29** , lower right). In bacteria and archaea, glycosylation can occur in a number of cell-surface proteins, including flagellar subunits, pili, adhesins, and the proteins that compose the S-layer. Glycoproteins can be involved in important microbial processes, such as biofilm formation, virulence, and colonization of the human gut. For example, glycosylation is important for *Bacteroides fragilis,* a member of the normal human gut microbiome. Critically, mutants unable to glycosylate proteins are outcompeted in the gut by wild-type *B. fragilis* . Glycosylation is also important for pathogens of the gut. *Clostridioides difficile* is a

pathogen that can exploit the loss of normal gut microflora after antibiotic treatment (see Chapter 26). One of its secreted toxins, TcdA, is a glycosyltransferase that can glycosylate important regulatory proteins in human cells. These glycosylations inactivate the proteins, causing a disruption in normal regulation in the human cells, and ultimately rendering the host gut more susceptible to infection by *C. difficile* .



The application of mass spectrometry to the investigation of cell proteins (proteomics) has led to major advances in our understanding of posttranslational modifications and has revealed that we have vastly underappreciated the extent to which bacterial proteins are modified after translation is complete. Mass spectrometry experimentally determines the exact mass of an unknown protein or peptide fragment. The experimentally determined mass can be compared to the predicted mass using genomic information for the organism, and by these comparisons the protein or protein fragment is identified.

To begin the analysis, proteins present in cell extracts are digested into distinct peptide fragments with a site-specific protease (trypsin; **Fig. 8.30** , steps 1 and 2). The fragment mixture is passed through an analytical column that separates peptides according to differences in hydrophobicity or some other parameter (step 3). The column effluent is then directly fed through tandem mass spectrometry (MS-MS) instrumentation (step 4). In MS-MS, each proteolytic fragment is subfragmented by ionization to produce progressively smaller secondary fragments missing one or more amino acids. Because the mass of each amino acid is distinct, MS-MS analysis can determine the amino acid sequence of the initial proteolytic fragment (step 5). Sophisticated computer programs identify the proteins by comparing the amino acid sequences of each protein fragment with the predicted sequences of proteolytic fragments from all ORFs in a genome (steps 6 and 7).

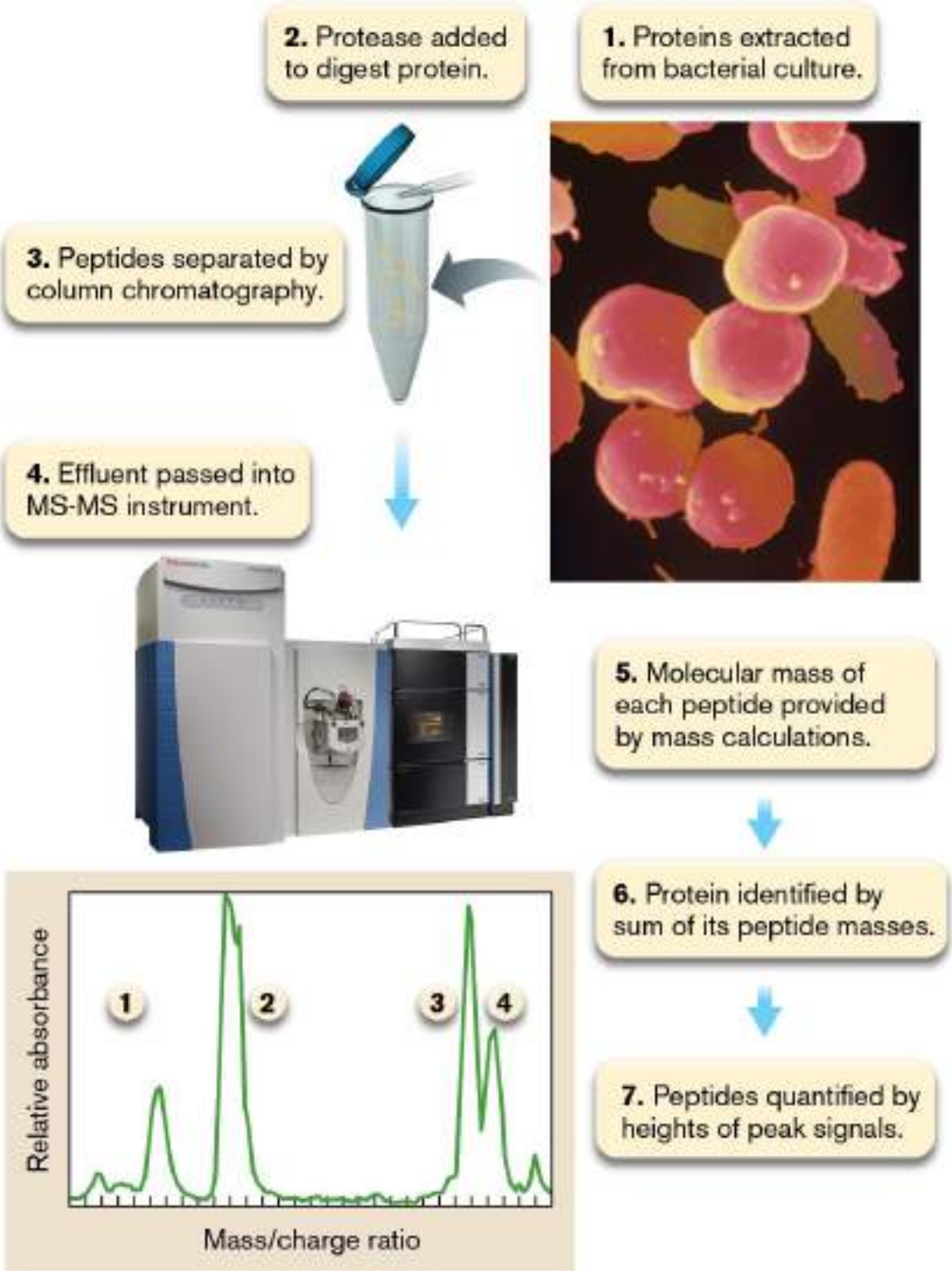


FIGURE 8.30 ■ Identifying proteins directly from whole-cell extracts by mass spectrometry. Proteins extracted from a bacterial culture are digested into peptides with trypsin. The peptides are separated by column chromatography and analyzed by mass spectrometry (here by the Thermo

Scientific Q Exactive hybrid quadrupole-Orbitrap mass spectrometer). In tandem mass spectrometry (MS-MS), the mass of each peptide is determined first (peaks 1–4 in the graph), and then selected peptides are subjected to additional fragmentation by ion spray (not shown). Each resulting peptide fragment will differ in size by one or more amino acids. Knowing the mass of each amino acid and the masses of the different peptide fragments enables extrapolation of the original peptide's sequence.

COURTESY OF THERMO FISHER SCIENTIFIC
SIMKO/VISUALS UNLIMITED, INC.



Peptide fragments that contain posttranslational modifications such as acetylations and phosphorylations can be isolated from unmodified fragments using special analytical columns. The samples are then subjected to mass spectrometry to identify the modified fragments on the basis of their amino acid sequence. From such analyses, hundreds of proteins have been determined to be acetylated in *Escherichia coli* . In a study of a strain of the marine cyanobacterium *Synechococcus* , mass spectrometry data revealed 2,230 proteins that have one or more posttranslational modifications. This study discovered many previously unknown modifications to proteins involved in photosynthetic light harvesting. An important implication of this discovery is that fundamental physiological processes such as the photosynthetic conversion of light energy to chemical energy are still only partially understood at the molecular level.

Protein Folding: Assume the Position

As a new protein emerges from the ribosome, how does it fold into exactly the correct shape to do its job? Christian Anfinsen (1916–1995) won the 1972 Nobel Prize in Chemistry for demonstrating that, for some proteins, folding is governed solely by the protein itself. In other words, the optimal 3D structure of a

protein is determined solely by the linear sequence of amino acid residues. But three decades later, other scientists discovered that the folding of many proteins requires assistance from other proteins. These helper proteins are called chaperones (or chaperonins). Chaperones associate with target proteins during some phase of the folding process and then dissociate, usually after folding of the target protein is completed. Although chaperones exhibit some specificity, a given chaperone can help fold many different types of proteins.

The major chaperone family in most species includes GroEL, GroES, DnaK, DnaJ, and trigger factor. Because their levels in *E. coli* increase in response to high-temperature stress, these chaperones were originally named heat-shock proteins (HSPs) , and they are, in fact, more resistant to heat denaturation than the average protein is. Representatives of these chaperones are found in all species. Because their molecular masses (in units of kilodaltons; kDa) are similar, DnaK examples throughout nature are called HSP70s (70 kDa), while homologs of GroEL and DnaJ are called, respectively, HSP60s (60 kDa) and HSP40s (40 kDa).

The GroEL and GroES chaperones form a stacked ring with a hollow center like a barrel (**Fig. 8.31A**). The chaperoned protein fits inside. The small, capping protein GroES controls entrance to the chamber, as shown in **Figure 8.31A** . Cycles of ATP binding and hydrolysis cause conformational changes within the chamber that can reconfigure target proteins. DnaK (HSP70) chaperones have a very different structure (**Fig. 8.31B**). They do not form rings like the GroEL and GroES chaperones but can clamp down on a peptide to assist folding. Proteins emerging from a bacterial ribosome enter a folding pathway that involves a hierarchy of these chaperones.

