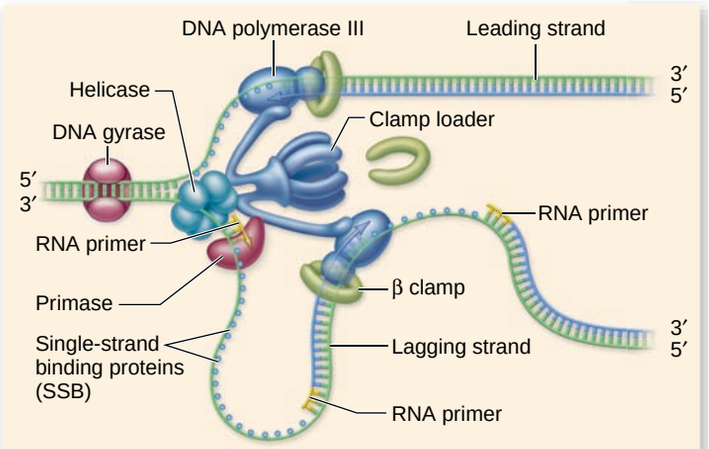
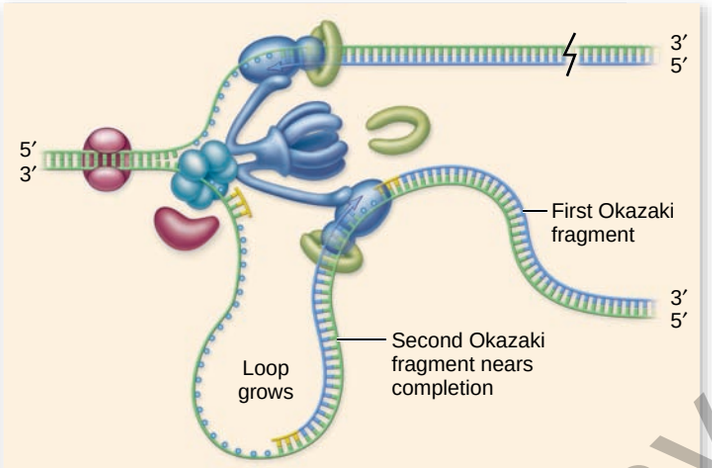
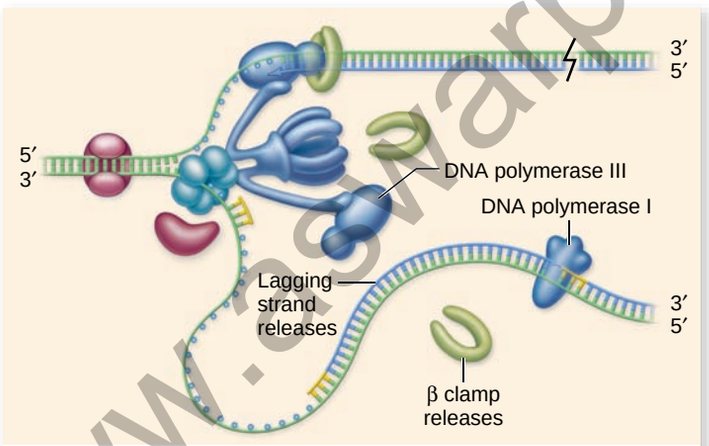


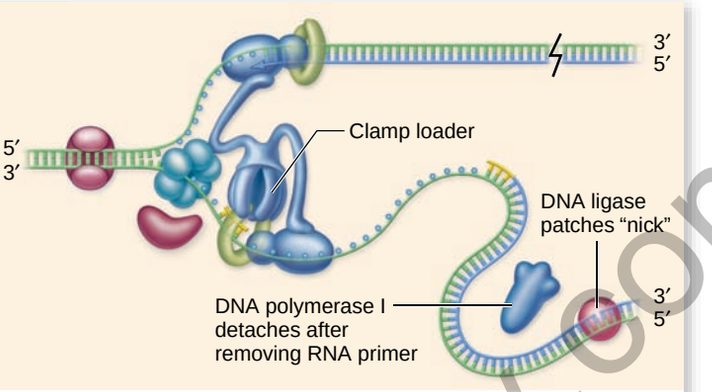
1. A DNA polymerase III enzyme is active on each strand. Primase synthesizes new primers for the lagging strand.



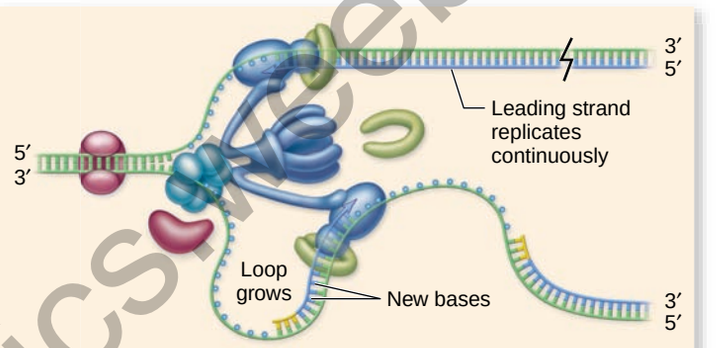
2. The "loop" in the lagging-strand template allows replication to occur 5′-to-3′ on both strands, with the complex moving to the left.



3. When the polymerase III on the lagging strand hits the previously synthesized fragment, it releases the β clamp and the template strand. DNA polymerase I attaches to remove the primer.



4. The clamp loader attaches the β clamp and transfers this to polymerase III, creating a new loop in the lagging-strand template. DNA ligase joins the fragments after DNA polymerase I removes the primers.



5. After the β clamp is loaded, the DNA polymerase III on the lagging strand adds bases to the next Okazaki fragment.

Figure 14.19 DNA synthesis by the replisome. The semidiscontinuous synthesis of DNA is illustrated in stages using the model from figure 14.18.

Even given the difficulties with lagging-strand synthesis, the two Pol III enzymes in the replisome are active on both leading and lagging strands simultaneously. How can the two strands be synthesized in the same direction when the strands are antiparallel? The model first proposed, still with us in some form, involves a loop formed in the lagging strand, so that the polymerases can move in the same direction (see figure 14.18). Current evidence also indicates that this replication complex is probably stationary, with the DNA strand moving through it like thread in a sewing machine, rather than the complex moving along the DNA strands. This stationary complex also pushes the newly synthesized DNA outward, which may aid in chromosome segregation. This process is summarized in figure 14.19.

Learning Outcomes Review 14.4

***E. coli* has three DNA polymerases: DNA Pol I, II, and III. Synthesis on one strand is discontinuous because DNA is antiparallel, and polymerases only synthesize in the 5′-to-3′ direction. Replication occurs at the replication fork, where the two strands are separated. Assembled here is a massive complex, the replisome, containing DNA polymerase III, primase, helicase, and other proteins. The lagging strand requires DNA polymerase I to remove the primers and replace them with DNA, and ligase to join Okazaki fragments.**

- ***How are the nuclease functions of the different polymerases used during replication?***

14.5 Eukaryotic Replication

Learning Outcomes

1. *Compare eukaryotic replication with prokaryotic.*
2. *Explain the function of telomeres.*
3. *Identify the link between telomerase and cell division.*

Eukaryotic replication is complicated by two main factors: the larger amount of DNA organized into multiple chromosomes, and the linear structure of the chromosomes. This process requires new enzymatic activities only for dealing with the ends of chromosomes; otherwise the basic enzymology is the same.

Eukaryotic replication requires multiple origins

The sheer amount of DNA and how it is packaged constitute a problem for eukaryotes (figure 14.20). Eukaryotes usually have multiple chromosomes that are each larger than the *E. coli* chromosome. If only a single unique origin existed for each chromosome, the length of time necessary for replication would be prohibitive. This problem is solved by the use of multiple origins of replication for each chromosome, resulting in multiple *replicons*.

Figure 14.20 DNA of a single human chromosome. This chromosome has been relieved of most of its packaging proteins, leaving the DNA in its native form. The residual protein scaffolding appears as the dark material in the lower part of the micrograph.

The origins are not as sequence-specific as *oriC*, and their recognition seems to depend on chromatin structure as well as on sequence. The number of origins used can also be adjusted during the course of development, so that early on, when cell divisions need to be rapid, more origins are activated. Each origin must be used only once per cell cycle.

The enzymology of eukaryotic replication is more complex

The replication machinery of eukaryotes is similar to that found in bacteria, but it is larger and more complex. The initiation phase of replication requires more factors to assemble both helicase and primase complexes onto the template, then load the polymerase with its sliding clamp unit.

The eukaryotic primase is interesting in that it is a complex of both an RNA polymerase and a DNA polymerase. It first makes short RNA primers, then extends these with DNA to produce the final primer. The reason for this added complexity is unclear.

The main replication polymerase itself is also a complex of two different enzymes that work together. One is called *DNA polymerase epsilon* (pol ε) and the other *DNA polymerase delta* (pol δ). The sliding clamp subunit that allows the enzyme complex to stay attached to the template is called PCNA (for proliferating cell nuclear antigen). This unusual name reflects the fact that PCNA was first identified as an antibody-inducing protein in proliferating (dividing) cells. The PCNA sliding clamp forms a trimer, but this structure is similar to the β subunit sliding clamp. The clamp loader is also similar to the bacterial structure. Despite the additional complexity, the action of the replisome is similar to that described earlier for *E. coli*, and the replication fork has essentially the same components.

Archaeal replication proteins are similar to eukaryotic proteins

Despite their lack of a membrane-bounded nucleus, Archaeal replication proteins are more similar to eukaryotes than to bacterial. The main replication polymerase is most similar to eukaryotic pol δ, and the sliding clamp is similar to the PCNA protein. The clamp loading complex is also more similar to eukaryotic than bacterial. The most interesting conclusion from all of these data are that all three domains of life have similar functions involved in replicating chromosomes. All three domains assemble similar protein complexes with clamp loader, sliding clamp, two polymerases, helicase, and primase at the replication fork.

Linear chromosomes have specialized ends

The specialized structures found on the ends of eukaryotic chromosomes are called **telomeres.** These structures protect

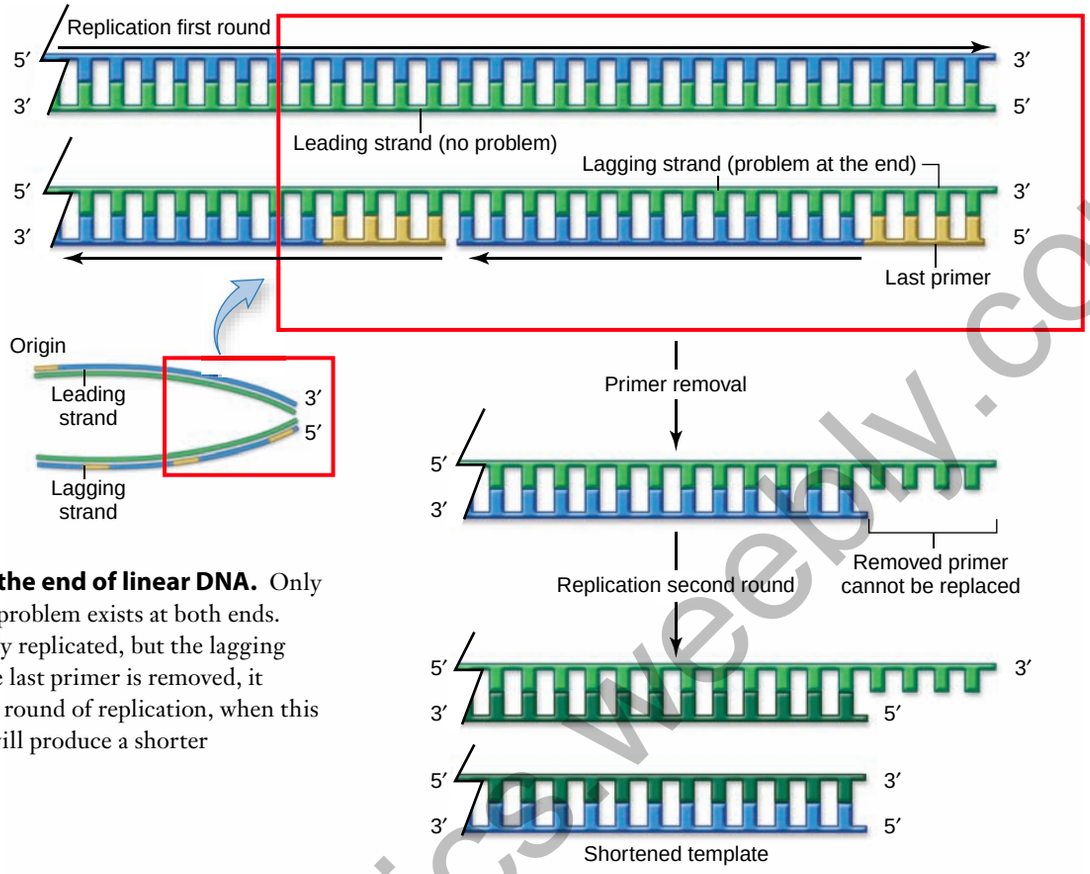


Figure 14.21 Replication of the end of linear DNA. Only one end is shown for simplicity; the problem exists at both ends. The leading strand can be completely replicated, but the lagging strand cannot be finished. When the last primer is removed, it cannot be replaced. During the next round of replication, when this shortened template is replicated, it will produce a shorter chromosome.

the ends of chromosomes from nucleases and maintain the integrity of linear chromosomes. These telomeres are composed of specific DNA sequences, but they are not made by the replication complex.

Replicating ends

The very structure of a linear chromosome causes a cell problems in replicating the ends. The directionality of polymerases, combined with their requirement for a primer, create this problem.

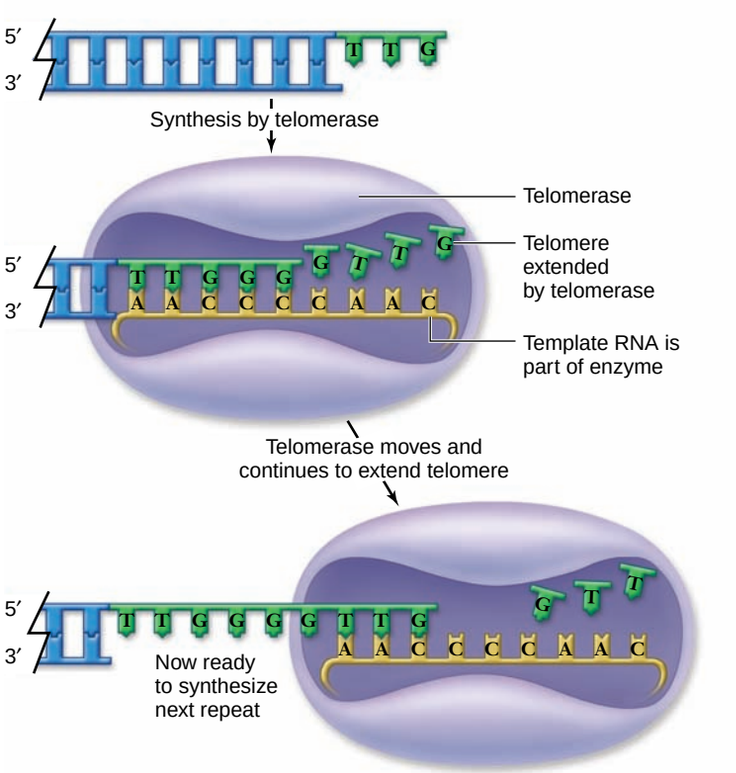
Consider a simple linear molecule like the one in figure 14.21. Replication of one end of each template strand is simple, namely the 5′ end of the leading-strand template. When the polymerase reaches this end, synthesizing in the 5′-to-3′ direction, it eventually runs out of template and is finished.

But on the other strand's end, the 3′ end of the lagging strand, removal of the last primer on this end leaves a gap. This gap cannot be primed, meaning that the polymerase complex cannot finish this end properly. The result would be a gradual shortening of chromosomes with each round of cell division (see figure 14.21).

The action of telomerase

When the sequence of telomeres was determined, they were found to be composed of short repeated sequences of DNA. This repeating nature is easily explained by their synthesis. They are made by an enzyme called **telomerase,** which uses an internal RNA as a template and not the DNA itself (figure 14.22).

Figure 14.22 Action of telomerase. Telomerase contains an internal RNA that the enzyme uses as a template to extend the DNA of the chromosome end. Multiple rounds of synthesis by telomerase produce repeated sequences. This single strand is completed by normal synthesis using it as a template (not shown).



The use of the internal RNA template allows short stretches of DNA to be synthesized, composed of repeated nucleotide sequences complementary to the RNA of the enzyme. The other strand of these repeated units is synthesized by the usual action of the replication machinery copying the strand made by telomerase.

Telomerase, aging, and cancer

A gradual shortening of the ends of chromosomes occurs in the absence of telomerase activity. During embryonic and childhood development in humans, telomerase activity is high, but it is low in most somatic cells of the adult. The exceptions are cells that must divide as part of their function, such as lymphocytes. The activity of telomerase in somatic cells is kept low by preventing the expression of the gene encoding this enzyme.

Evidence for the shortening of chromosomes in the absence of telomerase was obtained by producing mice with no telomerase activity. These mice appear to be normal for up to six generations, but they show steadily decreasing telomere length that eventually leads to nonviable offspring.

This finding indicates a relationship between cell senescence (aging) and telomere length. Normal cells undergo only a specified number of divisions when grown in culture. This limit is at least partially based on telomere length.

Support for the relationship between senescence and telomere length comes from experiments in which telomerase was introduced into fibroblasts in culture. These cells have their lifespan increased relative to controls that have no added telomerase. Interestingly, these cells do not show the hallmarks of malignant cells, indicating that activation of telomerase alone does not make cells malignant.

A relationship has been found, however, between telomerase and cancer. Cancer cells do continue to divide indefinitely, and this would not be possible if their chromosomes were being continually shortened. Cancer cells generally show activation of telomerase, which allows them to maintain telomere length; but this is clearly only one aspect of conditions that allow them to escape normal growth controls.

Inquiry question

How does the structure of eukaryotic genomes affect replication? Does this introduce problems that are not faced by prokaryotes?

Learning Outcomes Review 14.5

Eukarotic replication is complicated by a large amount of DNA organized into chromosomes, and by the linear nature of chromosomes. Eukaryotes replicate a large amount of DNA in a short time by using multiple origins of replication. Linear chromosomes end in telomeres, and the length of telomeres is correlated with the ability of cells to divide. The enzyme telomerase synthesizes the telomeres. Cancer cells show activation of telomerase, which extends the ability of the cells to divide.

- ***What might be the result of abnormal shortening of telomeres or a lack of telomerase activity?***

14.6 *DNA Repair*

Learning Outcomes

1. *Explain why DNA repair is critical for cells.*
2. *Describe the different forms of DNA repair.*

As you learned earlier, many DNA polymerases have 3′-to-5′exonuclease activity that allows "proofreading" of added bases. This action increases the accuracy of replication, but errors still occur. Without error correction mechanisms, cells would accumulate errors at an unacceptable rate, leading to high levels of deleterious or lethal mutations. A balance must exist between the introduction of new variation by mutation, and the effects of deleterious mutations on the individual.

Cells are constantly exposed to DNA-damaging agents

In addition to errors in DNA replication, cells are constantly exposed to agents that can damage DNA. These agents include radiation, such as UV light and X-rays, and chemicals in the environment. Agents that damage DNA can lead to mutations, and any agent that increases the number of mutations above background levels is called a **mutagen.**

The number of potentially mutagenic agents that organisms encounter is huge. Sunlight itself includes radiation in the UV range and is thus mutagenic. Ozone normally screens out much of the harmful UV radiation in sunlight, but some remains. The relationship between sunlight and mutations is shown clearly by the increase in skin cancer in regions of the southern hemisphere that are underneath a seasonal "ozone hole."

Organisms also may encounter mutagens in their diet in the form of either contaminants in food or natural plant products that can damage DNA. When a simple test was designed to detect mutagens, screening of possible sources indicated an amazing diversity of mutagens in the environment and in natural sources. As a result, consumer products are now screened to reduce the load of mutagens we are exposed to, but we cannot escape natural sources.

DNA repair restores damaged DNA

Cells cannot escape exposure to mutagens, but systems have evolved that enable cells to repair some damage. These DNA repair systems are vital to continued existence, whether a cell is a free-living, single-celled organism or part of a complex multicellular organism.

The importance of DNA repair is indicated by the multiplicity of repair systems that have been discovered and characterized. All cells that have been examined show multiple pathways for repairing damaged DNA and for reversing errors that occur during replication. These systems are not perfect, but they do reduce the mutational load on organisms to an acceptable level. In the rest of this section, we illustrate the action of DNA repair by concentrating on two examples drawn from these multiple repair pathways.

Repair can be either specific or nonspecific

DNA repair falls into two general categories: specific and nonspecific. Specific repair systems target a single kind of lesion in DNA and repair only that damage. Nonspecific forms of repair use a single mechanism to repair multiple kinds of lesions in DNA.

Photorepair: A specific repair mechanism

Photorepair is specific for one particular form of damage caused by UV light, namely the *thymine dimer*. Thymine dimers are formed by a photochemical reaction of UV light and adjacent thymine bases in DNA. The UV radiation causes the thymines to react, covalently linking them together: a thymine dimer (figure 14.23).

Repair of these thymine dimers can be accomplished by multiple pathways, including photorepair. In photorepair, an enzyme called a *photolyase* absorbs light in the visible range and uses this energy to cleave the thymine dimer. This action restores the two thymines to their original state (see figure 14.23). It is interesting that sunlight in the UV range can cause this damage, and sunlight in the visible range can be used to repair the damage. Photorepair does not occur in cells deprived of visible light.

The photolyase enzyme has been found in many different species, ranging from bacteria, to single-celled eukaryotes, to humans. The ubiquitous nature of this enzyme illustrates the importance of this form of repair. For as long as cells have existed on Earth, they have been exposed to UV light and its potential to damage DNA.

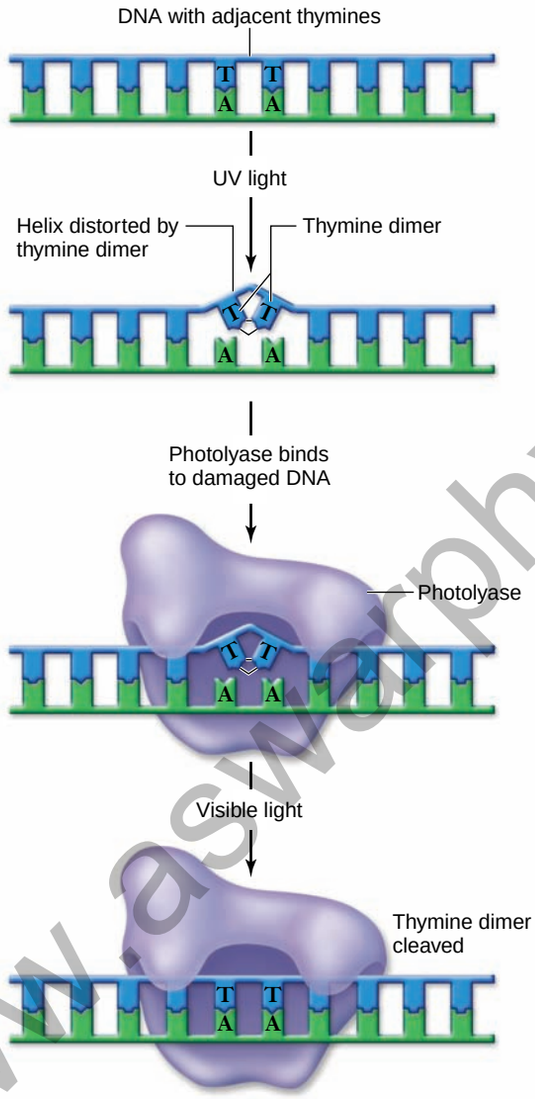


Figure 14.23 Repair of thymine dimer by photorepair. UV light can catalyze a photochemical reaction to form a covalent bond between two adjacent thymines, thereby creating a thymine dimer. A photolyase enzyme recognizes the damage and binds to the thymine dimer. The enzyme absorbs visible light and uses the energy to cleave the thymine dimer.

Excision repair: A nonspecific repair mechanism

A common form of nonspecific repair is **excision repair.** In this pathway, a damaged region is removed, or excised, and is then replaced by DNA synthesis (figure 14.24). In *E. coli*, this action is accomplished by proteins encoded by the *uvr A*, *B*, and

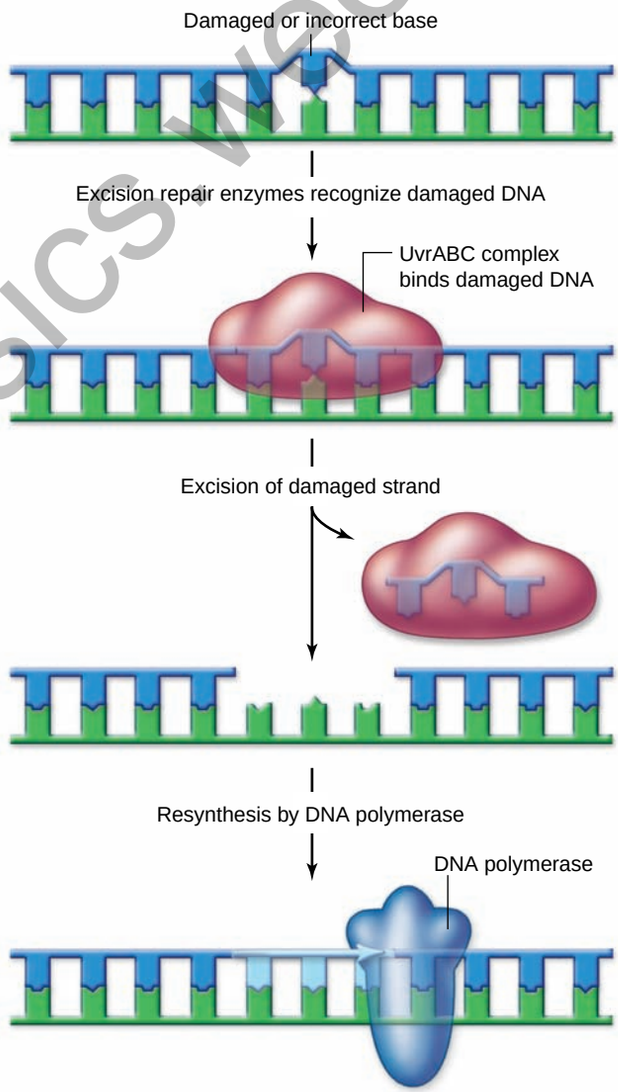


Figure 14.24 Repair of damaged DNA by excision repair. Damaged DNA is recognized by the uvr complex, which binds to the damaged region and removes it. Synthesis by DNA polymerase replaces the damaged region. DNA ligase finishes the process (not shown).

C genes. Although these genes were identified based on mutations that increased sensitivity of the cell to UV light (hence the "uvr" in their names), their proteins can act on damage due to other mutagens.

Excision repair follows three steps: (1) recognition of damage, (2) removal of the damaged region, and (3) resynthesis using the information on the undamaged strand as a template (see figure 14.24). Recognition and excision are accomplished by the UvrABC complex. The UvrABC complex binds to damaged DNA and then cleaves a single strand on either side of the damage, removing it. In the synthesis stage, DNA pol I or pol II replaces the damaged DNA. This restores the original information in the damaged strand by using the information in the complementary strand.

Other repair pathways

Cells have other forms of nonspecific repair, and these fall into two categories: error-free and error-prone. It may seem strange to have an error-prone pathway, but it can be thought of as a last-ditch effort to save a cell that has been exposed to such massive damage that it has overwhelmed the error-free systems. In fact, this system in *E. coli* is part of what is called the "SOS response."

Cells can also repair damage that produces breaks in DNA. These systems use enzymes related to those that are involved in recombination during meiosis (see chapter 11). It is thought that recombination uses enzymes that originally evolved for DNA repair.

The number of different systems and the wide spectrum of damage that can be repaired illustrate the importance of maintaining the integrity of the genome. Accurate replication of the genome is useless if a cell cannot reverse errors that can occur during this process or repair damage due to environmental causes.

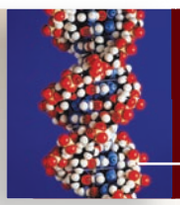
Inquiry question

Cells are constantly exposed to DNA-damaging agents, ranging from UV light to by-products of oxidative metabolism. How does the cell deal with this, and what would happen if the cell had no way of dealing with this?

Learning Outcomes Review 14.6

The ability to repair DNA is critical because of replication errors and the constant presence of damaging agents that can cause mutation. Cells have multiple repair pathways; some of these systems are specific for a single type of damage, such as photorepair that reverses thymine dimers caused by UV light. Other systems are nonspecific, such as excision repair that removes and replaces damaged regions.

- ***Could a cell survive with no form of DNA repair?***



Chapter Review

14.1 The Nature of the Genetic Material

Griffith finds that bacterial cells can be transformed.

Nonvirulent *S. pneumoniae* could take up an unknown substance from a virulent strain and become virulent.

Avery, MacLeod, and McCarty identify the transforming principle.

The transforming substance could be inactivated by DNA-digesting enzymes, but not by protein-digesting enzymes.

Hershey and Chase demonstrate that phage genetic material is DNA.

Radioactive labeling showed that the infectious agent of phage is its DNA, and not its protein.

14.2 DNA Structure

DNA's components were known, but its three-dimensional structure was a mystery.

The nucleotide building blocks for DNA contain deoxyribose and the bases adenine (A), guanine (G), cytosine (C), and thymine (T). Phosphodiester bonds are formed between the 5′ phosphate of one nucleotide and the 3′ hydroxyl of another nucleotide (see figure 14.4).

Chargaff, Franklin, and Wilkins obtained some structural evidence.

Chargaff found equal amounts of adenine and thymine, and of cytosine and guanine, in DNA. The bases exist primarily in keto and enol forms that exhibit hydrogen bonding. X-ray diffraction studies by Franklin and Wilkins indicated that DNA had a helical structure.

The Watson-Crick model fits the available evidence (see figures 14.8 and 14.9).

DNA consists of two antiparallel polynucleotide strands wrapped about a common helical axis. These strands are held together by hydrogen bonds forming specific base pairs (A/T and G/C). The two strands are complementary; one strand can specify the other.

14.3 Basic Characteristics of DNA Replication

Meselson and Stahl demonstrate the semiconservative mechanism (see figure 14.11).

Semiconservative replication uses each strand of a DNA molecule to specify the synthesis of a new strand. Meselson and Stahl showed this by using a heavy isotope of nitrogen and separating the replication products. Replication produces two new molecules each composed of one new strand and one old strand.

DNA replication requires a template, nucleotides, and a polymerase enzyme.

All new DNA molecules are produced by DNA polymerase copying a template. All known polymerases synthesize new DNA in the 5′-to-3′ direction. These enzymes also require a primer. The building blocks used in replication are deoxynucleotide triphosphates with high-energy bonds; they do not require any additional energy.

14.4 Prokaryotic Replication

Prokaryotic replication starts at a single origin.

The *E. coli* origin has AT-rich sequences that are easily opened. The chromosome and its origin form a replicon.

E. coli *has at least three different DNA polymerases.*

Some DNA polymerases can also degrade DNA from one end, called exonuclease activity. Pol I, II, and III all have 3′-to-5′ exonuclease activity that can remove mispaired bases. Pol I can remove bases in the 5′-to-3′ direction, important to removing RNA primers.

Unwinding DNA requires energy and causes torsional strain.

DNA helicase uses energy from ATP to unwind DNA. The torsional strain introduced is removed by the enzyme DNA gyrase.

Replication is semidiscontinuous.

Replication is discontinuous on one strand (see figure 14.15). The continuous strand is called the leading strand, and the discontinuous strand is called the lagging strand.

Synthesis occurs at the replication fork.

The partial opening of a DNA strand forms two single-stranded regions called the replication fork. At the fork, synthesis on the leading strand requires a single primer, and the polymerase stays attached to the template because of the β subunit that acts as a sliding clamp. On the lagging strand, DNA primase adds primers periodically, and DNA Pol III synthesizes the Okazaki fragments. DNA Pol I removes primer segments, and DNA ligase joins the fragments.

The replisome contains all the necessary enzymes for replication.

The replisome consists of two copies of Pol III, DNA primase, DNA helicase, and a number of accessory proteins. It moves in one direction by creating a loop in the lagging strand, allowing the antiparallel template strands to be copied in the same direction (see figures 14.18 and 14.19).

14.5 Eukaryotic Replication

Eukaryotic replication requires multiple origins.

The sheer size and organization of eukaryotic chromosomes requires multiple origins of replication to be able to replicate DNA in the time available in S phase.

The enzymology of eukaryotic replication is more complex.