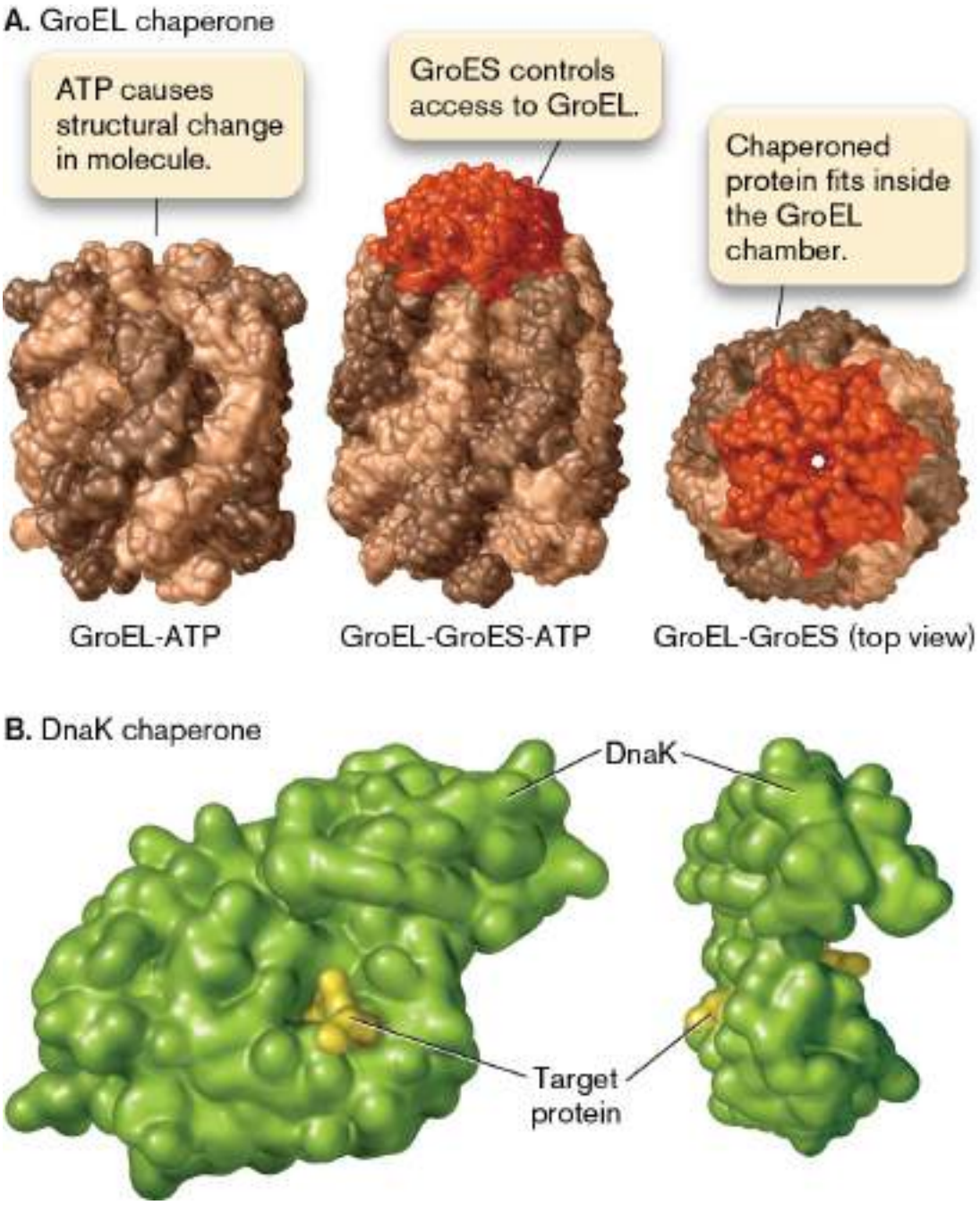


FIGURE 8.31 ■ *E. coli* GroEL, GroES, and DnaK structures. A. Three-dimensional reconstructions of GroEL-ATP, GroEL-GroES-ATP, and GroEL-GroES from cryo-electron microscopy. The first two panels are side views; the third panel is a top view. GroES is red. (PDB codes: 2C7E, 1PCQ) **B.** DnaK (HSP70) clamping down on a peptide (yellow). (PDB code: 1DKX)

Protein Degradation: Cleaning House

What happens when a cell no longer needs a specific protein or when a cell synthesizes a protein with incorrect amino acids? Because the cell's needs constantly change, the presence of useless proteins can adversely affect the cell and they must be destroyed. This is particularly true of regulatory proteins, whose concentrations must change with time or in response to alterations in the cellular condition.

Many normal proteins contain degradation signals called degrons that dictate the stability of a protein. The N-terminal rule describes one type of degron. Recall that for some proteins, the N-terminal fMet is removed after translation, leaving a new amino acid at the N terminus. The N-terminal rule states that the identity of the N-terminal amino acid correlates with the stability of the protein. For example, proteins beginning with leucine, phenylalanine, tryptophan, or tyrosine experience a short half-life (2 minutes or less), whereas proteins with aspartic acid, glutamic acid, or cysteine in the lead position have a longer half-life. A protein called ClpS facilitates degradation of these short-lived proteins. ClpS recognizes the destabilizing N-terminal amino acids and then presents the protein to the bacterial ClpAP protease.

Abnormally folded proteins are recognized by proteases in part because hydrophobic regions that are normally buried within the protein's 3D structure become exposed. The protein is progressively degraded into smaller and smaller pieces by a series of these proteases. Initial cuts, usually involving ATP-dependent endoproteases like Lon protein or ClpP, are followed by digestion with tripeptidases and dipeptidases. Endopeptidases cleave proteins somewhere within the sequence, but not from the ends of the sequence. Many peptidases use ATP hydrolysis to help unfold the target protein prior to digestion. Unfolding is necessary for the target protein to slide into a barrel-shaped protease such as ClpAP, ClpXP, or ClpYQ in bacteria (**Fig. 8.32**).

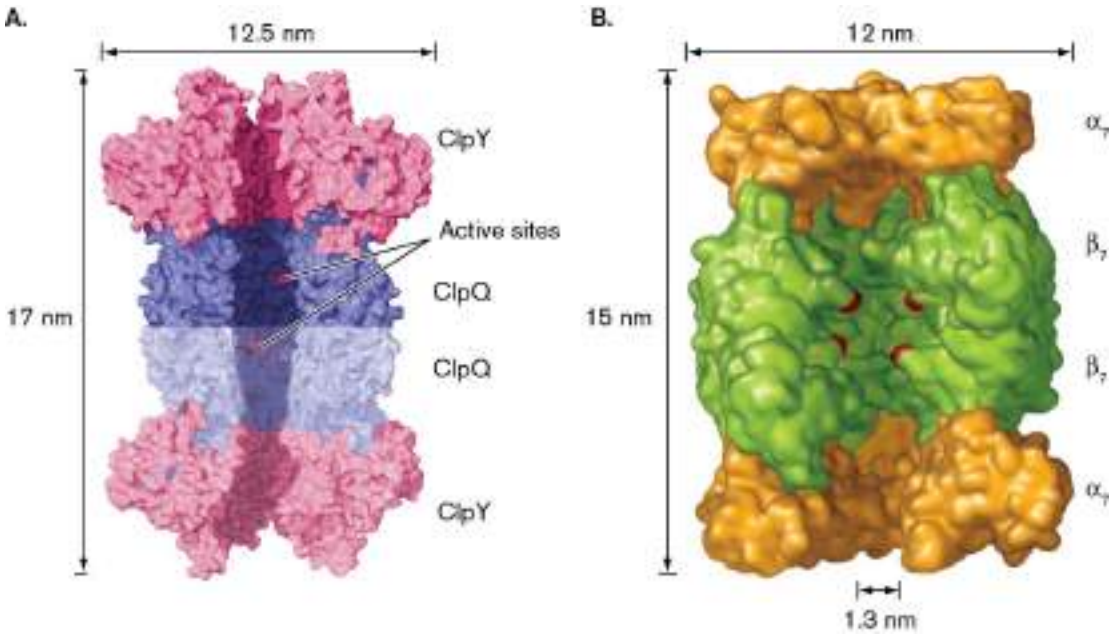


FIGURE 8.32 ■ Protein degradation machines. A. Bacterial ClpY ATPase and ClpQ protease (*Haemophilus influenzae*). (PDB code 1G3I) Two of the six subunits from each ring were removed to reveal the interior cavity. The active sites involved in peptide bond cleavage are indicated in pink. **B.** The 20S proteasome from the methanoarchaeon *Methanosarcina thermophila* . (PDB code: 1G0U)

Bacterial Clp proteases have a proteolytic core made of two homoheptameric protein rings of either ClpP or ClpY. The ClpP protease has interchangeable homohexameric ATPase caps made of ClpX, ClpA, ClpB, or ClpC, each of which recognizes different substrates. ClpY plays a similar capping role for ClpQ. The accessory proteins recognize and present different substrate proteins to the ClpP protease, thereby regulating which proteins are degraded. Protein-degrading enzymes are classified as serine, cysteine, or threonine proteases, depending on the key residue in their active sites.

Bacterial Clp proteases are structurally similar to eukaryotic proteasomes, which are even more complex protein-degrading machines. Proteasomes are found primarily in eukaryotes and

archaea, although a few bacteria, such as the pathogen *Mycobacterium tuberculosis* , have them. Eukaryotic proteasomes recognize and then degrade proteins tagged by ubiquitin, a 76-amino-acid peptide. Some bacterial and viral pathogens exploit ubiquitination to reroute host metabolism.

What happens to proteins damaged by stress? Are they always degraded? Microbes are constantly exposed to environmental insults such as high temperature or pH extremes, which damage proteins and cause them to misfold. As an energy-saving device and to prevent interruption of protein function, injured proteins go through a kind of triage process that evaluates whether they are salvageable or must be destroyed before they can endanger the cell. Chaperones constantly hunt for misfolded (or otherwise damaged) proteins and attempt to refold them. But if the protein is released from a chaperone and remains misfolded, it can, by chance, either reengage the chaperone or bind a protease that destroys it (**Fig. 8.33**). This fold-or-destroy triage system is essential if a microbe is to survive environmental stress.

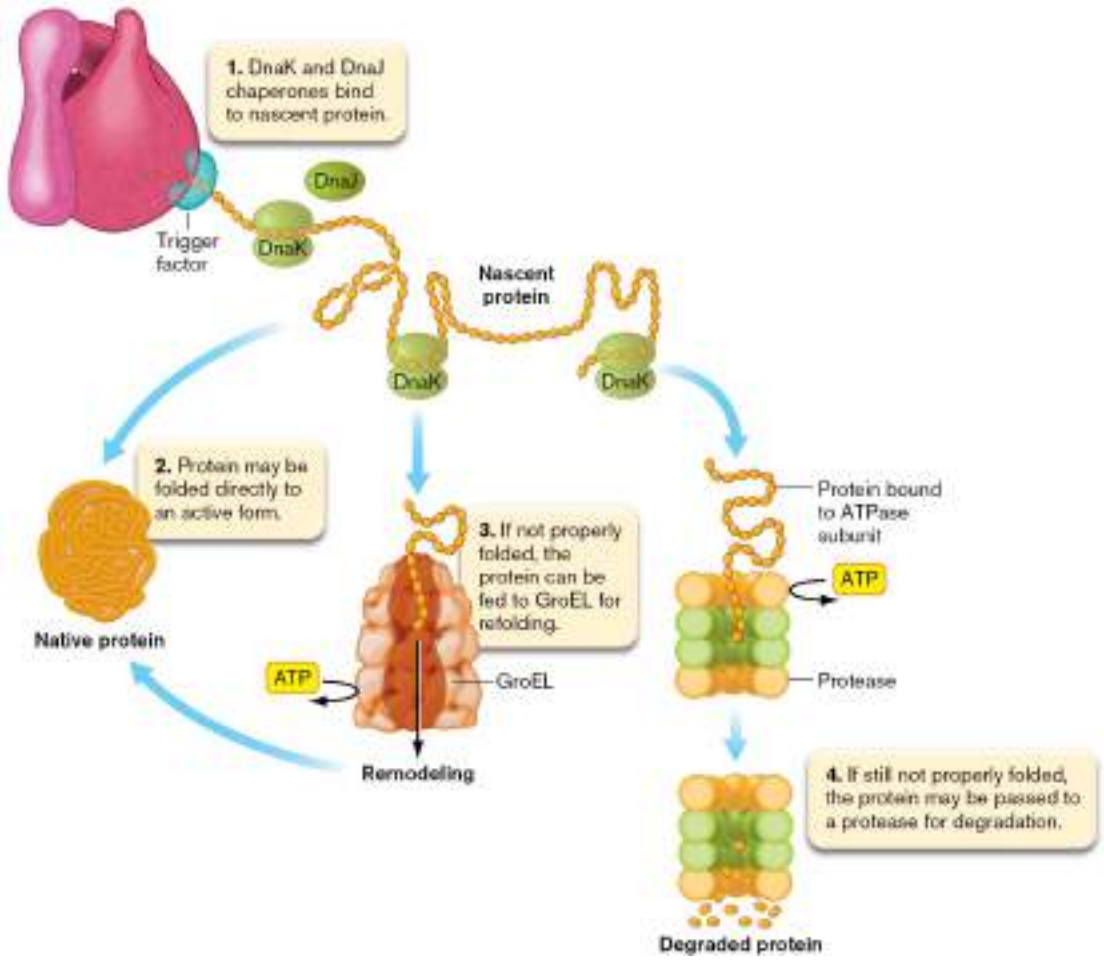


FIGURE 8.33 ■ *E. coli* protein folding-versus-degradation triage pathways. The diagram depicts what can happen to a newly synthesized protein. However, a protein that unfolds in response to environmental stress (for example, heat) will undergo the same triage process.