The eukaryotic primase synthesizes a short stretch of RNA and then switches to making DNA. This primer is extended by the main replication polymerase, which is a complex of two enzymes. The sliding clamp subunit was originally identified as protein produced by proliferating cells and is called PCNA.

Archaeal replication proteins are similar to eukaryotic proteins.

The replication proteins of archaea, including the sliding clamp, clamp loader, and DNA polymerases, are more similar to those of eukaryotes than to prokaryotes.

Linear chromosomes have specialized ends.

The ends of linear chromosomes are called telomeres. They are made by telomerase, not by the replication complex. Telomerase contains an internal RNA that acts as a template to extend the DNA of the chromosome end. Adult cells lack telomerase activity, and telomere shortening correlates with senescence.

14.6 DNA Repair

Cells are constantly exposed to DNA-damaging agents.

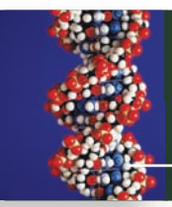
Errors from replication and damage induced by agents such as UV light and chemical mutagens can lead to mutations.

DNA repair restores damaged DNA.

Without repair mechanisms, cells would accumulate mutations until inviability occurred.

Repair can be either specific or nonspecific.

The enzyme photolyase uses energy from visible light to cleave thymine dimers caused by UV light. Excision repair is nonspecific. In prokaryotes, the uvr system can remove a damaged region of DNA.



Review Questions

UNDERSTAND

1. What was the key finding from Griffith's experiments using live and heat-killed pathogenic bacteria?
   a. Bacteria with a smooth coat could kill mice.
   b. Bacteria with a rough coat are not lethal.
   c. DNA is the genetic material.
   d. Genetic material can be transferred from dead to live bacteria.
2. Which of the following is *not* a component of DNA?
   a. The pyrimidine uracil
   b. Five-carbon sugars
   c. The purine adenine
   d. Phosphate groups
3. Chargaff studied the composition of DNA from different sources and found that
   a. the number of phosphate groups always equals the number of five-carbon sugars.
   b. the proportions of A equal that of C and G equals T.
   c. the proportions of A equal that of T and G equals C.
   d. purines bind to pyrimidines.
4. The bonds that hold two complementary strands of DNA together are
   a. hydrogen bonds.
   b. peptide bonds.
   c. ionic bonds.
   d. phosphodiester bonds.

5. The basic mechanism of DNA replication is semiconservative with two new molecules,
   a. each with new strands.
   b. one with all new strands and one with all old strands.
   c. each with one new and one old strand.
   d. each with a mixture of old and new strands.
6. One common feature of all DNA polymerases is that they
   a. synthesize DNA in the 3′-to-5′ direction.
   b. synthesize DNA in the 5′-to-3′ direction.
   c. synthesize DNA in both directions by switching strands.
   d. do not require a primer.
7. Which of the following is *not* part of the Watson–Crick model of the structure of DNA?
   a. DNA is composed of two strands.
   b. The two DNA strands are oriented in parallel (5′-to-3′).
   c. Purines bind to pyrimidines.
   d. DNA forms a double helix.

APPLY

1. If one strand of a DNA is: 5′ ATCGTTAAGCGAGTCA 3′, then the complementary strand would be:
   a. 5′ TAGCAATTCGCTCAGT 3′.
   b. 5′ ACTGAGCGAATTGCTA 3′.
   c. 5′ TGACTCGCTTAACGAT 3′.
   d. 5′ ATCGTTAAGCGAGTCA 3′.
2. Hershey and Chase used radioactive phosphorus and sulfur to
   a. label DNA and protein with the same molecule.
   b. differentially label DNA and protein.
   c. identify the transforming principle.
   d. Both (b) and (c) are correct.
3. The Meselson and Stahl experiment used a density label to be able to
   a. determine the directionality of DNA replication.
   b. differentially label DNA and protein.
   c. distinguish between newly replicated and old strands.
   d. distinguish between replicated DNA and RNA primers.
4. The difference in leading- versus lagging-strand synthesis is a consequence of
   a. only the physical structure of DNA.
   b. only the activity of DNA polymerase enzymes.
   c. both the physical structure of DNA and the action of polymerase enzyme.
   d. the larger size of the lagging strand.
5. If the activity of DNA ligase was removed from replication, this would have a greater affect on
   a. synthesis on the lagging strand versus the leading strand.
   b. synthesis on the leading strand versus the lagging strand.
   c. priming of DNA synthesis versus actual DNA synthesis.
   d. photorepair of DNA versus DNA replication.
6. Successful DNA synthesis requires all of the following *except*
   a. helicase.
   b. endonuclease.
   c. DNA primase.
   d. DNA ligase.
7. The synthesis of telomeres
   a. uses DNA polymerase, but without the sliding clamp.
   b. uses enzymes involved in DNA repair.
   c. requires telomerase, which does not need a template.
   d. requires telomerase, which uses an internal RNA as a template.
8. Which type of enzyme is involved in excision repair?
   a. Photolyase
   b. DNA polymerase III
   c. Endonuclease
   d. Telomerase

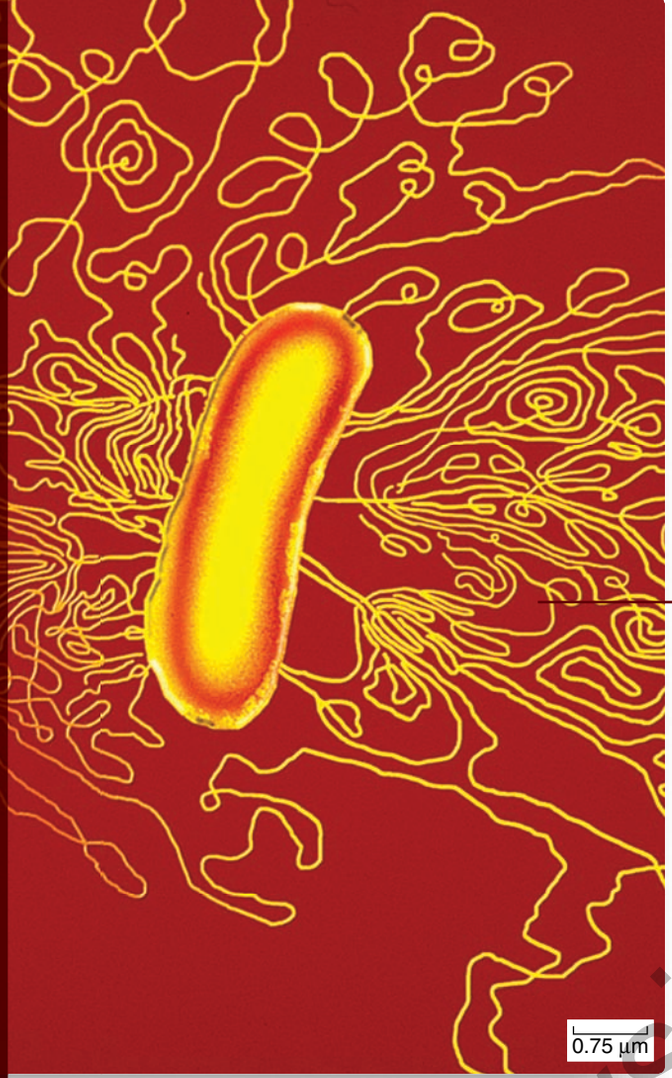
SYNTHESIZE

1. The work by Griffith provided the first indication that DNA was the genetic material. Review the four experiments outlined in figure 14.1. Predict the likely outcome for the following variations on this classic research.
   a. Heat-killed pathogenic and heat-killed nonpathogenic
   b. Heat-killed pathogenic and live nonpathogenic in the presence of an enzyme that digests proteins (proteases)
   c. Heat-killed pathogenic and live nonpathogenic in the presence of an enzyme that digests DNA (endonuclease)
2. In the Meselson–Stahl experiment, a control experiment was done to show that the hybrid bands after one round of replication were in fact two complete strands, one heavy and one light. Using the same experimental setup as detailed in the text, how can this be addressed?
3. Enzyme function is critically important for the proper replication of DNA. Predict the consequence of a loss of function for each of the following enzymes.
   a. DNA gyrase
   b. DNA polymerase III
   c. DNA ligase
   d. DNA polymerase I

ONLINE RESOURCE

www.ravenbiology.com

Understand, Apply, and Synthesize—enhance your study with animations that bring concepts to life and practice tests to assess your understanding. Your instructor may also recommend the interactive eBook, individualized learning tools, and more.



Chapter 15

Genes and How They Work

Chapter Outline

15.1 The Nature of Genes
15.2 The Genetic Code
15.3 Prokaryotic Transcription
15.4 Eukaryotic Transcription
15.5 Eukaryotic pre-mRNA Splicing
15.6 The Structure of tRNA and Ribosomes
15.7 The Process of Translation
15.8 Summarizing Gene Expression
15.9 Mutation: Altered Genes

Introduction

You've seen how genes specify traits and how those traits can be followed in genetic crosses. You've also seen that the information in genes resides in the DNA molecule; the picture above shows all the DNA within the entire E. coli *chromosome. Information in DNA is replicated by the cell and then partitioned equally during the process of cell division. The information in DNA is much like a blueprint for a building. The construction of the building uses the information in the blueprint, but requires building materials and carpenters and other skilled laborers using a variety of tools working together to actually construct the building. Similarly, the information in DNA requires nucleotide and amino acid building blocks, multiple forms of RNA, and many proteins acting in a coordinated fashion to make up the structure of a cell.*

We now turn to the nature of the genes themselves and how cells extract the information in DNA in the process of gene expression. *Gene expression can be thought of as the conversion of genotype into the phenotype.*

15.1 The Nature of Genes

Learning Outcomes

1. *Describe the evidence for the one-gene/one-polypeptide hypothesis.*
2. *Distinguish between transcription and translation.*
3. *List the roles played by RNA in gene expression.*

We know that DNA encodes proteins, but this knowledge alone tells us little about how the information in DNA can control cellular functions. Researchers had evidence that genetic mutations affected proteins, and in particular enzymes, long before the structure and code of DNA was known. In this section we review the evidence of the link between genes and enzymes.

Garrod concluded that inherited disorders can involve specific enzymes

In 1902, the British physician Archibald Garrod noted that certain diseases among his patients seemed to be more prevalent in

particular families. By examining several generations of these families, he found that some of the diseases behaved as though they were the product of simple recessive alleles. Garrod concluded that these disorders were Mendelian traits, and that they had resulted from changes in the hereditary information in an ancestor of the affected families.

Garrod investigated several of these disorders in detail. In alkaptonuria, patients produced urine that contained homogentisic acid (alkapton). This substance oxidized rapidly when exposed to air, turning the urine black. In normal individuals, homogentisic acid is broken down into simpler substances. With considerable insight, Garrod concluded that patients suffering from alkaptonuria lack the enzyme necessary to catalyze this breakdown. He speculated that many other inherited diseases might also reflect enzyme deficiencies.

Beadle and Tatum showed that genes specify enzymes

From Garrod's finding, it took but a short leap of intuition to surmise that the information encoded within the DNA of chromosomes acts to specify particular enzymes. This point was not actually established, however, until 1941, when a series of experiments by George Beadle and Edward Tatum at Stanford University provided definitive evidence. Beadle and Tatum deliberately set out to create mutations in chromosomes and verified that they behaved in a Mendelian fashion in crosses. These alterations to single genes were analyzed for their effects on the organism (figure 15.1).

Neurospora crassa, *the bread mold*

One of the reasons Beadle and Tatum's experiments produced clear-cut results was their choice of experimental organism, the bread mold *Neurospora crassa*. This fungus can be grown readily in the laboratory on a defined medium consisting of only a carbon source (glucose), a vitamin (biotin), and inorganic salts. This type of medium is called "minimal" because it represents the minimal requirements to support growth. Any cells that can grow on minimal medium must be able to synthesize all necessary biological molecules.

Beadle and Tatum exposed *Neurospora* spores to X-rays, expecting that the DNA in some of the spores would experience damage in regions encoding the ability to make compounds needed for normal growth (see figure 15.1). Such a mutation would cause cells to be unable to grow on minimal medium. Such mutations are called **nutritional mutations** because cells carrying them grow only if the medium is supplemented with additional nutrients.

Nutritional mutants

To identify mutations causing metabolic deficiencies, Beadle and Tatum placed subcultures of individual fungal cells grown on a rich medium onto minimal medium. Any cells that had lost the ability to make compounds necessary for growth would not grow on minimal medium. Using this approach, Beadle and Tatum succeeded in isolating and identifying many nutritional mutants.

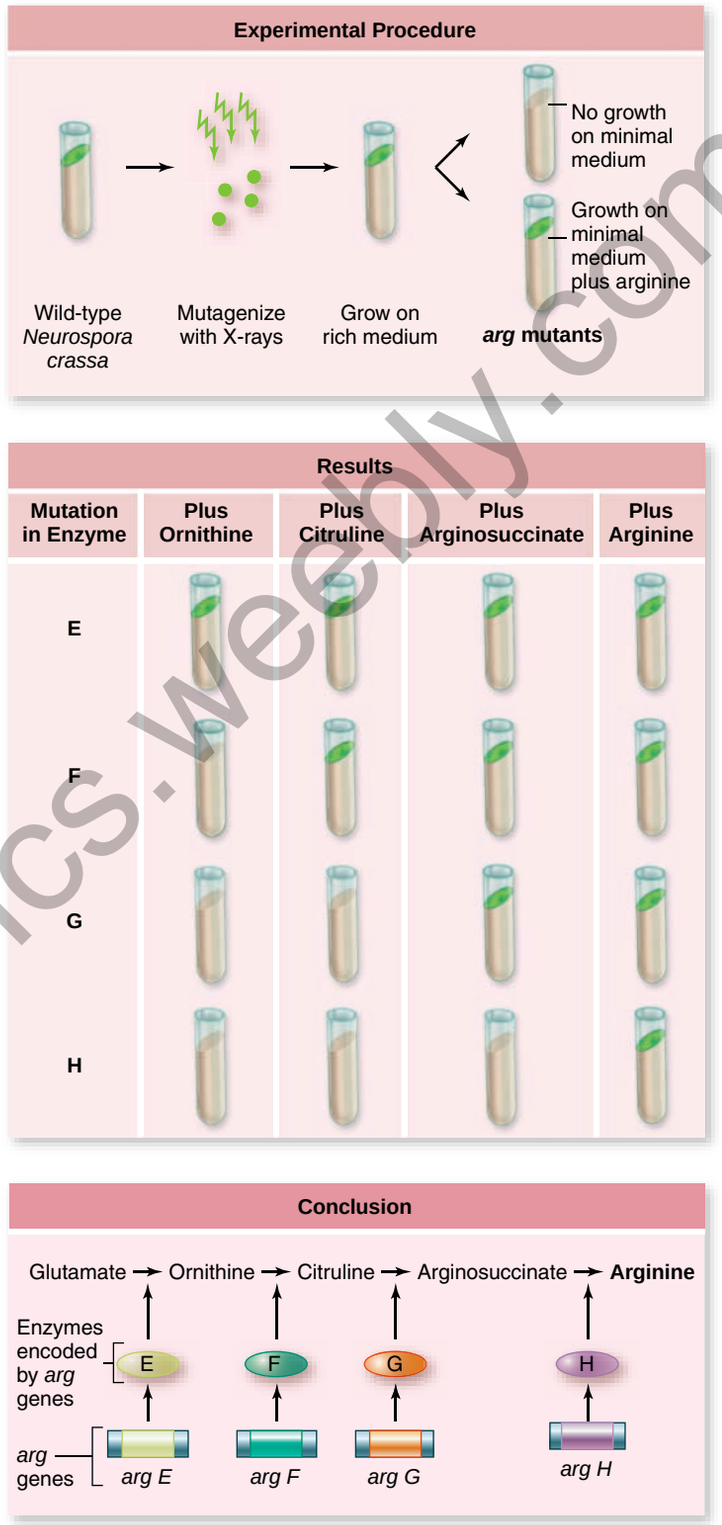


Figure 15.1 The Beadle and Tatum experiment.
Wild-type *Neurospora* were mutagenized with X-rays to produce mutants deficient in the synthesis of arginine (top panel). The specific defect in each mutant was identified by growing on medium supplemented with intermediates in the biosynthetic pathway for arginine (middle panel). A mutant will grow only on media supplemented with an intermediate produced after the defective enzyme in the pathway for each mutant. The enzymes in the pathway can then be correlated with genes on chromosomes (bottom panel).

Next, the researchers supplemented the minimal medium with different compounds known to be intermediates in this biochemical pathway to identify the deficiency in each mutant. This step allowed them to pinpoint the nature of the strain's biochemical deficiency. They concentrated in particular on mutants that would grow only in the presence of the amino acid arginine, dubbed *arg* mutants. These were shown to define four genes they named *argE*, *argF*, *argG*, and *argH*. When their chromosomal positions were located, the *arg* mutations were found to cluster in three areas.

One gene/one polypeptide

The next step was to determine where each mutation was blocked in the biochemical pathway for arginine biosynthesis. To do this, they supplemented the medium with each intermediate in the pathway to see which intermediate would support a mutant's growth. If the mutation affects an enzyme in the pathway that acts prior to the intermediate used as a supplement, then growth should be supported—but not if the mutation affects a step after the intermediate used (see figure 15.1). For each enzyme in the arginine biosynthetic pathway, Beadle and Tatum were able to isolate a mutant strain with a defective form of that enzyme. The mutation was always located at one of a few specific chromosomal sites, and each mutation had a unique location. Thus, each of the mutants they examined had a defect in a single enzyme, caused by a mutation at a single site on a chromosome.

Beadle and Tatum concluded that genes specify the structure of enzymes, and that each gene encodes the structure of one enzyme (see figure 15.1). They called this relationship the *one-gene/one-enzyme hypothesis*. Today, because many enzymes contain multiple polypeptide subunits, each encoded by a separate gene, the relationship is more commonly referred to as the **one-gene/one-polypeptide hypothesis.** This hypothesis clearly states the molecular relationship between genotype and phenotype.

As you learn more about genomes and gene expression, you'll find that this clear relationship is overly simple. As described later in this chapter, eukaryotic genes are more complex than those of prokaryotes. In addition, some enzymes are composed, at least in part, of RNA, itself an intermediate in the production of proteins. Nevertheless, this one-gene/one-polypeptide concept is a useful starting point for thinking about gene expression.

The central dogma describes information flow in cells as DNA to RNA to protein

The conversion of genotype to phenotype requires information stored in DNA to be converted to protein. The nature of information flow in cells was first described by Francis Crick as the **central dogma of molecular biology.** Information passes in one direction from the gene (DNA) to an RNA copy of the gene, and the RNA copy directs the sequential assembly of a chain of amino acids into a protein (figure 15.2). Stated briefly,

$$\text{DNA} \longrightarrow \text{RNA} \longrightarrow \text{protein}$$

The central dogma provides an intellectual framework that describes information flow in biological systems. We call the

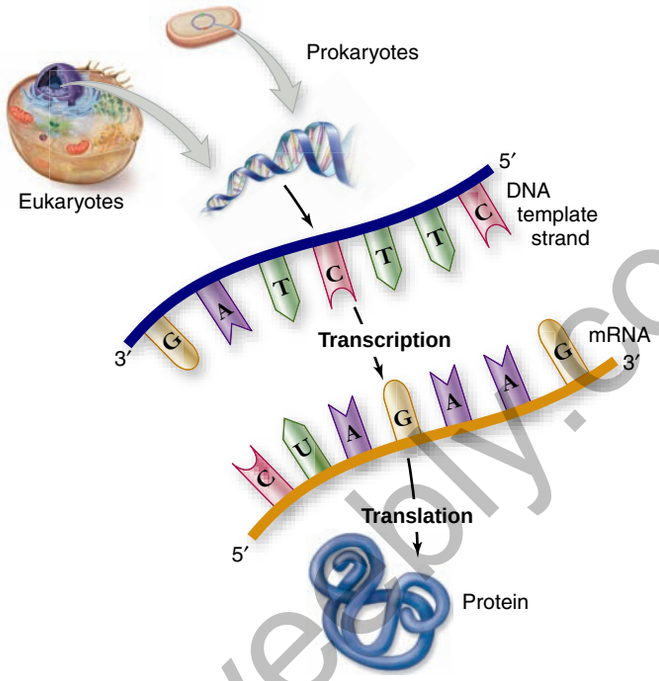
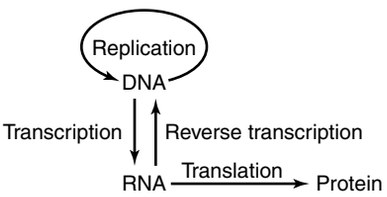


Figure 15.2 The central dogma of molecular biology.
DNA is transcribed to make mRNA, which is translated to make a protein.

DNA-to-RNA step **transcription** because it produces an exact copy of the DNA, much as a legal transcription contains the exact words of a court proceeding. The RNA-to-protein step is termed **translation** because it requires translating from the nucleic acid to the protein "languages."

Since the original formulation of the central dogma, a class of viruses called **retroviruses** was discovered that can convert their RNA genome into a DNA copy, using the viral enzyme **reverse transcriptase.** This conversion violates the direction of information flow of the central dogma, and the discovery forced an updating of the possible flow of information to include this "reverse" flow from RNA to DNA.



Transcription makes an RNA copy of DNA

The process of transcription produces an RNA copy of the information in DNA. That is, transcription is the DNA-directed synthesis of RNA by the enzyme RNA polymerase (figure 15.3). This process uses the principle of complementarity, described in chapter 14, to use DNA as a template to make RNA.

Because DNA is double-stranded and RNA is single-stranded, only one of the two DNA strands needs to be copied. We call the strand that is copied the **template strand.** The RNA transcript's sequence is complementary to the template strand. The strand of DNA not used as a template is called the **coding strand.** It has the same sequence as the RNA transcript,

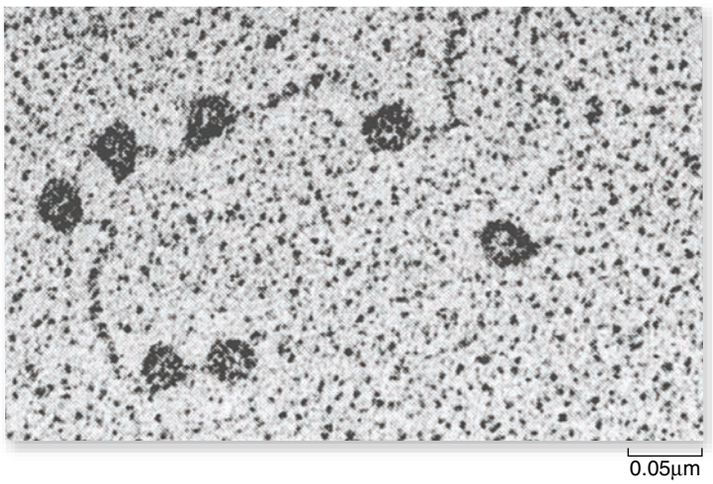


Figure 15.3 RNA polymerase. In this electron micrograph, the dark circles are RNA polymerase molecules synthesizing RNA from a DNA template.

except that U (uracil) in the RNA is T (thymine) in the DNA-coding strand. Another naming convention for the two strands of the DNA is to call the coding strand the sense strand, as it has the same "sense" as the RNA. The template strand would then be the antisense strand.

The RNA transcript used to direct the synthesis of polypeptides is termed *messenger RNA (mRNA)*. Its name reflects its use by the cell to carry the DNA message to the ribosome for processing.

Inquiry question

It is widely accepted that RNA polymerase has no proofreading capacity. Would you expect high or low levels of error in transcription compared with DNA replication? Why do you think it is more important for DNA polymerase than for RNA polymerase to proofread?

Translation uses information in RNA to synthesize proteins

The process of translation is by necessity much more complex than transcription. In this case, RNA cannot be used as a direct template for a protein because there is no complementarity—that is, a sequence of amino acids cannot be aligned to an RNA template based on any kind of "chemical fit." Molecular geneticists suggested that some kind of adapter molecule must exist that can interact with both RNA and amino acids, and *transfer RNA (tRNA)* was found to fill this role. This need for an intermediary adds a level of complexity to the process that is not seen in either DNA replication or transcription of RNA.

Translation takes place on the ribosome, the cellular protein-synthesis machinery, and it requires the participation of multiple kinds of RNA and many proteins. Here we provide an outline of the processes; all are described in detail in the sections that follow.

RNA has multiple roles in gene expression

All RNAs are synthesized from a DNA template by transcription. Gene expression requires the participation of multiple kinds of RNA, each with different roles in the overall process. Here is a brief summary of these roles, which are described in detail later.

Messenger RNA. Even before the details of gene expression were unraveled, geneticists recognized that there must be an intermediate form of the information in DNA that can be transported out of the eukaryotic nucleus to the cytoplasm for ribosomal processing. This hypothesis was called the "messenger hypothesis," and we retain this language in the name *messenger RNA* (mRNA).

Ribosomal RNA. The class of RNA found in ribosomes is called **ribosomal RNA (rRNA).** There are multiple forms of rRNA, and rRNA is found in both ribosomal subunits. This rRNA is critical to the function of the ribosome.

Transfer RNA. The intermediary adapter molecule between mRNA and amino acids, as mentioned earlier, is **transfer RNA (tRNA).** Transfer RNA molecules have amino acids covalently attached to one end and an anticodon that can base-pair with an mRNA codon at the other. The tRNAs act to interpret information in mRNA and to help position the amino acids on the ribosome.

Small nuclear RNA. Small nuclear RNAs (snRNAs) are part of the machinery that is involved in nuclear processing of eukaryotic "pre-mRNA." We discuss this splicing reaction later in the chapter.

SRP RNA. In eukaryotes, where some proteins are synthesized by ribosomes on the rough endoplasmic reticulum (RER), this process is mediated by the **signal recognition particle,** or **SRP,** described later in the chapter. The SRP contains both RNA and proteins.

Micro-RNA. This class of RNA was discovered relatively recently. These very short RNAs, called **micro-RNAs,** or **miRNA,** have gone from unknown to a major class of regulatory molecules in a very short period of time.

Learning Outcomes Review 15.1

Garrod showed that altered enzymes can cause metabolic disorders. Beadle and Tatum demonstrated that each gene encodes a unique enzyme. Genetic information flows from DNA (genes) to protein (enzymes) using messenger RNA as an intermediate. Transcription converts information in DNA into an RNA transcript, and translation converts this information into protein. RNA comes in several varieties having different functions; these include mRNA (the transcript), tRNA (the intermediary), and rRNA (in ribosomes), as well as snRNA, SNP RNA, and micro-RNA.

- ***Why do cells need an adapter molecule like tRNA between RNA and protein?***

15.2 The Genetic Code

Learning Outcomes

1. *Summarize the experiments that revealed the genetic code.*
2. *Describe the characteristics of the genetic code.*
3. *Identify the relationship between codons and amino acids.*

How does the order of nucleotides in a DNA molecule encode the information that specifies the order of amino acids in a polypeptide? The answer to this essential question came in 1961, through an experiment led by Francis Crick and Sydney Brenner. That experiment was so elegant and the result so critical to understanding the genetic code that we describe it here in detail.

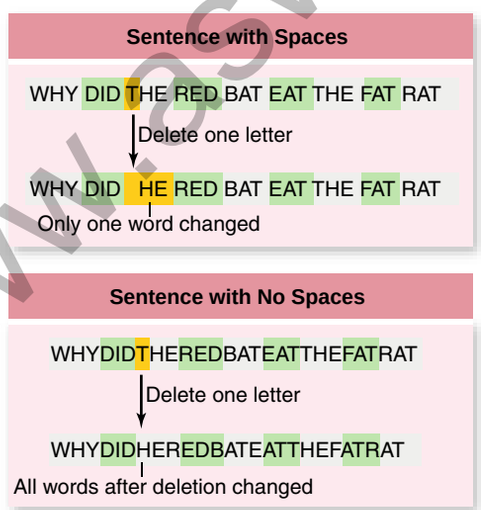
The code is read in groups of three

Crick and Brenner reasoned that the genetic code most likely consisted of a series of blocks of information called **codons,** each corresponding to an amino acid in the encoded protein. They further hypothesized that the information within one codon was probably a sequence of three nucleotides. With four DNA nucleotides (G, C, T, and A), using two in each codon can produce only 4^2, or 16, different codons—not enough to code for 20 amino acids. However, three nucleotides results in 4^3, or 64, different combinations of three, more than enough.

Spaced or unspaced codons?

In theory, the sequence of codons in a gene could be punctuated with nucleotides between the codons that are not used, like the spaces that separate the words in this sentence. Alternatively, the codons could lie immediately adjacent to each other, forming a continuous sequence of nucleotides.

If the information in the genetic message is separated by spaces, then altering any single word would not affect the entire sentence. In contrast, if all of the words are run together but read in groups of three, then any alteration that is not in groups of three would alter the entire sentence. These two ways of using information in DNA imply different methods of translating the information into protein.



Determining that codons are unspaced

To choose between these alternative mechanisms, Crick and his colleagues used a chemical to create mutations that caused single-base insertions or deletions from a viral DNA molecule. They then showed that combining an insertion with a deletion restored function even though either one individually displayed loss of function. In this case, only the region between the insertion or deletion would be altered. By choosing a region of the gene that encoded a part of the protein not critical to function, this small change did not cause a change in phenotype.

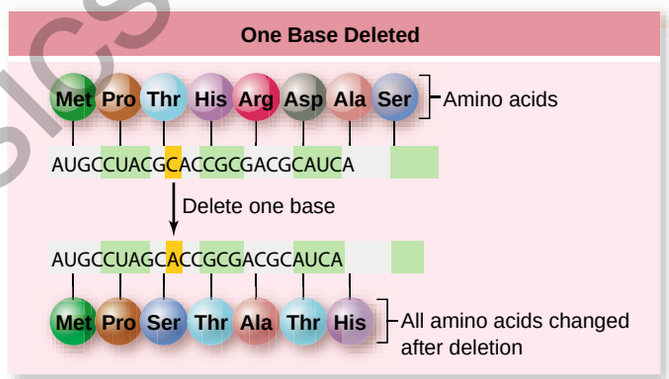
When they combined a single deletion or two deletions near each other, the genetic message shifted, altering all of the amino acids after the deletion. When they made three deletions,

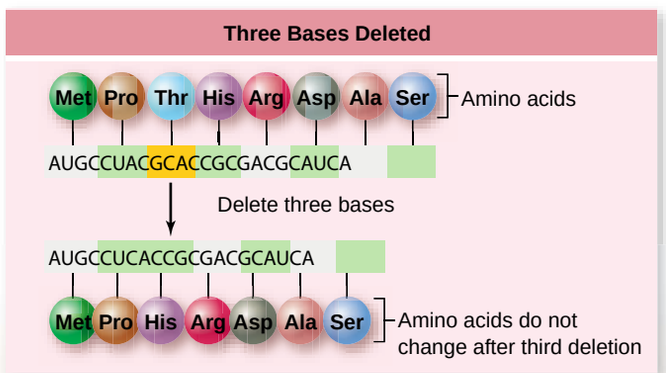
SCIENTIFIC THINKING

Hypothesis: *The genetic code is read in groups of three bases.*

Prediction: *If the genetic code is read in groups of three, then deletion of one or two bases would shift the reading frame after the deletion. Deletion of three bases, however, would produce a protein with a single amino acid deleted but no change downstream.*

Test: *Single-base deletion mutants are collected, each of which exhibits a mutant phenotype. Three of these deletions in a single region are combined to assess the effect of deletion of three bases.*





Result: *The combination of three deletions does not have the same drastic effect as the loss of one or two bases.*

Conclusion: *The genetic code is read in groups of three.*

Further Experiments: *If you also had mutants with one base additions, what would be the effect of combining a deletion and an addition?*

Figure 15.4 **The genetic code is triplet.**

however, the protein after the deletions was normal. They obtained the same results when they made additions to the DNA consisting of 1, 2, or 3 nt (nucleotides).

Thus, Crick and Brenner concluded that the genetic code is read in increments of three nucleotides (in other words, it is a triplet code), and that reading occurs continuously without punctuation between the 3-nt units (figure 15.4).

These experiments indicate the importance of the **reading frame** for the genetic message. Because there is no punctuation, the reading frame established by the first codon in the sequence determines how all subsequent codons are read. We now call the kinds of mutations that Crick and Brenner used **frameshift mutations** because they alter the reading frame of the genetic message.

Nirenberg and others deciphered the code

The determination of which of the 64 possible codons encoded each particular amino acids was one of the greatest triumphs of 20th-century biochemistry. Accomplishing this decryption required two main developments to succeed: (1) cell-free biochemical systems that would support protein synthesis from a defined RNA and (2) the ability to produce synthetic, defined RNAs that could be used in the cell-free system.