To Summarize

- **Protein modifications** are made after translation is complete.
- **The N-terminal amino acid** (fMet or Met) can be removed by methionine aminopeptidase, or, in bacteria, just the formyl group can be removed by methionine deformylase.

- **Covalent additions** to proteins can enhance stability, add localization motifs, and regulate changes in activity.
- **Chaperone proteins** help translated proteins fold properly.
- **All proteins in all cells are eventually degraded** by specific devices such as proteases or proteasomes.
- **The N-terminal rule** describes one type of degradation signal (degron) that marks the half-life of a protein (that is, how long it takes 50% of the protein to degrade).
- **ATP-dependent proteases** such as Lon or ClpP usually initiate the degradation of a large protein.
- **Damaged proteins** enter chaperone-based refolding pathways or degradation pathways until the protein is repaired or destroyed.

Glossary

mass spectrometry
: An analytical technique that measures the mass of molecules. Molecules are ionized and sorted according to their mass-to-charge (m / z) ratio.

chaperone or chaperonin
: A protein that helps other proteins fold into their correct tertiary structure.

heat-shock protein (HSP)
: A chaperone protein whose synthesis is induced by high-temperature stress.

N-terminal rule
: The tendency of the N-terminal amino acid of a protein to influence protein stability.

8.5 Secretion: Protein Traffic Control

Microorganisms, especially Gram-negative bacteria, face a challenge in delivering proteins to different target locations of the cell. Recall that Gram-negative microbes are surrounded by two layers of membrane (the inner membrane, or cell membrane, and an outer membrane), between which lies a periplasmic space (see Section 3.3). Many proteins are specifically destined for one or another of these cell compartments. Other proteins are secreted completely out of the cell into the surrounding environment (for example, hemolysins that lyse red blood cells). But how do these diverse proteins know where to go? Protein traffic out of the cell is directed by an elaborate set of protein secretion systems. Each system selectively delivers a set of proteins originally made in the cytoplasm to various extracytoplasmic locations.

The term "secretion" is used to describe movement of a protein out of the cytoplasm. Some protein secretion systems move proteins out of the cytoplasm into the cytoplasmic membrane and across the membrane to the periplasm, others move proteins to the outer membrane, and still others deliver proteins across both of the membranes and into the surrounding environment. An added complication of protein export is that periplasmic proteins are usually delivered unfolded into the periplasm and require another set of chaperones to fold properly in this cell compartment.

Protein Export Out of the Cytoplasm

Proteins destined for the bacterial cell membrane (such as membrane transport proteins), periplasm (binding proteins), outer membrane (porins), or extracellular spaces (proteases) require special export systems. These systems manage to move hydrophilic proteins through one or more hydrophobic membrane barriers. Proteins destined for export are identified by hydrophobic N-terminal

signal sequences of 15–30 amino acids. The hydrophobic nature of these signal sequences allows them to become embedded within the membrane once they are released by the transmembrane secretion machinery. For proteins exported beyond the inner membrane, a protease cleaves this signal sequence to release the protein into the periplasm. For inner membrane proteins, the signal sequences are retained and tether the nascent proteins to the membrane. Some inner membrane proteins, such as nutrient transporters, have multiple membrane-spanning regions that weave back and forth across the membrane. These proteins contain additional hydrophobic regions (20–25 amino acids) within the polypeptide that serve as the transmembrane domains.

Protein Export to the Cell Membrane

The pathway leading proteins to the inner (cell) membrane begins with a complex called the signal recognition particle (SRP) . In *Escherichia coli* , the SRP consists of a 54-kDa protein (Ffh) complexed with a small RNA molecule (4.5S RNA). Inner membrane proteins contain very hydrophobic signal sequences, and their location at the N terminus of the protein means that they are synthesized early during translation. SRP binds to these hydrophobic signal sequences as they are being translated (**Fig. 8.34**) and halts further translation in the cytoplasm. The nascent protein with its stalled, nontranslating ribosome is delivered by SRP to the membrane-embedded protein FtsY, which then shuttles the complex to the transmembrane SecYEG secretion machinery (translocon). Contact with SecYEG stimulates GTP hydrolysis by SRP, which then releases the signal sequence and transfers it and the attached ribosome to SecYEG. The signal sequence enters the translocon and is released into the membrane through a lateral gate of the channel. No longer blocked by SRP, translation of the protein resumes to completion, and any additional transmembrane domains are released through the lateral gate.

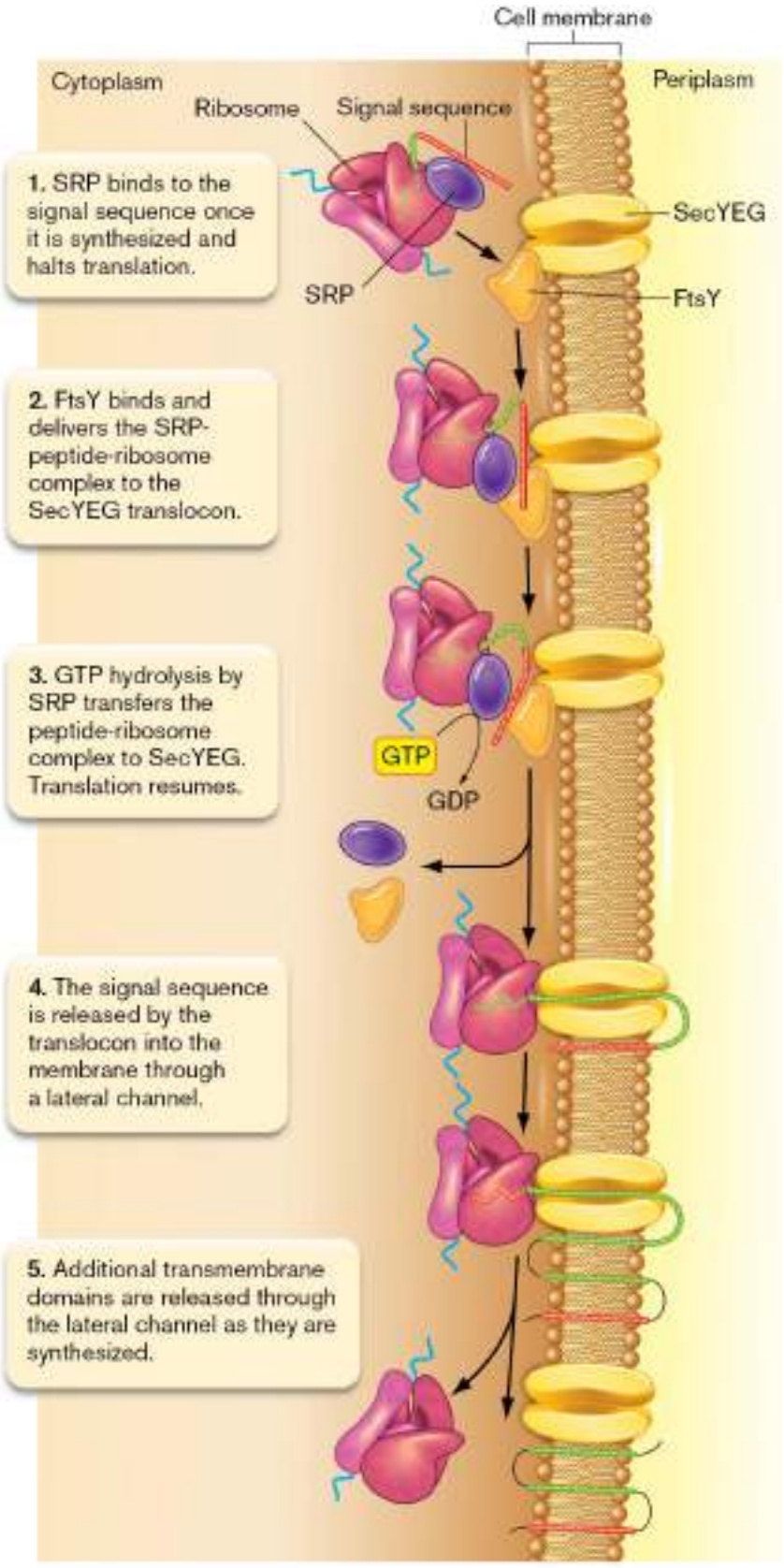

Cell membrane
Cytoplasm
Periplasm
Ribosome
Signal sequence
1. SRP binds to the signal sequence once it is synthesized and halts translation.
SecYEG
SRP
FtsY
2. FtsY binds and delivers the SRP-peptide-ribosome complex to the SecYEG translocon.
3. GTP hydrolysis by SRP transfers the peptide-ribosome complex to SecYEG. Translation resumes.
GTP
GDP
4. The signal sequence is released by the translocon into the membrane through a lateral channel.
5. Additional transmembrane domains are released through the lateral channel as they are synthesized.

FIGURE 8.34 ■ SRP and cotranslational export in *E. coli*. A ribosome "paralyzed" by an SRP does not resume translating protein until it encounters FtsY in the membrane. Translation can then recommence, with the translated peptide fed into the SecYEG translocon. Transmembrane sections of the peptide are released into the membrane by a lateral gate.

Protein Export to the Periplasm: The General Secretion Pathway

The periplasm contains important proteins that bind nutrients for transport into the cell and other proteins that carry out enzymatic reactions. For example, one form of superoxide dismutase (SOD), an enzyme that degrades superoxide, is a periplasmic protein in *Salmonella enterica* and other Gram-negative bacteria. Many periplasmic proteins, such as SOD and maltose-binding protein (which imports the sugar maltose), are delivered to the periplasm by a common pathway called the general secretion pathway.