During a five-year period from 1961 to 1966, work performed primarily in Marshall Nirenberg's laboratory led to the elucidation of the genetic code. Nirenberg's group first showed that adding the synthetic RNA molecule polyU (an RNA molecule consisting of a string of uracil nucleotides) to their cell-free systems produced the polypeptide polyphenylalanine (a string of phenylalanine amino acids). Therefore, UUU encodes phenylalanine.

Next they used enzymes to produce RNA polymers with more than one nucleotide. These polymers allowed them to infer the composition of many of the possible codons, but not the order of bases in each codon.

The researchers then were able to use enzymes to synthesize defined 3-base sequences that could be tested for binding to the protein-synthesis machinery. This so-called *triplet-binding assay* allowed them to identify 54 of the 64 possible triplets.

The organic chemist H. Gobind Khorana provided the final piece of the puzzle by using organic synthesis to produce artificial RNA molecules of defined sequence, and then examining what polypeptides they directed in cell-free systems. The combination of all of these methods allowed the determination of all 64 possible 3-nt sequences, and the full genetic code was determined (table 15.1).

The code is degenerate but specific

Some obvious features of the code jump out of table 15.1. First, 61 of the 64 possible codons are used to specify amino acids. Three codons, UAA, UGA, and UAG, are reserved for another function: they signal "stop" and are known as **stop codons.** The only other form of "punctuation" in the code is that AUG is used to signal "start" and is therefore the **start codon.** In this case the codon has a dual function because it also encodes the amino acid methionine (Met).

TABLE 15.1 The Genetic Code

First Letter	SECOND LETTER U	C	A	G	Third Letter
U	UUU Phe Phenylalanine	UCU Ser Serine	UAU Tyr Tyrosine	UGU Cys Cysteine	U
	UUC	UCC	UAC	UGC	C
	UUA Leu Leucine	UCA	**UAA** "Stop"	**UGA** "Stop"	A
	UUG	UCG	**UAG** "Stop"	UGG Trp Tryptophan	G
C	CUU Leu Leucine	CCU Pro Proline	CAU His Histidine	CGU Arg Arginine	U
	CUC	CCC	CAC	CGC	C
	CUA	CCA	CAA Gln Glutamine	CGA	A
	CUG	CCG	CAG	CGG	G
A	AUU Ile Isoleucine	ACU Thr Threonine	AAU Asn Asparagine	AGU Ser Serine	U
	AUC	ACC	AAC	AGC	C
	AUA	ACA	AAA Lys Lysine	AGA Arg Arginine	A
	AUG Met Methionine; "Start"	ACG	AAG	AGG	G
G	GUU Val Valine	GCU Ala Alanine	GAU Asp Aspartate	GGU Gly Glycine	U
	GUC	GCC	GAC	GGC	C
	GUA	GCA	GAA Glu Glutamate	GGA	A
	GUG	GCG	GAG	GGG	G

A codon consists of three nucleotides read in the sequence shown. For example, ACU codes for threonine. The first letter, A, is in the First Letter column; the second letter, C, is in the Second Letter column; and the third letter, U, is in the Third Letter column. Each of the mRNA codons is recognized by a corresponding anticodon sequence on a tRNA molecule. Many amino acids are specified by more than one codon. For example, threonine is specified by four codons, which differ only in the third nucleotide (ACU, ACC, ACA, and ACG).

Figure 15.5 Transgenic pig. The piglet on the right is a conventional piglet. The piglet on the left was engineered to express a gene from jellyfish that encodes green fluorescent protein. The color of this piglet's nose is due to expression of this introduced gene. Such transgenic animals indicate the universal nature of the genetic code.

You can see that 61 codons are more than enough to encode 20 amino acids. That leaves lots of extra codons. One way to deal with this abundance would be to use only 20 of the 61 codons, but this is not what cells do. In reality, all 61 codons are used, making the code **degenerate,** which means that some amino acids are specified by more than one codon. The reverse, however, in which a single codon would specify more than one amino acid, is never found.

This degeneracy is not uniform. Some amino acids have only one codon, and some have up to six. In addition, the degenerate base usually occurs in position 3 of a codon, such that the first two positions are the same, and two or four of the possible nucleotides at position 3 encode the same amino acid. (The nature of protein synthesis on ribosomes explains how this codon usage works, and it is discussed later.)

The code is practically universal, but not quite

The genetic code is the same in almost all organisms. The universality of the genetic code is among the strongest evidence that all living things share a common evolutionary heritage. Because the code is universal, genes can be transferred from one organism to another and can be successfully expressed in their new host (figure 15.5). This universality of gene expression is central to many of the advances of genetic engineering discussed in chapter 17.

In 1979, investigators began to determine the complete nucleotide sequences of the mitochondrial genomes in humans, cattle, and mice. It came as something of a shock when these investigators learned that the genetic code used by these mammalian mitochondria was not quite the same as the "universal code" that has become so familiar to biologists.

In the mitochondrial genomes, what should have been a stop codon, UGA, was instead read as the amino acid tryptophan; AUA was read as methionine rather than as isoleucine; and AGA and AGG were read as stop codons rather than as arginine. Furthermore, minor differences from the universal code have also been found in the genomes of chloroplasts and in ciliates (certain types of protists).

Thus, it appears that the genetic code is not quite universal. Some time ago, presumably after they began their endosymbiotic existence, mitochondria and chloroplasts began to read the code differently, particularly the portion associated with "stop" signals.

Inquiry question

The genetic code is almost universal. Why do you think it is nearly universal?

Learning Outcomes Review 15.2

The genetic code was shown to be nucelotide base triplets with two forms of punctuation and no spaces: three bases code for an amino acid, and the groups of three are read in order. Sixty-one codons specify amino acids, one of which also codes for "start," and three codons indicate "stop," for 64 total. Because some amino acids are specified by more than one codon, the code is termed degenerate. All codons encode only one amino acid, however.

- ***What would be the outcome if a codon specified more than one amino acid?***

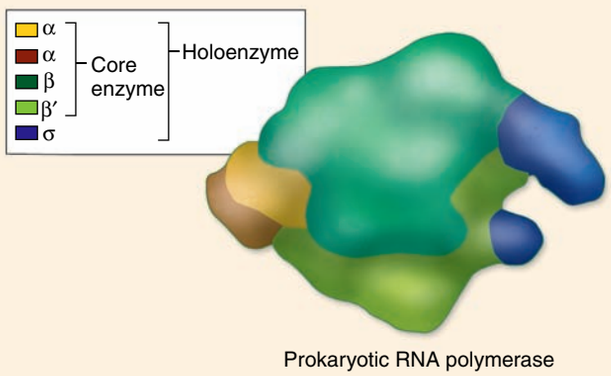
15.3 *Prokaryotic Transcription*

Learning Outcomes

1. ***Describe the transcription process in bacteria.***
2. ***Differentiate features of initiation from those of elongation.***
3. ***Define the unique features of prokaryotic transcription.***

We begin an examination of gene expression by describing the process of transcription in prokaryotes. The later description of eukaryotic transcription will concentrate on their differences from prokaryotes.

Figure 15.6 Bacterial RNA polymerase and transcription initiation. ***a.*** RNA polymerase has two forms: core polymerase and holoenzyme. ***b.*** The σ subunit of the holoenzyme recognizes promoter elements at –35 and –10 and binds to the DNA. The helix is opened at the –10 region, and transcription begins at the start site at +1.



a.

Prokaryotes have a single RNA polymerase

The single **RNA polymerase** of prokaryotes exists in two forms: the *core polymerase* and the *holoenzyme*. The core polymerase can synthesize RNA using a DNA template, but it cannot initiate synthesis accurately. The holoenzyme can accurately initiate synthesis.

The core polymerase is composed of four subunits: two identical α subunits, a β subunit, and a β′ subunit (figure 15.6*a*). The two α subunits help to hold the complex together and can bind to regulatory molecules. The active site of the enzyme is formed by the β and β′ subunits, which bind to the DNA template and the ribonucleotide triphosphate precursors.

The *holoenzyme*, which can properly initiate synthesis, is formed by the addition of a σ (sigma) subunit to the core polymerase (see figure 15.6*a*). Its ability to recognize specific signals in DNA allows RNA polymerase to locate the beginning of genes, which is critical to its function. Note that initiation of mRNA synthesis does not require a primer, in contrast to DNA replication.

Initiation occurs at promoters

Accurate initiation of transcription requires two sites in DNA: one called a **promoter** that forms a recognition and binding site for the RNA polymerase, and the actual **start site.** The polymerase also needs a signal to end transcription, which we call a **terminator.** We then refer to the region from promoter to terminator as a **transcription unit.**

The action of the polymerase moving along the DNA can be thought of as analogous to water flowing in a stream. We can speak of sites on the DNA as being "upstream" or "downstream" of the start site. We can also use this comparison to form a simple system for numbering bases in DNA to refer to positions in the transcription unit. The first base transcribed is called **+1,** and this numbering continues downstream until the last base is transcribed. Any bases upstream of the start site receive negative numbers, starting at **–1.**

The promoter is a short sequence found upstream of the start site and is therefore not transcribed by the polymerase. Two 6-base sequences are common to bacterial promoters: One is located 35 nt upstream of the start site (–35), and the other is located 10 nt upstream of the start site (–10) (figure 15.6*b*). These two sites provide the promoter with asymmetry; they indicate not only the site of initiation, but also the direction of transcription.

The binding of RNA polymerase to the promoter is the first step in transcription. Promoter binding is controlled by the σ subunit of the RNA polymerase holoenzyme, which recognizes the –35 sequence in the promoter and positions the RNA polymerase at the correct start site, oriented to transcribe in the correct direction.

Once bound to the promoter, the RNA polymerase begins to unwind the DNA helix at the –10 site (see figure 15.6*b*). The polymerase covers a region of about 75 bp but only unwinds about 12–14 bp.

Inquiry question

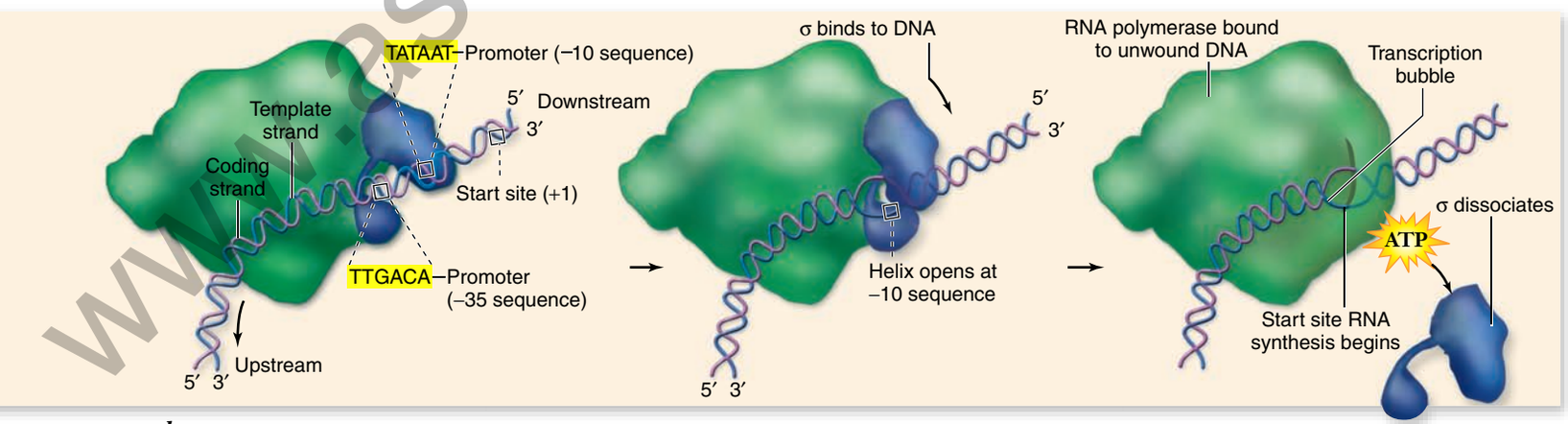
The prokaryotic promoter has two distinct elements that are not identical. How is this important to the initiation of transcription?

Elongation adds successive nucleotides

In prokaryotes, the transcription of the RNA chain usually starts with ATP or GTP. One of these forms the 5′ end of the chain, which grows in the 5′-to-3′ direction as ribonucleotides are added. As the RNA polymerase molecule leaves the promoter region, the σ factor is no longer required, although it may remain in association with the enzyme.

This process of leaving the promoter, called *clearance*, or *escape*, involves more than just synthesizing the first few nucleotides of the transcript and moving on, because the enzyme has made strong contacts to the DNA during initiation. It is necessary to break these contacts with the promoter region to be able to move progressively down the template. The enzyme goes through conformational changes during this clearance stage, and subsequently contacts less of the DNA than it does during the initial promoter binding.

The region containing the RNA polymerase, the DNA template, and the growing RNA transcript is called the **transcription bubble** because it contains a locally unwound



b.

"bubble" of DNA (figure 15.7). Within the bubble, the first 9 bases of the newly synthesized RNA strand temporarily form a helix with the template DNA strand. This stabilizes the positioning of the 3′ end of the RNA so it can interact with an incoming ribonucleotide triphosphate. The enzyme itself covers about 50 bp of DNA around this transcription bubble.

The transcription bubble created by RNA polymerase moves down the bacterial DNA at a constant rate, about 50 nt/sec, with the growing RNA strand protruding from the bubble. After the transcription bubble passes, the now-transcribed DNA is rewound as it leaves the bubble.

Termination occurs at specific sites

The end of a bacterial transcription unit is marked by terminator sequences that signal "stop" to the polymerase. Reaching these sequences causes the formation of phosphodiester bonds to cease, the RNA–DNA hybrid within the transcription bubble to dissociate, the RNA polymerase to release the DNA, and the DNA within the transcription bubble to rewind.

The simplest terminators consist of a series of G–C base-pairs followed by a series of A–T base-pairs. The RNA transcript of this stop region can form a double-stranded structure in the GC region called a *hairpin*, which is followed by four or more uracil (U) ribonucleotides (figure 15.8). Formation of the hairpin causes the RNA polymerase to pause, placing it directly over the run of four uracils. The pairing of U with the DNA's A is the weakest of the four hybrid base-pairs, and it is not strong enough to hold the hybrid strands when the polymerase pauses. Instead, the RNA strand dissociates from the DNA within the transcription bubble, and transcription stops. A variety of protein factors also act at these terminators to aid in terminating transcription.

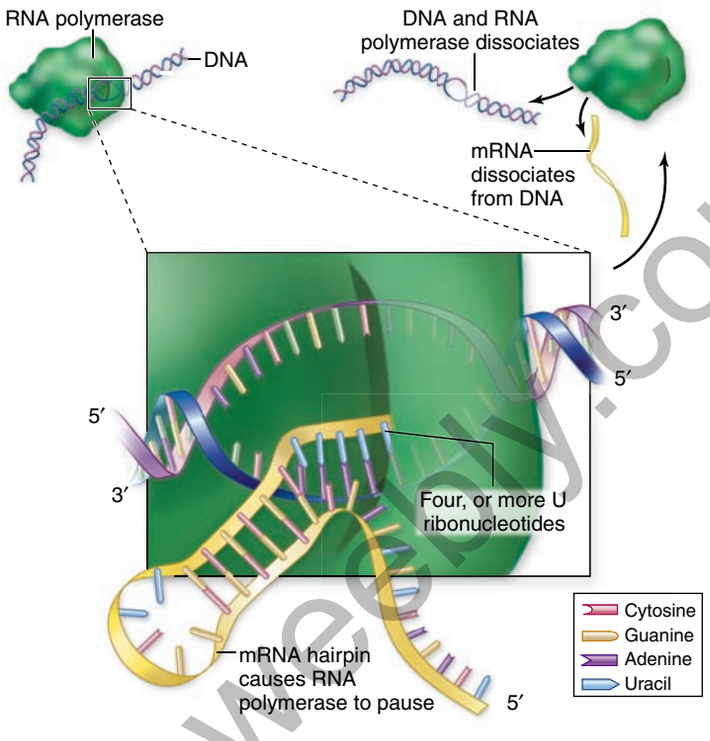


Figure 15.8 Bacterial transcription terminator. The self-complementary G–C region forms a double-stranded stem with a single-stranded loop called a hairpin. The stretch of U's forms a less stable RNA–DNA hybrid that falls off the enzyme.

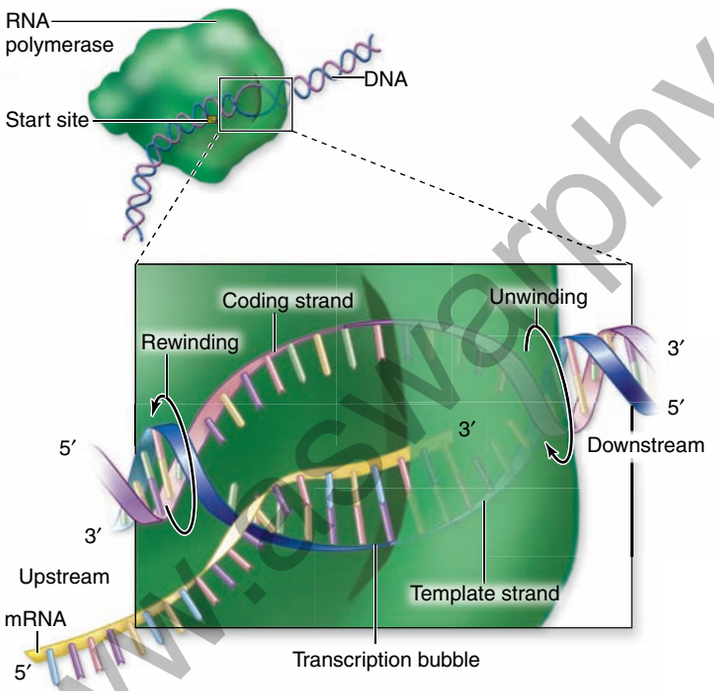


Figure 15.7 Model of a transcription bubble. The DNA duplex is unwound by the RNA polymerase complex, rewinding at the end of the bubble. One of the strands of DNA functions as a template, and nucleotide building blocks are added to the 3′ end of the growing RNA. There is a short region of RNA–DNA hybrid within the bubble.

Prokaryotic transcription is coupled to translation

In prokaryotes, the mRNA produced by transcription begins to be translated before transcription is finished—that is, they are *coupled* (figure 15.9). As soon as a 5′ end of the mRNA becomes available, ribosomes are loaded onto this to begin translation. (This coupling cannot occur in eukaryotes because transcription occurs in the nucleus, and translation occurs in the cytoplasm.)

Another difference between prokaryotic and eukaryotic gene expression is that the mRNA produced in prokaryotes may contain multiple genes. Prokaryotic genes are often organized such that genes encoding related functions are clustered together. This grouping of functionally related genes is referred to as an **operon.** An operon is a single transcription unit that encodes multiple enzymes necessary for a biochemical pathway. By clustering genes by function, they can be regulated together, a topic that we return to in the next chapter.

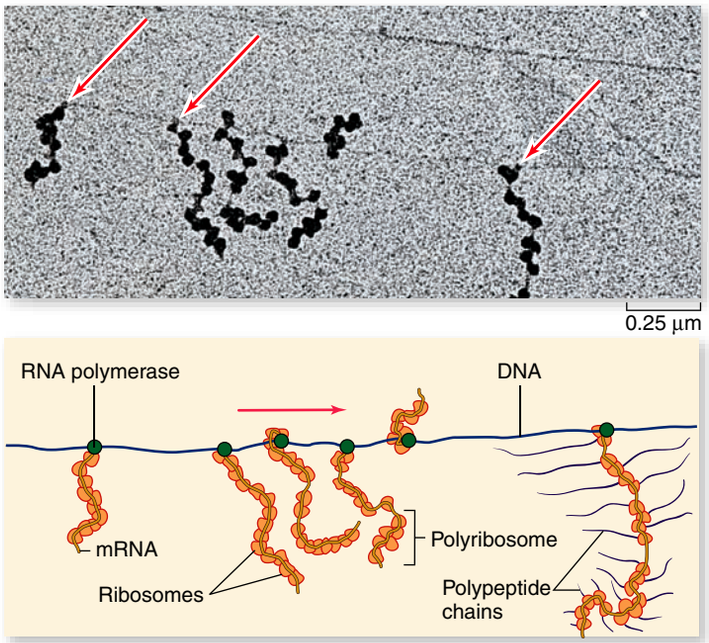


Figure 15.9 Transcription and translation are coupled in prokaryotes. In this micrograph of gene expression in *E. coli*, translation is occurring during transcription. The arrows point to RNA polymerase enzymes, and ribosomes are attached to the mRNAs extending from the polymerase. Polypeptides being synthesized by ribosomes, which are not visible in the micrograph, have been added to the last mRNA in the drawing.

Learning Outcomes Review 15.3

Transcription in bacteria is accomplished by RNA polymerase, which has two forms: a core polymerase and a holoenzyme. Initiation is accomplished by the holoenzyme form, which can accurately recognize promoter sequences. Elongation consists of RNA synthesis by the core enzyme, which adds RNA nucleotides in sequence until it reaches a terminator where synthesis stops, then the transcript is released. In prokaryotes, translation of the RNA transcript begins before transcription is finished, making the processes coupled.

- ***Yeast are unicellular organism like bacteria; would you expect them to have the same transcription/translation coupling?***

15.4 Eukaryotic Transcription

Learning Outcomes

1. ***List the different eukaryotic RNA polymerases.***
2. ***Distinguish between the promoters of the RNA polymerases.***
3. ***Define the processing that occurs to eukaryotic transcripts.***

The basic mechanism of transcription by RNA polymerase is the same in eukaryotes as in prokaryotes; however, the details of the two processes differ enough that it is necessary to consider them separately. Here we concentrate only on how eukaryotic systems differ from prokaryotic systems, such as the bacterial system just discussed. All other features may be assumed to be the same.

Eukaryotes have three RNA polymerases

Unlike prokaryotes, which have a single RNA polymerase enzyme, eukaryotes have three different RNA polymerases, which are distinguished in both structure and function. The enzyme **RNA polymerase I** transcribes rRNA, **RNA polymerase II** transcribes mRNA and some small nuclear RNAs, and **RNA polymerase III** transcribes tRNA and some other small RNAs. Together, these three enzymes accomplish all transcription in the nucleus of eukaryotic cells.

Each polymerase has its own promoter

The existence of three different RNA polymerases requires different signals in the DNA to allow each polymerase to recognize where to begin transcription. Each polymerase recognizes a different promoter structure.

RNA polymerase I promoters

RNA polymerase I promoters at first puzzled biologists, because comparisons of rRNA genes between species showed no similarities outside the coding region. The current view is that these promoters are also specific for each species, and for this reason, cross-species comparisons do not yield similarities.

RNA polymerase II promoters

The RNA polymerase II promoters are the most complex of the three types, probably a reflection of the huge diversity of genes that are transcribed by this polymerase. When the first eukaryotic genes were isolated, many had a sequence called the **TATA box** upstream of the start site. This sequence was similar to the prokaryotic –10 sequence, and it was assumed that the TATA box was the primary promoter element. With the sequencing of entire genomes, many more genes have been analyzed, and this assumption has proved too simple. It has been replaced by the idea of a "core promoter" that can be composed of a number of different elements, including the TATA box. Additional control elements allow for tissue-specific and developmental time–specific expression (see chapter 16).

RNA polymerase III promoters

Promoters for RNA polymerase III also were a source of surprise for biologists in the early days of molecular biology who were examining the control of eukaryotic gene expression. A common technique for analyzing regulatory regions was to make successive deletions from the 5′ end of genes until enough was deleted to abolish specific transcription. The logic followed experiences with prokaryotes, in which the regulatory regions had been found at the 5′ end of genes. But in the case of tRNA genes, the 5′ deletions had no effect on

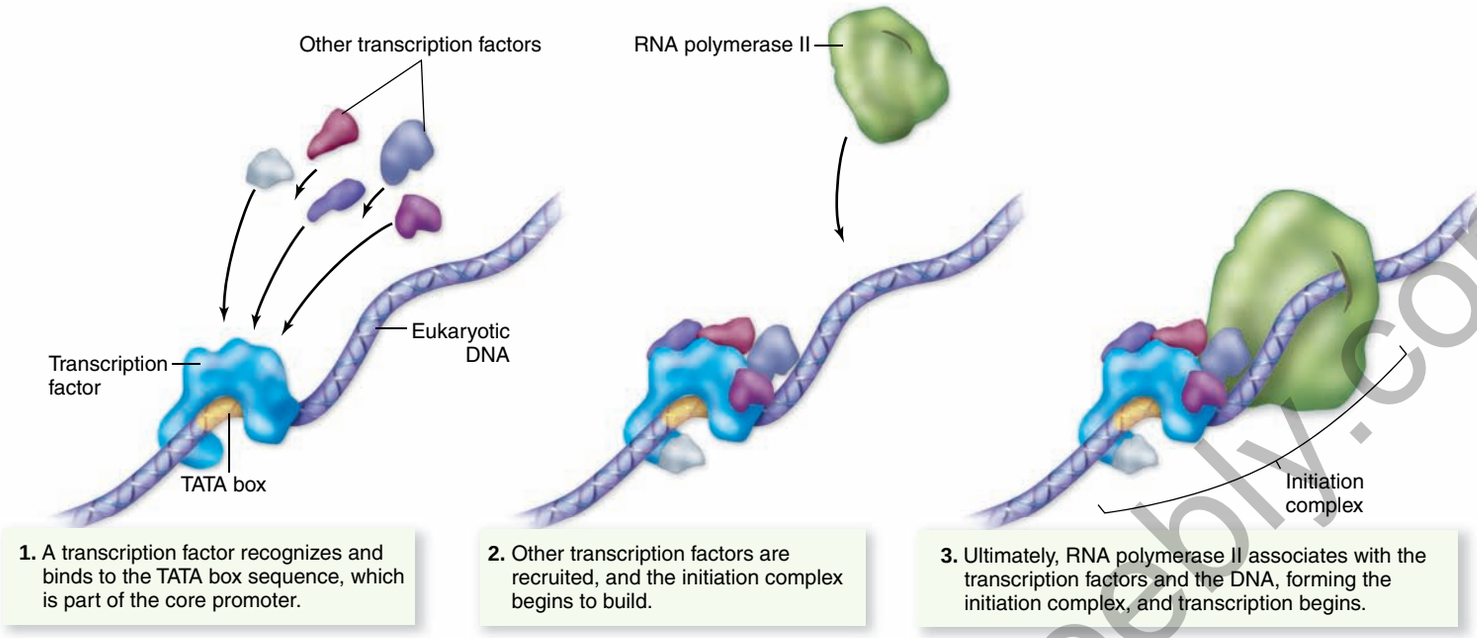


Figure 15.10 Eukaryotic initiation complex. Unlike transcription in prokaryotic cells, in which the RNA polymerase recognizes and binds to the promoter, eukaryotic transcription requires the binding of transcription factors to the promoter before RNA polymerase II binds to the DNA. The association of transcription factors and RNA polymerase II at the promoter is called the initiation complex.

expression! The promoters were found to actually be internal to the gene itself. This has not proved to be the case for all polymerase III genes, but appears to be for most.

Initiation and termination differ from that in prokaryotes

The initiation of transcription at RNA polymerase II promoters is analogous to prokaryotic initiation but is more complex. Instead of a single factor allowing promoter recognition, eukaryotes use a host of **transcription factors.** These proteins are necessary to get the RNA polymerase II enzyme to a promoter and to initiate gene expression. A number of these transcription factors interact with RNA polymerase II to form an *initiation complex* at the promoter (figure 15.10). We explore this complex in detail in chapter 16 when we describe the control of gene expression.

The termination of transcription for RNA polymerase II also differs from that in prokaryotes. Although termination sites exist, they are not as well defined as are prokaryotic terminators. The end of the mRNA is also not formed by RNA polymerase II because the primary transcript is modified after transcription.

Eukaryotic transcripts are modified

A primary difference between prokaryotes and eukaryotes is the fate of the transcript itself. Between transcription in the nucleus and export of a mature mRNA to the cytoplasm, a number of modifications occur to the initial transcripts made by RNA polymerase II. We call the RNA synthesized by RNA polymerase II the **primary transcript** and the final processed form the **mature mRNA.**

The 5′ cap

Eukaryotic transcripts have an unusual structure that is added to the 5′ end of mRNAs. The first base in the transcript is usually an adenine (A) or a guanine (G), and this is further modified by the addition of GTP to the 5′ PO_4 group, forming what is known as a **5′ cap** (figure 15.11). The G nucleotide in the cap is joined to the transcript by its 5′ end; the only such 5′-to-5′ bond found in nucleic acids. The G in the GTP is also modified by the addition of a methyl group, so it is often called a *methyl-G cap.* The cap is added while transcription is still in progress. This cap protects the 5′ end of the mRNA from degradation and is also involved in translation initiation.

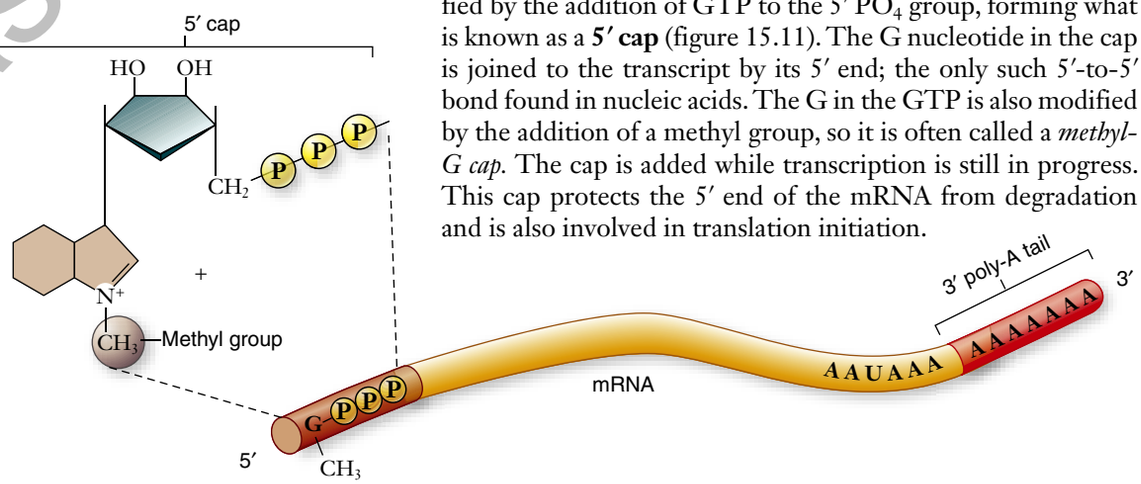


Figure 15.11 Posttranscriptional modifications to 5′ and 3′ ends. Eukaryotic mRNA molecules are modified in the nucleus with the addition of a methylated GTP to the 5′ end of the transcript, called the 5′ cap, and a long chain of adenine residues to the 3′ end of the transcript, called the 3′ poly-A tail.

The 3′ poly-A tail

A major difference between transcription in prokaryotes and eukaryotes is that in eukaryotes, the end of the transcript is not the end of the mRNA. The eukaryotic transcript is cleaved downstream of a specific site (AAUAAA) prior to the termination site for transcription. A series of adenine (A) residues, called the **3′ poly-A tail,** is added after this cleavage by the enzyme poly-A polymerase. Thus the end of the mRNA is not created by RNA polymerase II and is not the end of the transcript (see figure 15.11).

The enzyme poly-A polymerase is part of a complex that recognizes the poly-A site, cleaves the transcript, then adds 100–200 A's to the end. The poly-A tail appears to play a role in the stability of mRNAs by protecting them from degradation (see chapter 16).

Splicing of primary transcripts

Eukaryotic genes may contain noncoding sequences that have to be removed to produce the final mRNA. This process, called pre-mRNA splicing, is accomplished by an organelle called the *spliceosome*. This complex topic is discussed in the next section.

Learning Outcomes Review 15.4

Eukaryotes have three RNA polymerases called polymerase I, II, and III. Each synthesizes a different RNA and recognizes its own promoter. The RNA polymerase I promoter is species-specific. The polymerase II promoter is complex, but often includes a sequence called the TATA box. The polymerase III promoter is internal to the gene, rather than close to the 5′ end. Polymerase II is responsible for mRNA synthesis. The primary mRNA transcript is modified by addition of a 5′ cap and a 3′ poly-A tail consisting of 100–200 adenines. Noncoding regions are removed by splicing.

- ***Does the complexity of the eukaryotic genome require three polymerases?***

15.5 *Eukaryotic pre-mRNA Splicing*

Learning Outcomes

1. ***Explain the concept of colinearity of genes and proteins and how it applies in prokaryotes and eukaryotes.***
2. ***Describe the splicing reaction for pre-mRNA.***
3. ***Illustrate how splicing changes the nature of genes.***

The first genes isolated were prokaryotic genes found in *E. coli* and its viruses. A clear picture of the nature and some of the control of gene expression emerged from these systems before any eukaryotic genes were isolated. It was assumed that although details would differ, the outline of gene expression in eukaryotes would be similar. The world of biology was in for a shock with the isolation of the first genes from eukaryotic organisms.