The general secretion pathway has several steps. First, the peptide is completely translated in the cytoplasm and released by the ribosome (**Fig. 8.35** , step 1). The signal sequences of these proteins are less hydrophobic than those of inner membrane proteins, so the SRP protein does not bind them and halt translation. Trigger factor interacts with newly synthesized protein as the protein exits the ribosome and keeps pre-secreted proteins in a loosely folded conformation, awaiting interaction with the next component of the secretion machinery. In proteobacteria, a second chaperone, SecB, can also bind and protect pre-secreted proteins on their way to the membrane. Keeping a pre-secretion protein unfolded in the cytoplasm assists the secretion process because the narrow transmembrane channel of the general secretion pathway can only translocate proteins in their unfolded state. A homodimer of SecA proteins peripherally associated with the membrane-spanning

SecYEG translocon then binds to the N-terminal signal sequence and several downstream regions of the pre-protein (step 2). SecA can bind and hydrolyze ATP, which catalyzes its essential function of driving the translocation of the pre-protein through the translocon channel.

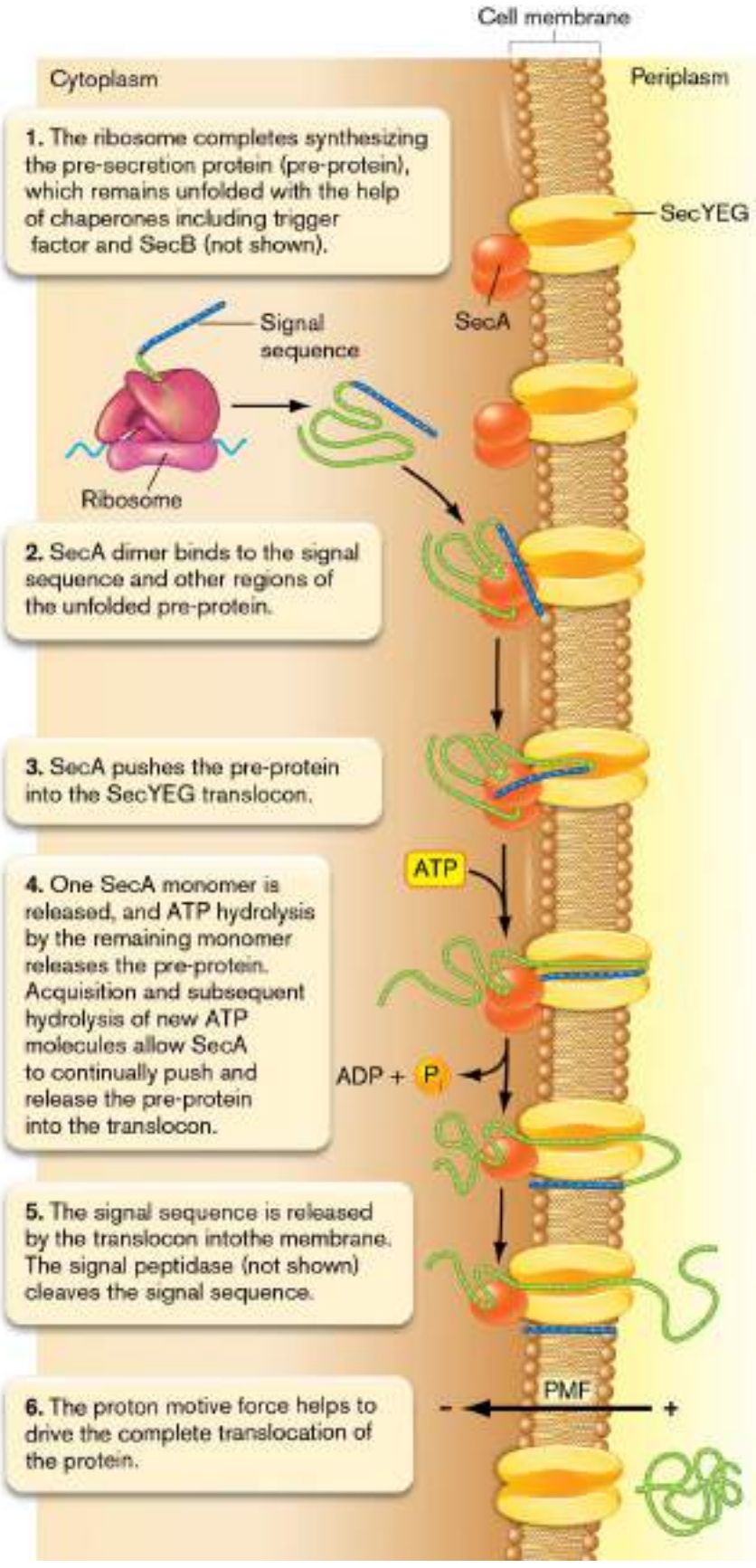

Cell membrane
Cytoplasm
Periplasm
1. The ribosome completes synthesizing the pre-secretion protein (pre-protein), which remains unfolded with the help of chaperones including trigger factor and SecB (not shown).
SecYEG
SecA
Signal sequence
Ribosome
2. SecA dimer binds to the signal sequence and other regions of the unfolded pre-protein.
3. SecA pushes the pre-protein into the SecYEG translocon.
ATP
4. One SecA monomer is released, and ATP hydrolysis by the remaining monomer releases the pre-protein. Acquisition and subsequent hydrolysis of new ATP molecules allow SecA to continually push and release the pre-protein into the translocon.
ADP + P
5. The signal sequence is released by the translocon intothe membrane. The signal peptidase (not shown) cleaves the signal sequence.
PMF
6. The proton motive force helps to drive the complete translocation of the protein.

FIGURE 8.35 ■ The SecA-dependent general secretion pathway. This pathway exports many proteins across the cell membranes of Gram-negative and Gram-positive bacteria. PMF = proton motive force.

The translocation process is not fully resolved. One model has SecA acting like a plunger (**Fig. 8.35** , step 3). It inserts deep into the SecYEG channel, shoving about 20 amino acids of the target export protein into the channel. At this stage the SecA dimer dissociates, and the SecA monomer remaining at the SecYEG channel continues the secretion process. ATP hydrolysis causes SecA to release the protein and withdraw (step 4). The SecA monomer can bind fresh ATP, rebind the target protein, and reinsert, pushing another 20 amino acids through SecYEG. The proton motive force at the cytoplasmic membrane (see Chapters 4 and 14) also contributes to the process and is thought to help drive translocation of the protein after SecA release (step 6).

When the N-terminal signal sequence of a periplasmic protein enters the SecYEG translocon, it is released into the membrane through a lateral gate in a manner similar to the signal sequence of inner membrane proteins. Next, a periplasmic signal peptidase enzyme such as LepB in *E. coli* cleaves the signal sequence, removing the transmembrane tether and allowing the mature protein to diffuse from the membrane (**Fig. 8.35** , step 5). Signal peptidases, however, will not cleave signals from proteins destined to stay embedded within the membrane (integral membrane proteins).

Note: "Translocation" can refer to the movement of a ribosome along mRNA or it can describe the movement of a protein from one cell compartment (cytoplasm) to another (periplasm).

Periplasmic proteins delivered by the Sec system arrive unfolded and inactive. Because the folding chaperones mentioned earlier are cytoplasmic, periplasmic proteins need a different set of dedicated chaperones to guide their tertiary folding. Another problem with periplasmic proteins is that the oxygen-rich environment of aerobic cells can oxidize cysteines within a protein and produce inappropriate cysteine disulfide bonds that destroy enzyme function. Special periplasmic disulfide reductases are required to reduce these S−S bonds back to two SH groups. Many periplasmic proteins, however, need certain disulfide bonds to be active, so the periplasm also contains a disulfide bond catalyst (DsbA) to make those bonds.

Protein Export in Gram-Positive Bacteria, Eukaryotes, and Archaea

Gram-positive bacteria must also export proteins across the cell membrane and then fold and process them once they are secreted. However, Gram-positive bacteria lack the periplasmic space needed to facilitate interactions between newly secreted proteins and the accessory processing proteins. Many streptococci solve this problem by clustering their secretion systems and accessory factors at a microdomain of the cytoplasmic membrane called the ExPortal. The ExPortal is located near the cell septum and appears linked to peptidoglycan synthesis (**Fig. 8.36**). Proteins in the ExPortal include HtrA (which assists in pili formation and covalent attachment of proteins to the cell wall) and sortase (which aids in maturation of secreted proteins), as well as Sec system components and chaperones. Some proteins that pass through the ExPortal are truly secreted; others are not. The membrane-embedded proteins have a transmembrane domain (missing in Gram-negative homologs) that anchors the proteins to the Gram-positive cytoplasmic membrane. The anchored proteins are extracellular but will not float away.

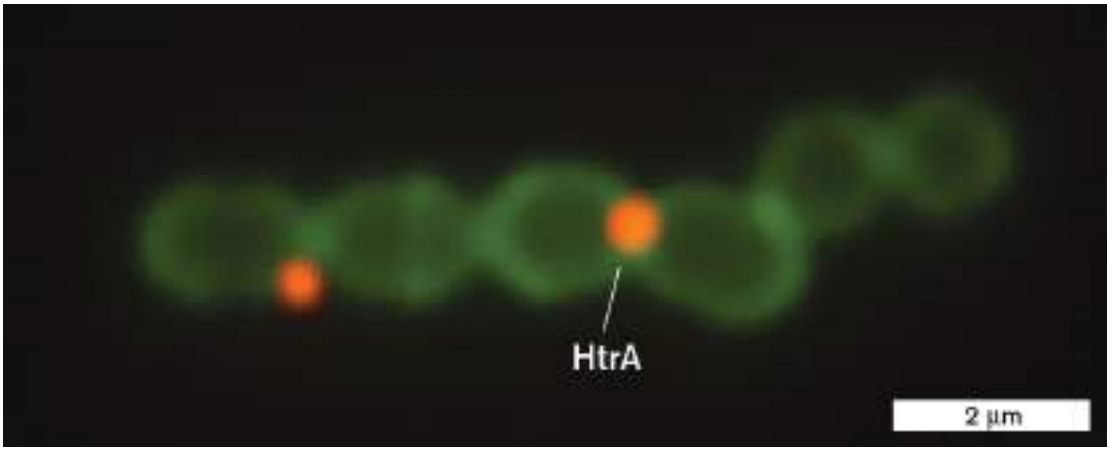


FIGURE 8.36 ■ Location of the ExPortal in *Streptococcus pyogenes*. The ExPortal protein HtrA was identified using immunofluorescence, a microscopy technique involving a fluorophore-attached antibody that binds a specific cell component. Note that HtrA is located at the septum.