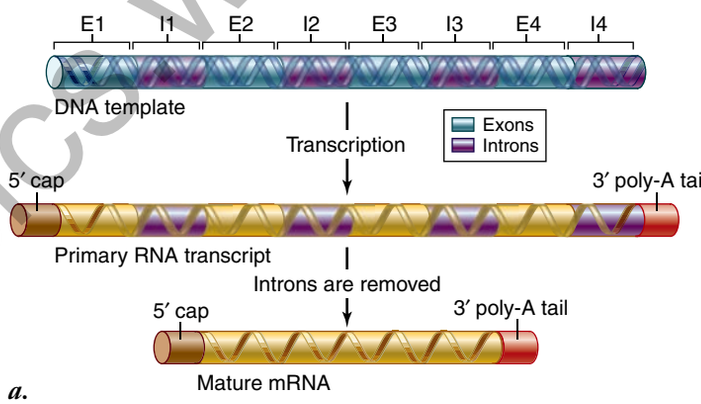
Eukaryotic genes may contain interruptions

Many eukaryotic genes appeared to contain sequences that were not represented in the mRNA. It is hard to exaggerate how unexpected this finding was. A basic tenet of molecular biology based on *E. coli* was that a gene was *colinear* with its protein product, that is, the sequence of bases in the gene corresponds to the sequence of bases in the mRNA, which in turn corresponds to the sequence of amino acids in the protein.

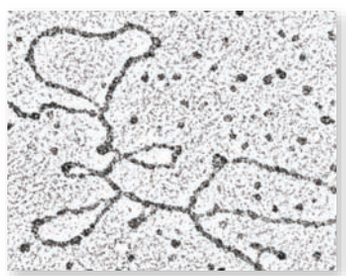
In the case of eukaryotes, genes can be interrupted by sequences that are not represented in the mRNA and the protein. The term "split genes" was used at the time, but the nomenclature that has stuck describes the unexpected nature of these sequences. We call the noncoding DNA that interrupts the sequence of the gene "intervening sequences," or **introns,** and we call the coding sequences **exons** because they are expressed (figure 15.12).

The spliceosome is the splicing organelle

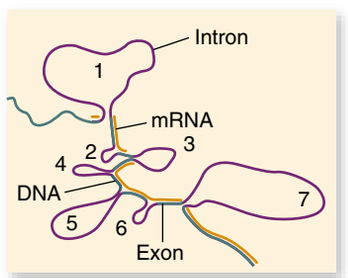
It is still true that the mature eukaryotic mRNA is colinear with its protein product, but a gene that contains introns is not. Imagine looking at an interstate highway from a satellite. Scattered randomly along the thread of concrete would be cars,



a.



b.



c.

Figure 15.12 Eukaryotic genes contain introns and exons. ***a.*** Eukaryotic genes contain sequences that form the coding sequence called exons and intervening sequences called introns. ***b.*** An electron micrograph showing hybrids formed with the mRNA and the DNA of the ovalbumin gene, which has seven introns. Introns within the DNA sequence have no corresponding sequence in the mRNA and thus appear as seven loops. ***c.*** A schematic drawing of the micrograph.

Inquiry question

How can the same gene encode different transcripts?

some moving in clusters, others individually; most of the road would be bare. That is what a eukaryotic gene is like—scattered exons embedded within much longer sequences of introns.

In humans, only 1 to 1.5% of the genome is devoted to the exons that encode proteins; 24% is devoted to the noncoding introns within which these exons are embedded.

The splicing reaction

The obvious question is—How do eukaryotic cells deal with the noncoding introns? The answer is that the primary transcript is cut and put back together to produce the mature mRNA. The latter process is referred to as **pre-mRNA splicing,** and it occurs in the nucleus prior to the export of the mRNA to the cytoplasm.

The intron–exon junctions are recognized by **small nuclear ribonucleoprotein particles,** called **snRNPs** (pronounced "snurps"). The snRNPs are complexes composed of snRNA and protein. These snRNPs then cluster together with other associated proteins to form a larger complex called the **spliceosome,** which is responsible for the splicing, or removal, of the introns.

For splicing to occur accurately, the spliceosome must be able to recognize intron–exon junctions. Introns all begin with the same 2-base sequence and end with another 2-base sequence that tags them for removal. In addition, within the intron there is a conserved A nucleotide, called the *branch point*, which is important for the splicing reaction (figure 15.13).

The splicing process begins with cleavage of the 5′ end of the intron. This 5′ end becomes attached to the 2′ OH of the branch point A, forming a branched structure called a *lariat* due to its resemblance to a cowboy's lariat in a rope (see figure 15.13). The 3′ end of the first exon is then used to displace the 3′ end of the intron, joining the two exons together and releasing the intron as a lariat.

The processes of transcription and RNA processing do not occur in a linear sequence, but are rather all part of a concerted process that produces the mature mRNA. The capping reaction occurs during transcription, as does the splicing process. The RNA polymerase II enzyme itself helps to recruit the other factors necessary for modification of the primary transcript, and in this way the process of transcription and pre-mRNA processing are coupled.

Distribution of introns

No rules govern the number of introns per gene or the sizes of introns and exons. Some genes have no introns; others may have 50. The sizes of exons range from a few nucleotides to 7500 nt, and the sizes of introns are equally variable. The presence of introns partly explains why so little of a eukaryotic genome is actually composed of "coding sequences" (see chapter 18 for results from the Human Genome Project).

One explanation for the existence of introns suggests that exons represent functional domains of proteins, and that the intron–exon arrangements found in genes represent the shuffling of these functional units over long periods of evolutionary time. This hypothesis, called *exon shuffling*, was proposed soon after the discovery of introns and has been the subject of much debate over the years.

The recent flood of genomic data has shed light on this issue by allowing statistical analysis of the placement of introns and on intron–exon structure. This analysis has provided support for the exon shuffling hypothesis for many genes; however, it is also clearly not universal, because all proteins do not show this kind of pattern. It is possible that introns do not have a single origin, and therefore cannot be explained by a single hypothesis.

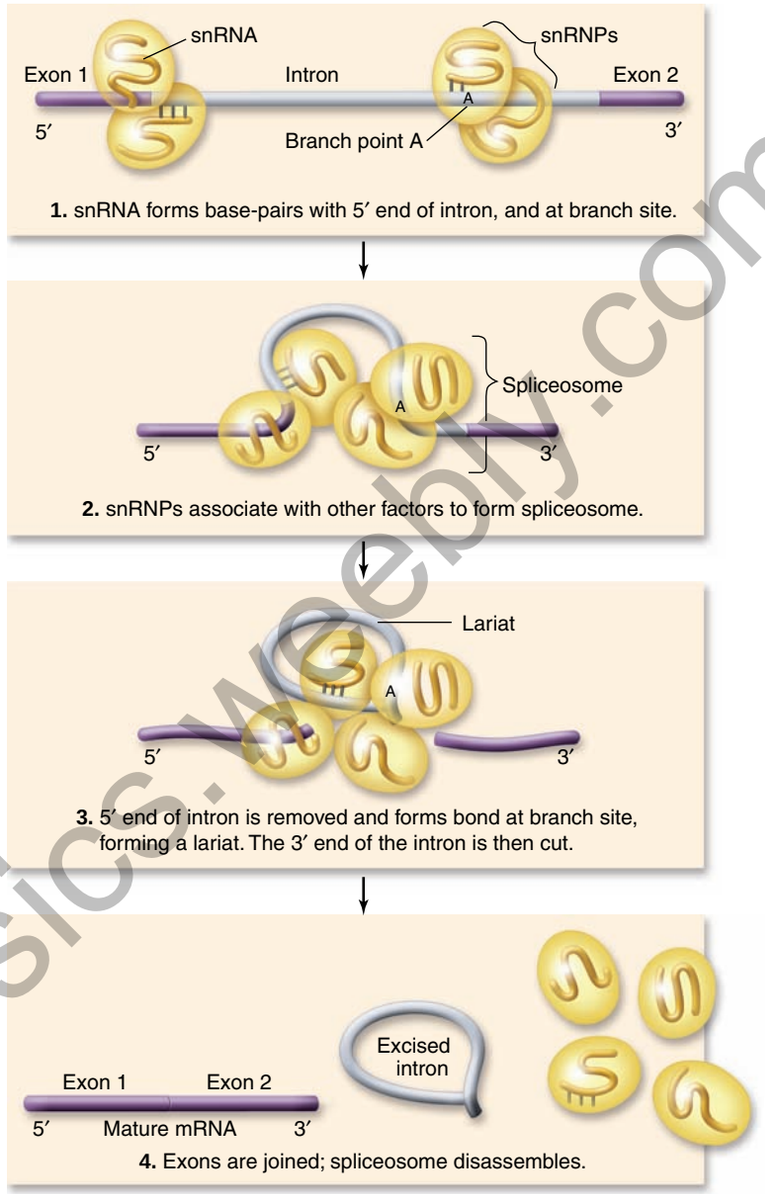


Figure 15.13 Pre-mRNA splicing by the spliceosome. Particles called snRNPs contain snRNA that interacts with the 5′ end of an intron and with a branch site internal to the intron. Several snRNPs come together with other proteins to form the spliceosome. As the intron forms a loop, the 5′ end is cut and linked to a site near the 3′ end of the intron. The intron forms a lariat that is excised, and the exons are spliced together. The spliceosome then disassembles and releases the spliced mRNA.

Splicing can produce multiple transcripts from the same gene

One consequence of the splicing process is greater complexity in gene expression in eukaryotes. A single primary transcript can be spliced into different mRNAs by the inclusion of different sets of exons, a process called **alternative splicing.**

Evidence indicates that the normal pattern of splicing is important to an organism's function. It has been estimated that 15% of known human genetic disorders are due to altered splicing. Mutations in the signals for splicing can introduce new splice sites or can abolish normal patterns of splicing. (In chapter 16 we consider how alternative splicing can be used to regulate gene expression.)

Although many cases of alternative splicing have been documented, the recent completion of the reference sequence of the human genome, along with other large data sets of expressed sequences, now allow large-scale comparisons between sequences found in mRNAs and in the genome. Three different computer-based analyses have been performed, producing results that are in rough agreement. These initial genomic assessments indicate a range of 35 to 59% for human genes that exhibit some form of alternative splicing. If we pick the middle ground of around 40%, this result still vastly increases the number of potential proteins encoded by the 25,000 genes in the human genome.

It is important to note that these analyses are primarily computer-based, and the functions of the possible spliced products have been investigated for only a small part of the potentially spliced genes. These analyses, however, do explain how the 25,000 genes of the human genome can encode the more than 80,000 different mRNAs reported to exist in human cells. The emerging field of proteomics addresses the number and functioning of proteins encoded by the human genome.

Learning Outcomes Review 15.5

In prokaryotes, genes appear to be colinear with their protein products. Eukaryotic genes, by contrast, contain exon regions, which are expressed, and intron sequences, which interrupt the exons. The introns are removed by the spliceosome in a process that leaves the exons joined together. Alternative splicing can generate different mRNAs, and thus different proteins, from the same gene. Recent estimates are that as many as half of human genes may be alternatively spliced.

- ***What advantages would alternative splicing confer on an organism?***

15.6 *The Structure of tRNA and Ribosomes*

Learning Outcomes

1. ***Explain why the tRNA charging reaction is critical to translation.***
2. ***Identify the tRNA-binding sites in the ribosome.***

The ribosome is the key organelle in translation, but it also requires the participation of mRNA, tRNA and a host of other factors. Critical to this process is the interaction of the ribosomes with tRNA and mRNA. To understand this, we first examine the structure of the tRNA adapter molecule and the ribosome itself.

Aminoacyl-tRNA synthetases attach amino acids to tRNA

Each amino acid must be attached to a tRNA with the correct anticodon for protein synthesis to proceed. This covalent attachment is accomplished by the action of activating enzymes called **aminoacyl-tRNA synthetases.** One of these enzymes is present for each of the 20 common amino acids.

tRNA structure

Transfer RNA is a bifunctional molecule that must be able to interact with mRNA and with amino acids. The structure of tRNAs is highly conserved in all living systems, and it can be formed into a cloverleaf type of structure based on intramolecular base-pairing that produces double-stranded regions. This primary structure is then folded in space to form an L-shaped molecule that has two functional ends: the **acceptor stem** and the **anticodon loop** (figure 15.14).

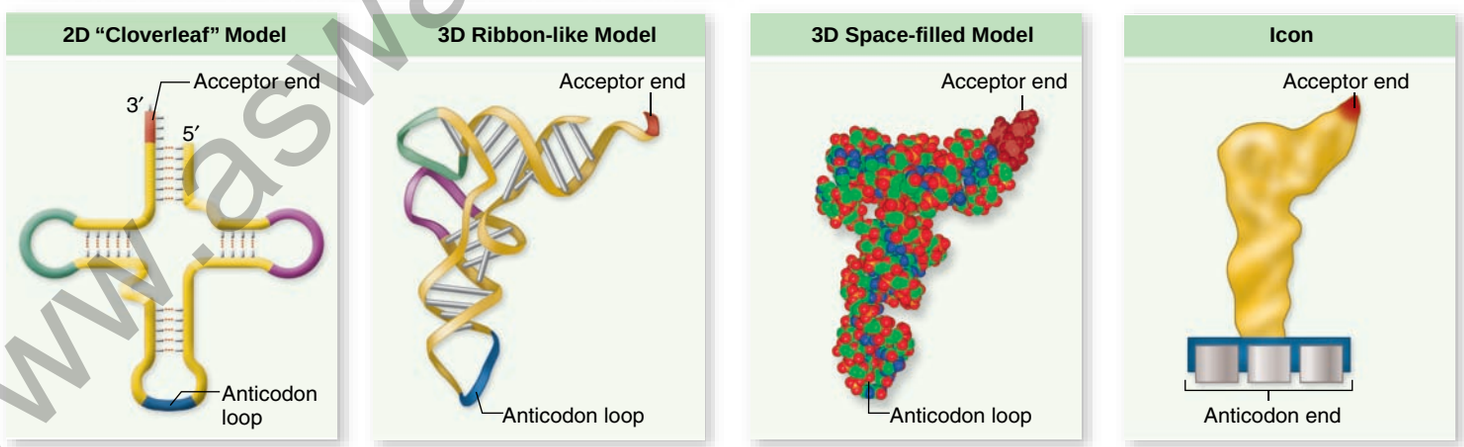


Figure 15.14 The structure of tRNA. Base-pairing within the molecule creates three stem and loop structures in a characteristic cloverleaf shape. The loop at the bottom of the cloverleaf contains the anticodon sequence, which can base-pair with codons in the mRNA. Amino acids are attached to the free, single-stranded —OH end of the acceptor stem. In its final three-dimensional structure, the loops of tRNA are folded into the final L-shaped structure.

The acceptor stem is the 3′ end of the molecule, which always ends in 5′ CCA 3′. The amino acid is attached to this end of the molecule. The anticodon loop is the bottom loop of the cloverleaf, and it can base-pair with codons in mRNA.

The charging reaction

The aminoacyl-tRNA synthetases must be able to recognize specific tRNA molecules as well as their corresponding amino acids. Although 61 codons code for amino acids, there are actually not 61 tRNAs in cells, although the number varies from species to species. Therefore, some aminoacyl-tRNA synthetases must be able to recognize more than one tRNA—but each recognizes only a single amino acid.

The reaction catalyzed by the enzymes is called the tRNA **charging reaction,** and the product is an amino acid joined to a tRNA, now called a *charged tRNA*. An ATP molecule provides energy for this endergonic reaction. The charged tRNA produced by the reaction is an activated intermediate that can undergo the peptide bond-forming reaction without an additional input of energy.

The charging reaction joins the acceptor stem to the carboxyl terminus of an amino acid (figure 15.15). Keeping this directionality in mind is critical to understanding the function of the ribosome, because each peptide bond will be formed between the amino group of one amino acid and the carboxyl group of another amino acid.

The correct attachment of amino acids to tRNAs is important because the ribosome does not verify this attachment. Ribosomes can only ensure that the codon–anticodon pairing is correct. In an elegant experiment, cysteine was converted chemically to alanine after the charging reaction, when the amino acid was already attached to tRNA. When this charged tRNA was used in an in vitro protein synthesis system, alanine was incorporated in the place of cysteine, showing that the ribosome cannot "proofread" the amino acids attached to tRNA.

In a very real sense, therefore, the charging reaction is the real translation step; amino acids are incorporated into a peptide based solely on the tRNA anticodon and its interaction with the mRNA.