M. G. CAPARON ET AL. 2005. *MOL MICROBIOL* . **58** :959–968

Eukaryotic microbes such as the yeast *Saccharomyces cerevisiae* also possess secretion systems that move proteins to the membrane and beyond. Most secreted proteins in eukaryotes are exported cotranslationally, using SRP to pause translation until the peptide is delivered to the SRP receptor and transferred to the Sec translocase. The Sec translocase of eukaryotes has homologs of the SecY and SecE proteins of bacteria, called Sec61α and Sec61γ, respectively. Their third component, Sec61β, however, is unrelated to the bacterial SecG. The translating ribosome provides the driving force for transport during cotranslational secretion (**Fig. 8.37A**). For posttranslational secretion, eukaryotes use the ATP-binding protein Bip, located in the lumen of the endoplasmic reticulum (ER), in coordination with a transmembrane complex of Sec62 and Sec63 to ratchet transport through the Sec complex (**Fig. 8.37B**).

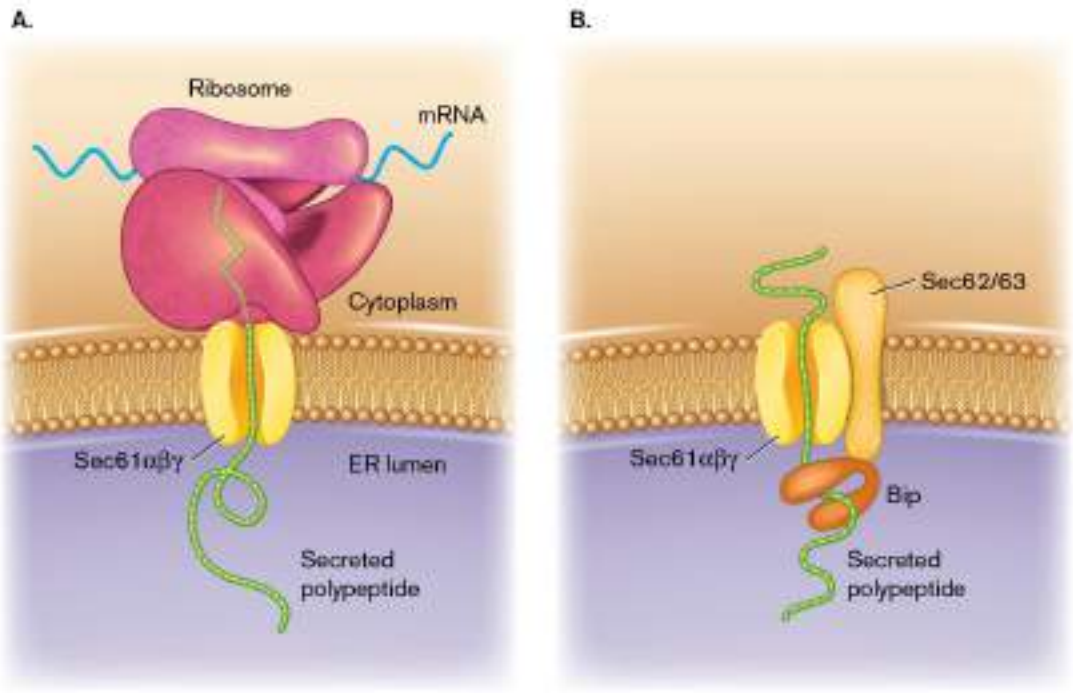


FIGURE 8.37 ■ Protein secretion in eukaryotes. A. Cotranslational secretion though the Sec61αβγ channel is driven by the ribosome. **B.** Posttranslational secretion through the channel is driven by the ratcheting action of Bip and Sec62/63.

Source: Modified from T. A. Rapoport et al. 2017. *Annu. Rev. Cell Dev. Biol.* **33** :369–390, fig. 8.

Archaeal secretion systems are more similar to those of eukaryotes than to those of bacteria. However, archaea lack both SecA and Bip; thus, the driving force of protein secretion in archaea is still unknown.

Export of Prefolded Proteins to the Periplasm

In a dramatic departure from Sec-dependent transport systems, bacterial proteins can also be transported fully folded across the membrane to their periplasmic destination. Prefolding of periplasmic proteins in the cytoplasm enables cells to carefully control the

insertion of cofactors. Cofactors such as flavins are important for proteins involved in respiration (see Chapter 14), and some of these cofactors are embedded within the protein structure as the protein folds after translation. Metalloproteins use metals as cofactors, but different metals can compete for the metal-binding sites within the protein. Prefolding of the metalloprotein in the cytoplasm ensures that the correct metal is inserted. In addition, other proteins without cofactors simply fold very quickly in the cytoplasm and need to be translocated in their fully folded state.

These proteins contain the amino acid motif RRXFXK within their N-terminal signal sequence (where R = arginine, F = phenylalanine, K = lysine, and X = any amino acid). This sequence, called the "twin arginine motif," targets the protein to the membrane-embedded twin arginine translocase (TAT), a transport complex that assembles on demand to ship fully folded proteins across the cell membrane to the periplasm (**Fig. 8.38**). Signal sequence binding triggers a conformational change in a complex of TatB and TatC proteins, which, with the assistance of the proton motive force, recruits and oligomerizes TatA, forming the translocase that enables fully folded protein substrates to pass through the membrane. After translocation the signal sequence is cleaved and the protein can diffuse into the periplasm. Once the protein leaves the translocase, the translocase dissociates into TatA monomers, and the system resets to receive the next secreted protein.

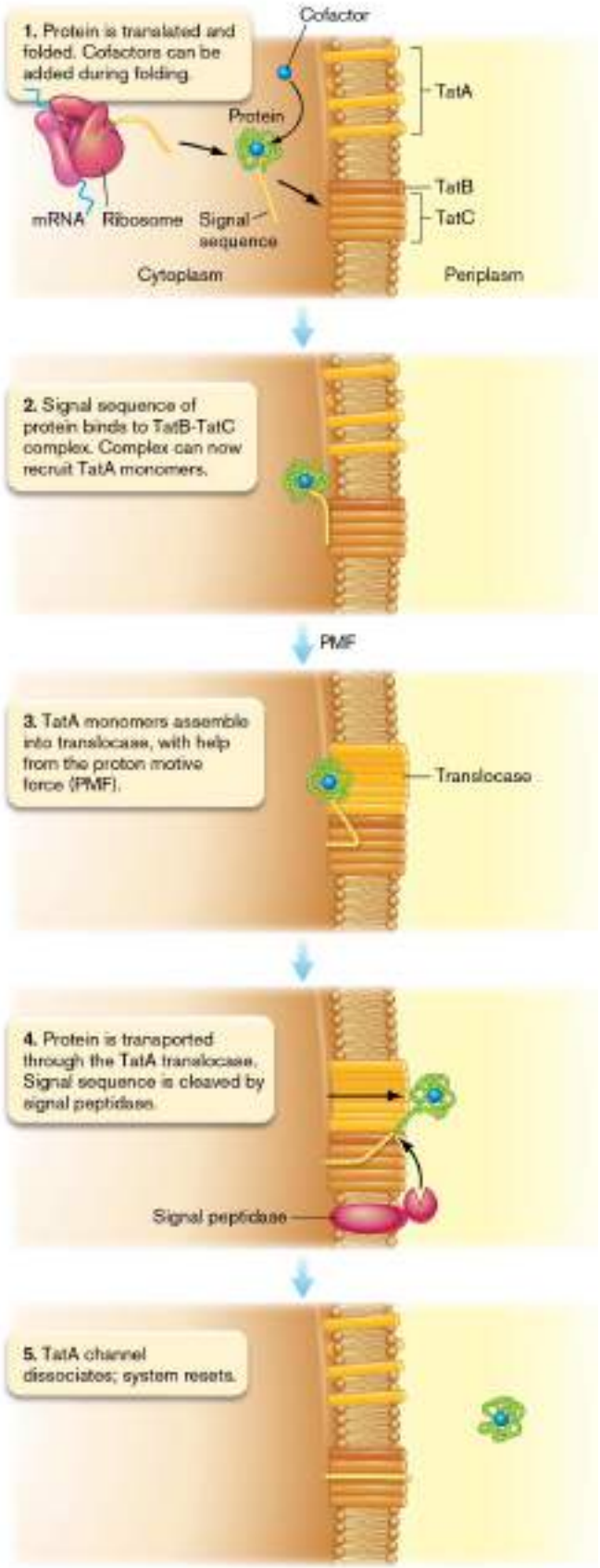

1. Protein is translated and folded. Cofactors can be added during folding.
Cofactor
TatA
Protein
TatB
TatC
mRNA
Ribosome
Signal sequence
Cytoplasm
Periplasm
2. Signal sequence of protein binds to TatB-TatC complex. Complex can now recruit TatA monomers.
PMF
3. TatA monomers assemble into translocase, with help from the proton motive force (PMF).
Translocase
4. Protein is transported through the TatA translocase. Signal sequence is cleaved by signal peptidase.
Signal peptidase
5. TatA channel dissociates; system resets.

FIGURE 8.38 ■ The twin arginine translocase (TAT). Model for the Tat protein translocase, which includes proteins TatA, TatB, and TatC.

Source: Modified from Tracy Palmer and Ben C. Berks. 2012. *Nat. Rev. Microbiol.* **10** :483–496, fig. 2.

Archaea also possess a TAT system, but eukaryotes do not. It is thought that a TAT system is unnecessary for eukaryotes, because translocation in eukaryotes occurs in the endoplasmic reticulum. In contrast to the extracytoplasmic space in bacteria and archaea, the lumen of the ER is ATP-rich, and is host to ATP-dependent enzymes that can provide transport assistance on the other side of the membrane.

Journeys to the Outer Membrane

Outer membrane proteins (OMPs) are made in the cytoplasm and exported to the periplasm by the SecA-dependent secretion system. Some OMPs have hydrophobic C-terminal signal sequences that facilitate insertion into the outer membrane, but all have a beta barrel structure that ultimately suits them for outer membrane placement; one example is TolC (see **Fig. 8.39**). Periplasmic chaperones prevent aggregation of OMPs as they traverse the periplasm and deliver the proteins to a multisubunit, outer membrane machine called the BAM (beta barrel assembly machine) complex that facilitates OMP assembly in the outer membrane. Although the players seem to be known, the mechanism by which beta barrels are folded and inserted into the outer membrane bilayer remains unclear.

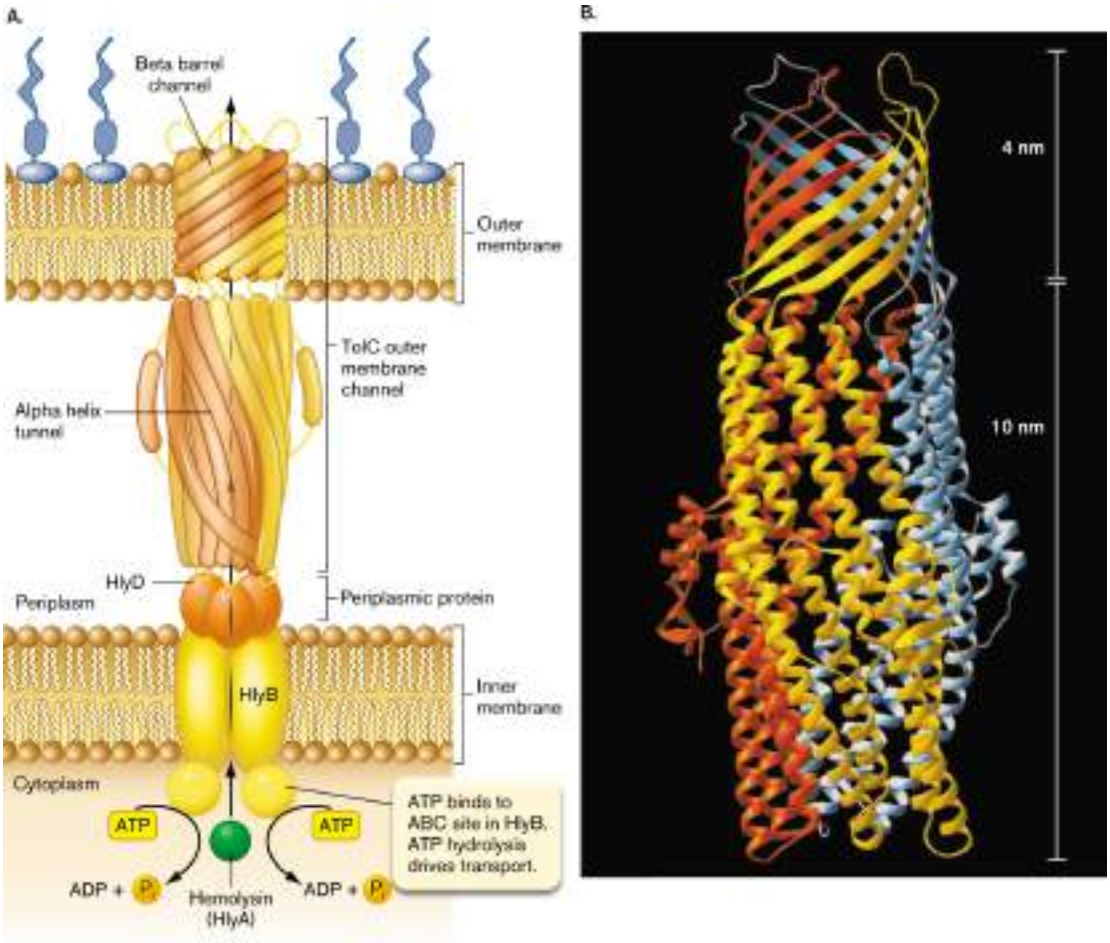


FIGURE 8.39 ■ Type I secretion: the Hly ABC transporter of *E. coli* . A. Hemolysin (HlyA) is transported directly from the cytoplasm into the extracellular medium through a multicomponent ABC transport system. The HlyB and HlyD proteins are dedicated to HlyA transport. TolC is shared with other transport systems. Not drawn to scale. **B.** Molecular model of TolC. The beta barrel channel spans the outer membrane, and the alpha helix tunnel extends into the periplasm. Three monomers (red, yellow, and blue) make up the channel.

Source: Part A modified from A. G. Moat et al. 2002. *Microbial Physiology* , 4th ed. Wiley-Liss.

VASSILIS KORONAKIS ET AL. 2000. *NATURE* **405:** 914–919

Journeys through the Outer Membrane

There are many reasons why bacteria need to export proteins completely out of the cell and into their surrounding environment. Some exported proteins digest extracellular peptides for carbon and nitrogen sources; others act as free-floating toxins that bind and kill host cells. Still others are injected directly into eukaryotic cells by pathogenic or symbiotic microbes to commandeer host metabolic processes. Nine secretion systems, identified as type I through type IX, have been discovered thus far to ship proteins out of the cell. Most are exclusive to Gram-negative microbes, which have a periplasmic space and an outer membrane in addition to the cytoplasmic membrane. A few secretion systems start with the Sec system just to get the protein into the periplasm, where dedicated outer membrane systems take over and complete export. Other systems provide nonstop service, delivering the protein directly from the cytoplasm to the extracellular space.

The diversity of system architecture is impressive. It is the result, in some instances, of selective evolutionary pressures appropriating established cellular processes (for example, pilus assembly). New systems evolve through the accidental duplication of one set of genes followed by random mutations that provide the duplicated set with a new function. We know this because the footprints of genetic divergence have been left behind in the DNA sequence. Type I secretion is described here. Other systems will be covered during the discussion of pathogenesis in Chapter 25 .

Type I Protein Secretion

Chapter 4 describes the family of ATP-binding cassette (ABC) influx transporters, whose signature is an amino acid motif that binds ATP. (The term “cassette” refers to a sequence of amino acids that is conserved in many proteins with similar functions.) In addition to ABC influx transporters, similar ABC transporters function in the opposite direction to export various toxins, proteases, and lipases, as well as antimicrobial drugs (multidrug efflux transporters). These

ABC transporters are the simplest of the protein secretion systems and make up what is called type I protein secretion (for example, the Hly system of *E. coli* that secretes hemolysin; **Fig. 8.39**). Type I systems all have three protein components, one of which contains an ATP-binding cassette. One component is an outer membrane channel, the second is an ABC protein at the inner membrane, and the third is a periplasmic protein lashed to the inner membrane.

Proteins secreted through type I systems never contact the periplasm, because they pass through a continuous channel that extends from the cytoplasm to the outer membrane. The inner membrane and periplasmic subunits are generally substrate specific, but numerous ABC export systems share the channel protein TolC, including multidrug efflux pumps that confer resistance. TolC is an intriguing protein composed of a beta barrel channel embedded in the outer membrane and an alpha helix tunnel spanning the periplasm (**Fig. 8.39B**). The type I transport system shown in **Figure 8.39A** exports a hemolysin (HlyA) from *E. coli* that lyses red blood cell membranes. HlyB and HlyD are the ABC and periplasmic components, respectively.

Some other protein secretion systems move proteins directly from the cytoplasm to the outside, similar to the type I system, whereas others pick up proteins deposited in the periplasm by the Sec system. They all play important roles in microbial pathogenesis and are more fully discussed in Chapter 25 .

To Summarize

- **Special protein export mechanisms** are used to move (secrete) proteins to the inner membrane, the periplasm, the outer membrane, and the extracellular surroundings.
- **N-terminal amino acid signal sequences** help target proteins to the membrane for secretion.
- **The general secretory system** , involving the SecYEG translocon, can move unfolded proteins to the inner

membrane or periplasm.

- **The signal recognition particle (SRP)** pauses the translation of a subset of proteins that will be placed into the membrane.
- **SecA protein** binds to unfolded proteins that will eventually end up in the periplasm and drives their secretion through the SecYEG translocon.
- **The twin arginine translocase (TAT)** can move a subset of fully folded proteins across the inner membrane and into the periplasm.
- **Type I secretion systems** are ATP-binding cassette (ABC) mechanisms that move certain secreted proteins directly from the cytoplasm to the extracellular environment.

Glossary

signal sequence
: A specific amino acid sequence on the amino terminus of proteins that directs them to the endoplasmic reticulum (of a eukaryote) or the cell membrane (of a prokaryote).

signal recognition particle (SRP)
: A receptor that recognizes the signal sequence of peptides undergoing translation. The complex attaches to the cell membrane of prokaryotes (or the rough endoplasmic reticulum of eukaryotes), where it docks the protein-ribosome complex to the membrane for protein membrane insertion or secretion.

eResearch Activity 8

How Do Cells of the Same Species Recognize Each Other for DNA Exchange?

Ultraviolet (UV) light creates thymidine dimers in DNA, which, left unchecked, can cause mutations in the chromosome. Homologous recombination (discussed in Chapter 9) can replace UV-damaged DNA if a second, undamaged copy is present in the cell. For the archaeon *Sulfolobus* , this undamaged copy can come from a neighbor cell. These cells possess special machines on their surface that exchange DNA, but only if the cells are close enough to touch. *Sulfolobus* responds to UV exposure by producing pili that extend beyond the cell and pull neighboring cells into close contact to let the DNA exchange process begin.

Because recombination-based DNA repair requires exchange of near identical, homologous DNA, it wouldn't be beneficial to use the pili to reel in cells of different species. How, then, can these UV-responsive pili target only cells of the same species when the organism lives in a highly diverse community? Sonja-Verena Albers and colleagues at the University of Freiburg in Germany found evidence that individual species of *Sulfolobus* have unique glycosylation signatures on cell-surface proteins, and that the pili of each species can recognize and bind to these glycosylations.

In a two-species mixed culture, cells of *S. acidocaldarius* and *S. tokodaii* remained isolated unless exposed to UV light. UV light induced cells of both species to aggregate, but these were strictly single-species aggregates (**Fig. ERA 8.1** , left column). The researchers then genetically engineered a strain of *S. acidocaldarius* to synthesize the pilus of *S. tokodaii* instead of its own. When exposed to UV light, this mutant was now able to aggregate with *S. tokodaii* cells (**Fig. ERA 8.1** , right column). This result confirmed that the pili made species-specific contact between neighboring

cells. From prior studies it was known that wild-type cells could still aggregate with cells missing their pili, which meant that the pili were attaching to something on the cell surface other than another copy of the pilus.

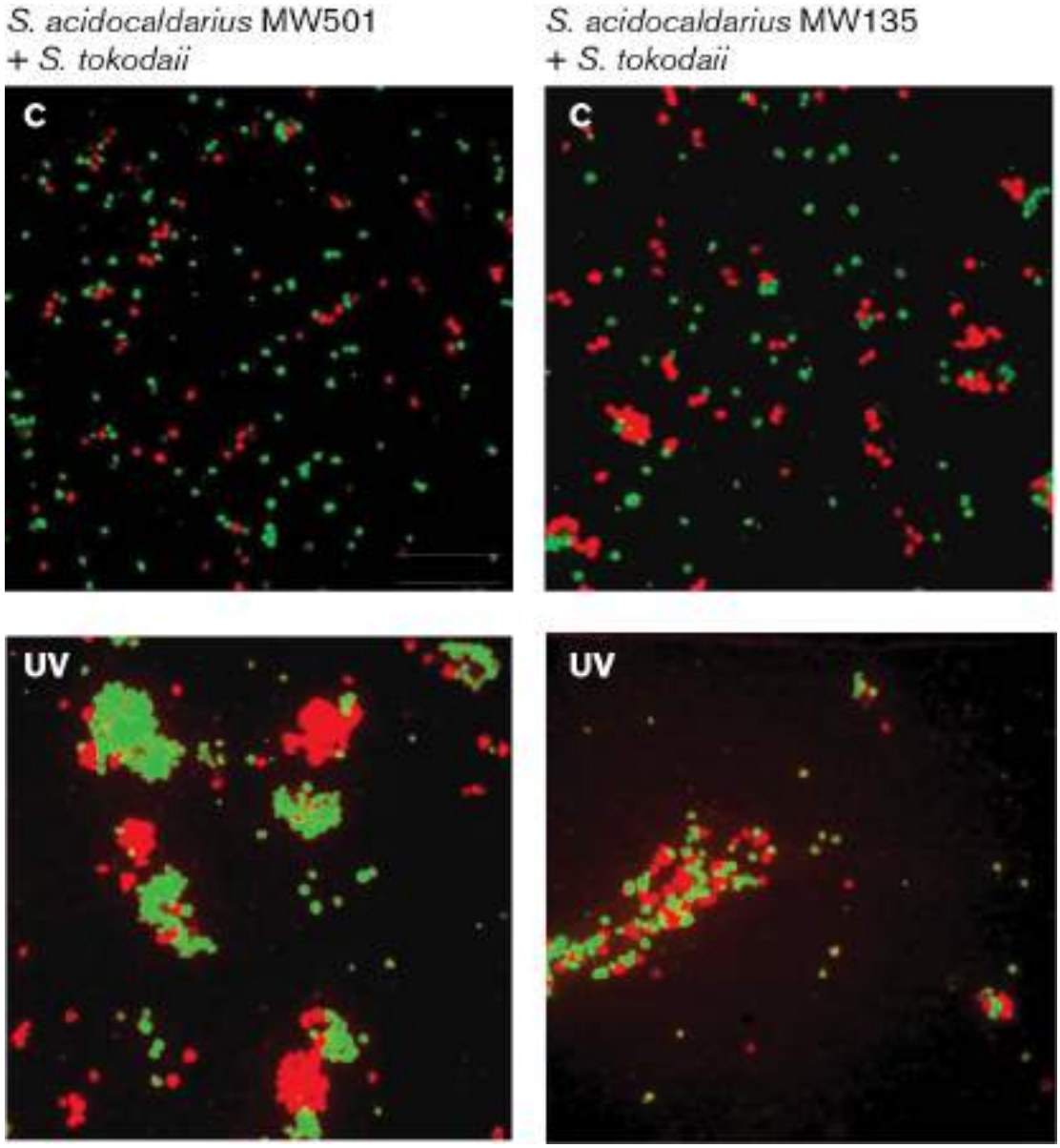


FIGURE ERA 8.1 ■ Pilus-mediated aggregation after exposure to UV light is species-specific in *Sulfolobus*. Strains of *S. acidocaldarius* (labeled green) were incubated with *S. tokodaii* (labeled red) with (UV) or without (C; control) exposure to UV light and visualized by epifluorescence

microscopy. *S. acidocaldarius* produced either its own pilus (left column; strain MW501) or the pilus from *S. tokodaii* (right column; strain MW135).

M. VAN WOLFEREN ET AL. 2020. *MBIO* . **11** :E03014–19

Sulfolobus cells have an S-layer on their cell exterior (discussed in Chapter 19), which is composed of SlaA and SlaB proteins. Both proteins were known to have posttranslational modifications in the form of glycosylations (see Section 8.4). The researchers hypothesized that the pili were binding to the glycosylation structures (glycans) on the S-layer proteins. They tested this hypothesis by treating cells of wild-type *S. acidocaldarius* with different types of monosaccharides that are common in glycans. If the pili bind to a specific monosaccharide within the glycan, then providing an overwhelming amount of that monosaccharide to the medium should saturate the binding sites of the pili and prevent them from attaching to the glycosylated S-layer proteins. The researchers found that, indeed, when mannose (but not glucose or *N* -acetylglucosamine) was added, the cells of wild-type *S. acidocaldarius* could no longer form aggregates when exposed to UV light (**Fig. ERA 8.2**). This result suggested that the pili of *S. acidocaldarius* bind to the mannose residues within the glycans attached to one or more of the surface proteins expressed on neighboring cells of the same species.

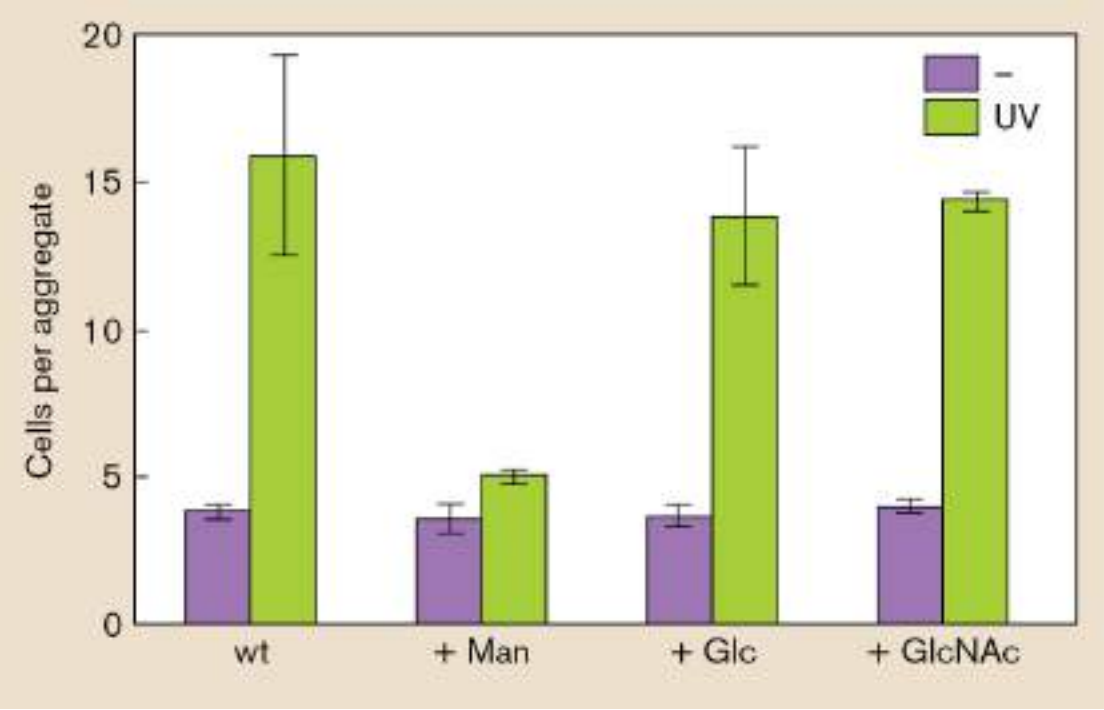


FIGURE ERA 8.2 ■ Mannose excess prevents aggregation. Wild-type (wt) cells of *S. acidocaldarius* have denser aggregates when exposed to ultraviolet light (UV) relative to the untreated control (−). UV light–induced aggregation by these wild-type cells was diminished in the presence of mannose (+Man) but not glucose (+Glc) or *N*-acetylglucosamine (+GlcNAc).

On the basis of the outcome of these first two experiments, one would predict that the glycosylation pattern on the S-layer proteins of *S. tokodaii* would be different from that of *S. acidocaldarius* . This prediction was confirmed when the researchers examined the glycan structures of these species by mass spectrometry. While the glycans of both species contained mannose, their linkages within the glycan were distinct (**Fig. ERA 8.3** , left).

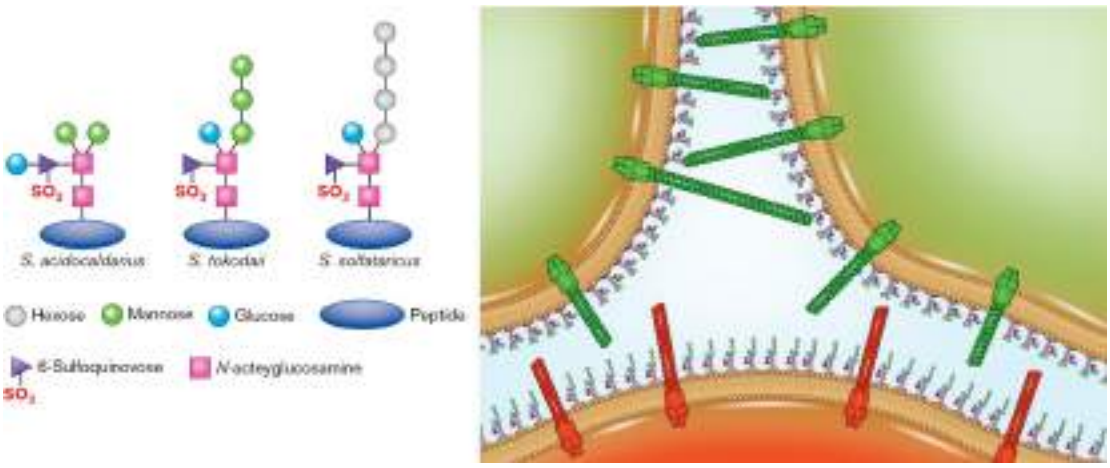


FIGURE ERA 8.3 ■ Posttranslational protein modification establishes strain-specific adhesion by pili. **Left:** Strain-specific glycosylations to the S-layer proteins exposed to the environment. **Right:** Model for the strain-specific binding of the *S. acidocaldarius* pili (green) to the glycans of other *S. acidocaldarius* cells but not to the glycans of *S. tokodaii* (red).

From the results of this study, the researchers proposed a model whereby each species of *Sulfolobus* produces unique glycosylation patterns on their S-layer proteins and unique UV-inducible pili that recognize only the glycosylations used by their own species (**Fig. ERA 8.3** , right). This specificity ultimately improves the ability of each species to survive UV exposure, as homologous recombination–based repair depends on high DNA-DNA similarity. This study highlights the importance of posttranslational modification of proteins, in this case as mediators of species recognition for a critical DNA repair mechanism.

Further Exploration

Conclusions from this work were supported by the results of genetic manipulation experiments in *S. acidocaldarius* (as in **Fig. ERA 8.1** , right column). Can you propose a genetic manipulation experiment

in *S. tokodaii* that would confirm species specificity in pilus binding to glycosylated S-layer proteins?

Source: **van Wolferen, Marleen, Asif Shajahan, Kristina Heinrich, Suzanne Brenzinger, Ian M. Black, et al.** 2020. Species-specific recognition of *Sulfolobales* mediated by UV-inducible pili and S-layer glycosylation patterns. *mBio* **11** :e03014–19.

CHAPTER REVIEW

Review Questions

1. What are some characteristics of an open reading frame?
2. What is a DNA sequence alignment, and what can it tell you?
3. What defines a promoter?
4. What are sigma factors, and what role do they play in gene expression?
5. Describe the three stages of transcription.
6. Explain the degeneracy of the genetic code. What is the wobble in codon-anticodon recognition?
7. Describe the stages of protein synthesis. Why is the ribosome classified as a ribozyme?
8. Discuss some antibiotics that affect transcription or translation.
9. What is coupled transcription and translation? Does it occur in eukaryotic cells?
10. How do bacterial cells release ribosomes that are stuck on damaged mRNA molecules lacking termination codons?
11. What kinds of posttranslational modifications can be made to proteins, and how do those modifications affect the proteins?
12. What can happen to misfolded proteins?
13. Why are only certain proteins secreted from the bacterial cell? What are some secretion mechanisms?

14. In what major way do proteins transported by the twin arginine translocase (TAT) differ from other exported proteins?
15. Compare protein degradation in eukaryotes and bacteria.

Thought Questions

1. The process of transcription generates positive supercoils in front of the polymerase as it moves along a DNA template. Why doesn't the DNA in front of the polymerase become so knotted that the polymerase can no longer separate the DNA strands?
2. Why do cells secrete certain proteins into their environments?
3. Type I protein secretion systems transport certain proteins from the cytoplasm of Gram-negative bacteria directly to the outside of the cell, across two membranes. How might the system "know" which proteins to transport?
4. Given a three-base codon, what is the maximum number of different amino acids that could be encoded by an organism's genetic code (keeping in mind at least one stop codon would be required), and what reasons might account for the far-smaller number encoded by the standard genetic code (20)?

Key Terms

- acceptor site (A site) (305)
- aminoacyl-tRNA synthetase (301)
- anticodon (300)

- catalytic RNA (296)
- chaperone (316)
- codon (299)
- consensus sequence (290)
- DNA-dependent RNA polymerase (289)
- exit site (E site) (305)
- expressed (288)
- heat-shock protein (HSP) (316)
- mass spectrometry (315)
- messenger RNA (mRNA) (288)
- N-terminal rule (317)
- open reading frame (ORF) (289)
- operon (289)
- peptidyltransferase (302)
- peptidyl-tRNA site (P site) (305)
- polysome (309)
- promoter (289)
- release factor (307)
- Rho factor (293)
- ribosomal RNA (rRNA) (296)
- ribosome (302)
- ribosome-binding site (304)
- ribozyme (296 , 302)
- RNA polymerase (288 , 289)
- Shine-Dalgarno sequence (304)
- sigma factor (289)
- signal recognition particle (SRP) (319)
- signal sequence (319)
- small RNA (sRNA) (296)
- start codon (304)
- stop codon (299)
- template strand (289)
- transcript (289)

- transcription (288)
- transfer messenger RNA (tmRNA) (296)
- transfer RNA (tRNA) (296)
- translation (288)
- translocation (306)

Recommended Reading

Bakshi, Somenath, Heejun Choi, Jagannath Mondal, and James C. Weisshaar. 2014. Time-dependent effects of transcription- and translation-halting drugs on the spatial distributions of the *Escherichia coli* chromosome and ribosomes. *Molecular Microbiology* **94** :871–887.

Bastos, P. A. D., J. P. da Costa, and R. Vitorino. 2017. A glimpse into the modulation of post-translational modifications of human-colonizing bacteria. *Journal of Proteomics* **152** :254–275.

Brandt, Florian, Stephanie A. Etchells, Julio O. Ortiz, Adrian H. Elcock, F. Ulrich Hartl, et al. 2009. The native 3D organization of bacterial polysomes. *Cell* **136** :261–271.

Burger, Adelle, Chris Whiteley, and Aileen Boshoff. 2011. Current perspectives of the *Escherichia coli* RNA degradosome. *Biotechnology Letters* **33** :2337–2350.

Costa, Tiago R., Catarina Felisberto-Rodrigues, Amit Meir, Marie S. Prevost, Adam Redzej, et al. 2015. Secretion systems in Gram-negative bacteria: Structural and mechanistic insights. *Nature Reviews. Microbiology* **13** :343–359.

Gehring, A. M., J. E. Walker, and T. J. Santangelo. 2016. Transcription regulation in Archaea. *Journal of Bacteriology* **198** :1906–1917.

Hui, Monica P., Patricia L. Foley, and Joel G. Belasco. 2014. Messenger RNA degradation in bacterial cells. *Annual Review of Genetics* **48** :537–559.

Johnson, Grace E., Jean Benôit Lalanne, Michelle L. Peters, and Gene-Wei Li. 2020. Functionally uncoupled transcription-translation in *Bacillus subtilis. Nature* **585** :134–138.

Keiler, Kenneth C. 2015. Mechanisms of ribosome rescue in bacteria. *Nature Reviews. Microbiology* **13** : 285–297.

Ling, J., P. O'Donoghue, and D. Soll. 2015. Genetic code flexibility in microorganisms: Novel mechanisms and impact on physiology. *Nature Reviews. Microbiology* **13** :707–721.

Murakami, Katsuhiko S., Shoko Masuda, Elizabeth A. Campbell, Oriana Muzzin, and Seth Darst. 2002. Structural basis of transcription initiation: RNA polymerase holoenzyme-DNA complex. *Science* **296** : 1285–1290.

Petrov, Anton S., Chad R. Bernier, Chiaolong Hsiao, Ashlyn M. Norris, Nicholas A. Kovacs, et al. 2014. Evolution of the ribosome at atomic resolution. *Proceedings of the National Academy of Sciences USA* **111** :10251–10256.

Preissler, Steffen, and Elke Deuerling. 2012. Ribosome associated chaperones as key players in proteostasis. *Trends in Biochemical Sciences* **37** :274–283.

Proshkin, Sergey, A. Rachid Rahmouni, Alexander Mironov, and Evgeny Nudler. 2010. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* **328** :504–508.

Saier, Milton H. 2019. Understanding the genetic code. *Journal of Bacteriology* **201** :e00091–19.

Schulze, Ryan J., Joanna Komar, Mathieu Botte, William J. Allen, Sarah Whitehouse, et al. 2014. Membrane protein insertion and proton-motive-force-dependent secretion through the bacterial holo-translocon SecYEG-SecDF-YajC-YidC. *Proceedings of the National Academy of Sciences USA* **111** :4844–4849.

Smets, Dries, Maria S. Loos, Spyridoula Karamanou, and Anastassios Economou. 2019. Protein transport across the

bacterial plasma membrane by the Sec pathway. *Protein Journal* **38** :262–273.

Vega, Luis A., Gary C. Port, and Michael G. Caparon. 2013. An association between peptidoglycan synthesis and organization of the *Streptococcus pyogenes* ExPortal. *mBio* **4** :e00485–13.

Washburn, Robert S., and Max E. Gottesman. 2015. Regulation of transcription elongation and termination. *Biomolecules* **5** :1063–1078.

Glossary

expression
: Synthesis of the products encoded by a gene. This includes RNA and, for mRNAs that are translated, protein.

transcription
: The synthesis of RNA complementary to a DNA template.

messenger RNA (mRNA)
: An RNA molecule that encodes a protein.

DNA-dependent RNA polymerase
: See **RNA polymerase** .

promoter
: A noncoding DNA regulatory region immediately upstream of a structural gene that is needed for transcription initiation.

operon
: A collection of genes that are in tandem on a chromosome and are transcribed into a single RNA.

open reading frame (ORF)
: A DNA sequence predicted to encode a protein.

RNA polymerase
: Also called *DNA-dependent RNA polymerase* . An enzyme that produces an RNA complementary to a template DNA strand.

transcript
: An RNA copy of a DNA template.

template strand

A DNA strand (or an RNA strand in some viruses) that is used as a template for the synthesis of mRNA.

sigma factor

A protein needed to bind RNA polymerase for the initiation of transcription in bacteria.

consensus sequence

A sequence of nucleotides or amino acids with a common function at many nucleic acid or protein positions. Consists of the base pair or amino acid most frequently found at each position in the sequence.

Rho factor

A bacterial protein involved in terminating transcription.

ribosomal RNA (rRNA)

An RNA molecule that includes the scaffolding and catalytic components of ribosomes.

transfer RNA (tRNA)

An RNA that carries an amino acid to the ribosome. The anticodon on the tRNA base-pairs with the codon on the mRNA.

small RNA (sRNA)

A non-protein-coding regulatory RNA molecule that modulates translation or mRNA stability.

transfer messenger RNA (tmRNA)

A molecule resembling both tRNA and mRNA that rescues ribosomes stalled on damaged mRNAs lacking a stop codon.

catalytic RNA

Also called *ribozyme* . An RNA molecule that is capable of catalyzing reactions.

ribozyme

See **catalytic RNA** .

codon

A set of three nucleotides that encodes a particular amino acid.

stop codon

One of three codons (UAA, UAG, UGA) that do not encode an amino acid, and thus trigger the end of translation.

anticodon

A set of three nucleotides in the middle loop of a tRNA that base-pairs with a codon in mRNA.

aminoacyl-tRNA synthetase

An enzyme that condenses a specific amino acid with the 3′ OH group of the correct tRNA, thereby charging the tRNA.

ribosome

A large enzyme, composed of RNA and protein subunits, that translates mRNA into protein.

peptidyltransferase

A ribozyme that catalyzes the formation of peptide bonds.

start codon

A codon (usually AUG, encoding methionine) that signals the first amino acid of a protein.

ribosome-binding site

Also called *Shine-Dalgarno sequence* . In bacteria, a stretch of nucleotides upstream of the start codon in an mRNA that hybridizes to the 16S rRNA of the ribosome, correctly positioning the mRNA for translation.

Shine-Dalgarno sequence

See **ribosome-binding site** .

acceptor site (A site)

The region of a ribosome that binds an incoming charged tRNA.

peptidyl-tRNA site (P site)

The region of a ribosome that contains the growing protein attached to a tRNA.

exit site (E site)

The region of a ribosome that holds the uncharged, exiting tRNA.

translocation

The energy-dependent movement of a ribosome to the next triplet codon along an mRNA.

release factor

A molecule that enters a ribosome A site containing an mRNA stop codon and initiates protein cleavage from the tRNA.

polysome

A cell structure consisting of multiple ribosomes performing translation on the same mRNA molecule.

mass spectrometry

An analytical technique that measures the mass of molecules. Molecules are ionized and sorted according to their mass-to-charge (m / z) ratio.

chaperone or chaperonin

A protein that helps other proteins fold into their correct tertiary structure.

heat-shock protein (HSP)

A chaperone protein whose synthesis is induced by high-temperature stress.

N-terminal rule

The tendency of the N-terminal amino acid of a protein to influence protein stability.

signal sequence

A specific amino acid sequence on the amino terminus of proteins that directs them to the endoplasmic reticulum (of a eukaryote) or the cell membrane (of a prokaryote).

signal recognition particle (SRP)

A receptor that recognizes the signal sequence of peptides undergoing translation. The complex attaches to the cell membrane of prokaryotes (or the rough endoplasmic reticulum of eukaryotes), where it docks the protein-ribosome complex to the membrane for protein membrane insertion or secretion.

translation

The ribosomal synthesis of proteins based on triplet codons present in mRNA.