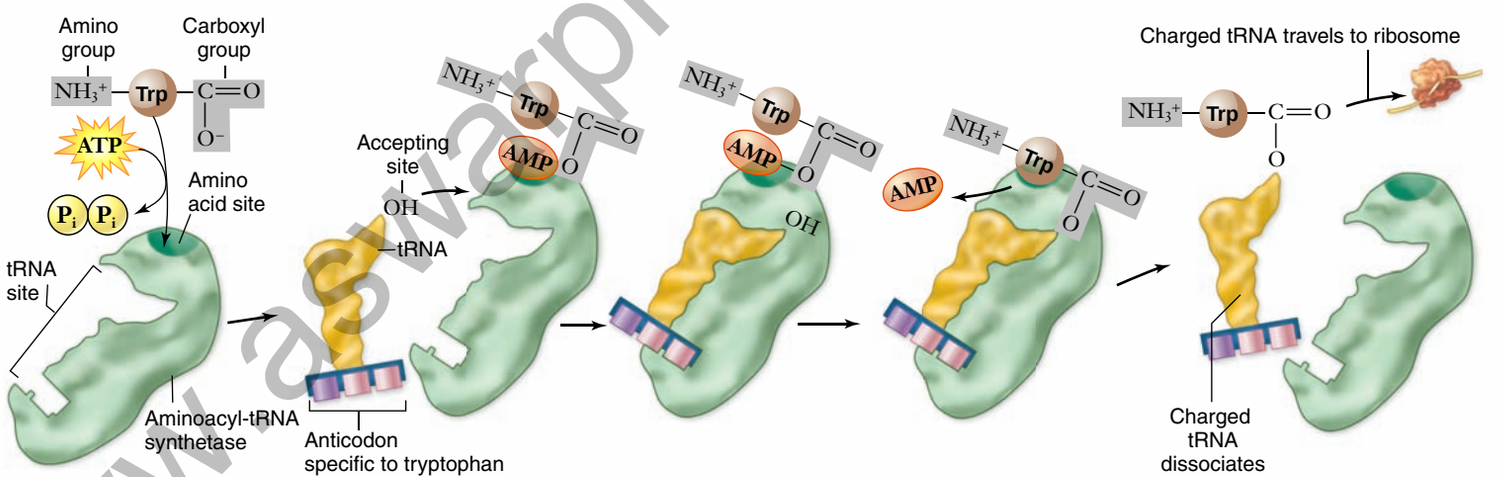
The ribosome has multiple tRNA-binding sites

The synthesis of any biopolymer can be broken down into initiation, elongation, and termination—you have seen this division for DNA replication as well as for transcription. In the case of translation, or protein synthesis, all three of these steps take place on the ribosome, a large macromolecular assembly consisting of rRNA and proteins. Details of the process by which the two ribosome subunits are assembled during initiation are described shortly.

For the ribosome to function it must be able to bind to at least two charged tRNAs at once so that a peptide bond can be formed between their amino acids, as described in the previous overview. The bacterial ribosome contains three binding sites, summarized in figure 15.16:

- The **P site** (peptidyl) binds to the tRNA attached to the growing peptide chain.
- The **A site** (aminoacyl) binds to the tRNA carrying the next amino acid to be added.
- The **E site** (exit) binds the tRNA that carried the previous amino acid added (see figure 15.16).

Transfer RNAs move through these sites successively during the process of elongation. Relative to the mRNA, the sites are arranged 5′ to 3′ in the order E, P, and A. The incoming



1. In the first step of the reaction, the amino acid is activated. The amino acid reacts with ATP to produce an intermediate with the carboxyl end of the amino acid attached to AMP. The two terminal phosphates (pyrophosphates) are cleaved from ATP in this reaction.

2. The amino acid-AMP complex remains bound to the enzyme. The tRNA next binds to the enzyme.

3. The second step of the reaction transfers the amino acid from AMP to the tRNA, producing a charged tRNA and AMP. The charged tRNA consists of a specific amino acid attached to the 3′ acceptor stem of its RNA.

Figure 15.15 tRNA charging reaction. There are 20 different aminoacyl-tRNA synthetase enzymes each specific for one amino acid, such as tryptophan (Trp). The enzyme must also recognize and bind to the tRNA molecules with anticodons specifying that amino acid, ACC for tryptophan. The reaction uses ATP and produces an activated intermediate that will not require further energy for peptide bond formation.

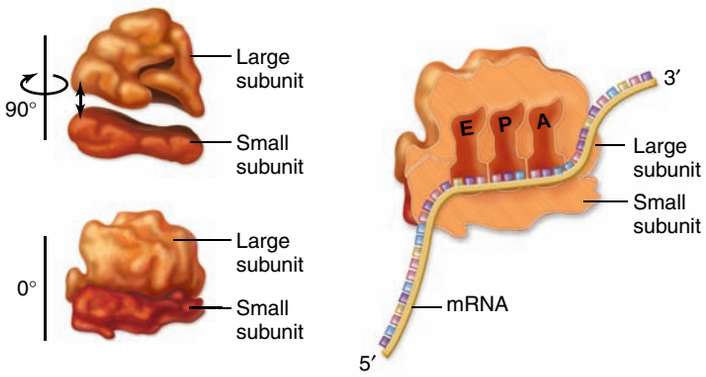


Figure 15.16 Ribosomes have two subunits.
Ribosome subunits come together and apart as part of a ribosome cycle. The smaller subunit fits into a depression on the surface of the larger one. Ribosomes have three tRNA-binding sites: aminoacyl site (A), peptidyl site (P), and empty site (E).

charged tRNAs enter the ribosome at the A site, transit through the P site, and then leave via the E site.

The ribosome has both decoding and enzymatic functions

The two functions of the ribosome involve decoding the transcribed message and forming peptide bonds. The decoding function resides primarily in the small subunit of the ribosome. The formation of peptide bonds requires the enzyme **peptidyl transferase,** which resides in the large subunit.

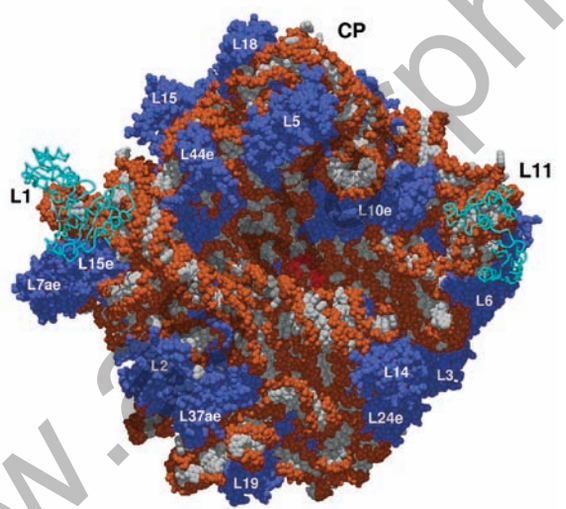


Figure 15.17 3-D structure of prokaryotic ribosome. The complete atomic structure of a prokaryotic large ribosomal subunit has been determined at 2.4-Å resolution. Bases of RNA are white, the polynucleotide backbone is red, and proteins are blue. The faces of each ribosomal subunit are lined with rRNA such that their interaction with tRNAs, amino acids, and mRNA all involve rRNA. Proteins are absent from the active site but abundant everywhere on the surface. The proteins stabilize the structure by interacting with adjacent RNA strands.

Our view of the ribosome has changed dramatically over time. Initially, molecular biologists assumed that the proteins in the ribosome carried out its function and that the rRNA was a structural scaffold necessary to hold the proteins in the correct position. Now this view has mostly been reversed; the ribosome is seen instead as rRNAs that are held in place by proteins. The faces of the two subunits that interact with each other are lined with rRNA, and the parts of both subunits that interact with mRNA, tRNA, and amino acids are also primarily rRNA (figure 15.17). It is now thought that the peptidyl transferase activity resides in an rRNA in the large subunit.

Learning Outcomes Review 15.6

Transfer RNA has two functional regions, one that bonds with an amino acid, and the other that can base-pair with mRNA. The tRNA charging reaction joins the carboxyl end of an amino acid to the 3′ acceptor stem of its tRNA; without charged tRNAs, translation cannot take place. This reaction is catalyzed by 20 different aminoacyl-tRNA synthetases, one for each amino acid. The ribosome has three different binding sites for tRNA, one for the tRNA adding to the growing peptide chain (P site), one for the next charged tRNA (A site), and one for the previous tRNA, which is now without an amino acid (E site). The ribosome can be thought of as having both a decoding function and an enzymatic function.

- ***What would be the effect on translation of a mutant tRNA that has an anticodon complementary to a STOP codon?***

15.7 The Process of Translation

Learning Outcomes

1. ***Distinguish between translation initiation and elongation.***
2. ***Explain the elongation cycle.***
3. ***Compare translation on the RER and in the cytoplasm.***

The process of translation is one of the most complex and energy-expensive tasks that cells perform. An overview of the process, as you saw earlier, is perhaps deceptively simple: The mRNA is threaded through the ribosome, while tRNAs carrying amino acids bind to the ribosome, where they interact with mRNA by base-pairing with the mRNA's codons. The ribosome and tRNAs position the amino acids such that peptide bonds can be formed between each new amino acid and the growing polypeptide.

Initiation requires accessory factors

As mentioned earlier, the start codon is AUG, which also encodes the amino acid methionine. The ribosome usually uses the first AUG it encounters in an mRNA strand to signal the start of translation.

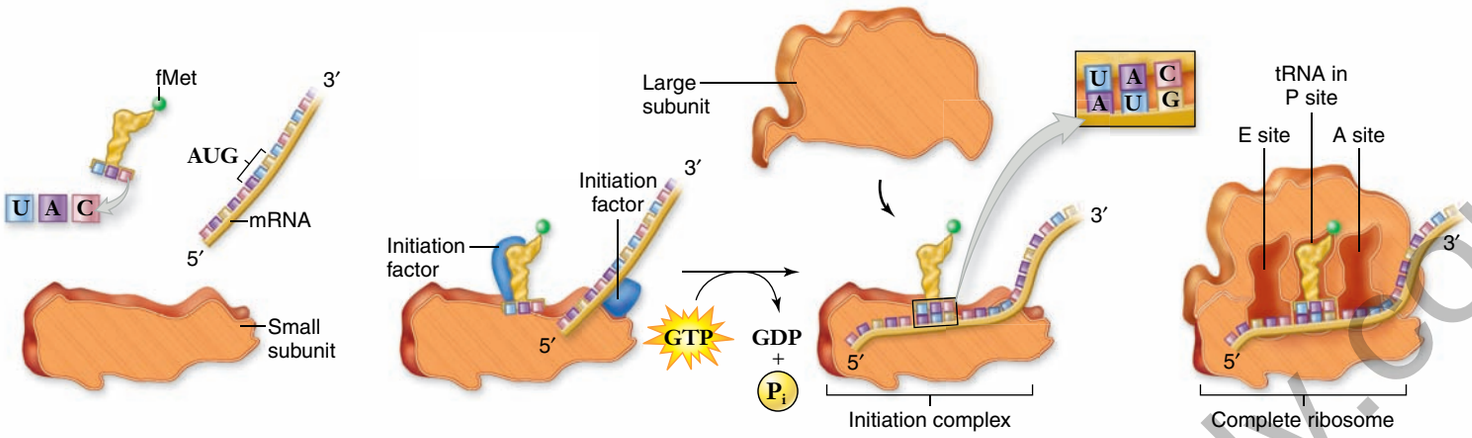


Figure 15.18 Initiation of translation. In prokaryotes, initiation factors play key roles in positioning the small ribosomal subunit, the initiator $tRNA^{fMet}$, and the mRNA. When the $tRNA^{fMet}$ is positioned over the first AUG codon of the mRNA, the large ribosomal subunit binds, forming the E, P, and A sites where successive tRNA molecules bind to the ribosomes, and polypeptide synthesis begins. Ribosomal subunits are shown as a cutaway sectioned through the middle.

Prokaryotic initiation

In prokaryotes, the **initiation complex** includes a special **initiator tRNA** molecule charged with a chemically modified methionine, *N-formylmethionine*. The initiator tRNA is shown as $tRNA^{fMet}$. The initiation complex also includes the small ribosomal subunit and the mRNA strand (figure 15.18). The small subunit is positioned correctly on the mRNA due to a conserved sequence in the 5′ end of the mRNA called the **ribosome-binding sequence (RBS)** that is complementary to the 3′ end of a small subunit rRNA.

A number of initiation factors mediate this interaction of the ribosome, mRNA, and $tRNA^{fMet}$ to form the initiation complex. These factors are involved in initiation only and are not part of the ribosome.

Once the complex of mRNA, initiator tRNA, and small ribosomal subunit is formed, the large subunit is added, and

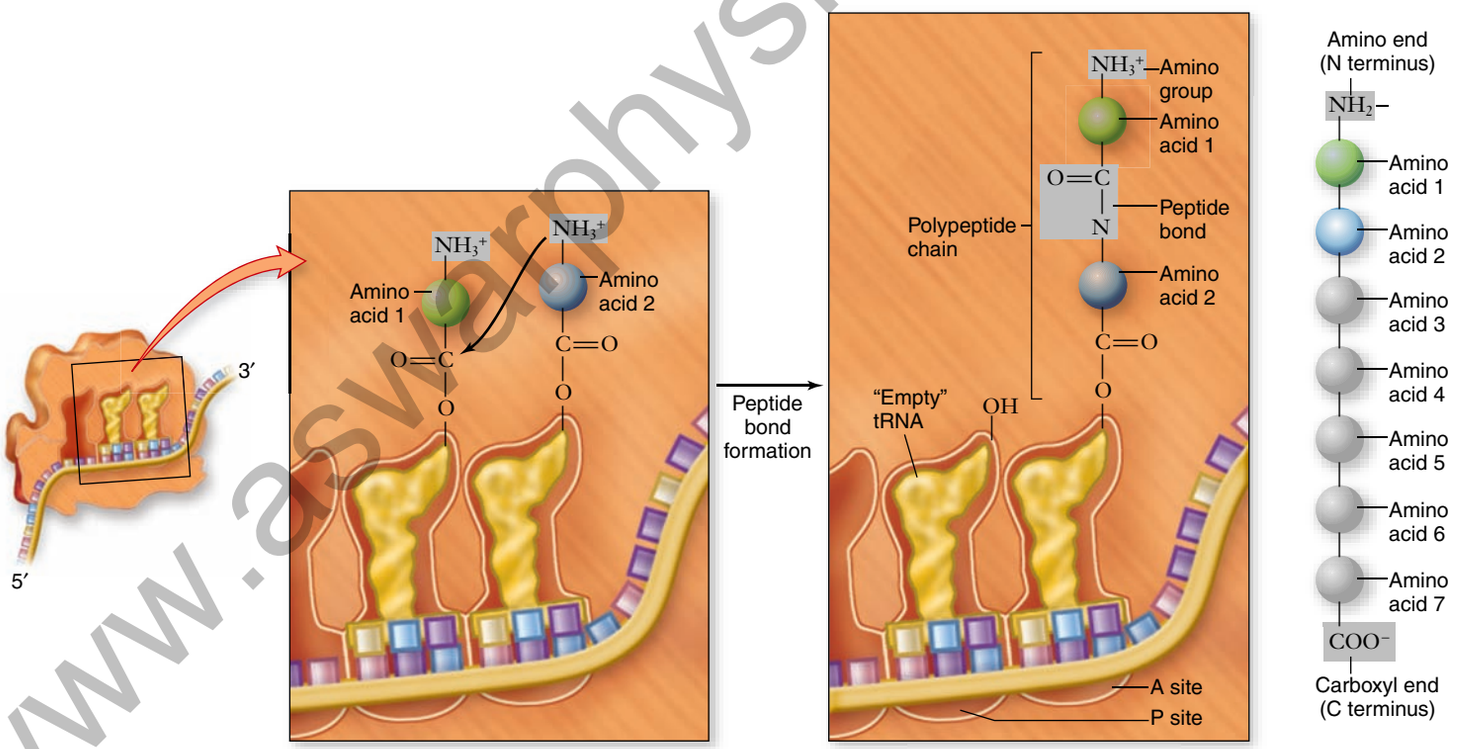


Figure 15.19 Peptide bond formation. Peptide bonds are formed between a "new" charged tRNA in the A site and the growing chain attached to the tRNA in the P site. The bond forms between the amino group of the new amino acid and the carboxyl group of the growing chain. This breaks the bond between the growing chain and its tRNA, transfering it to the A site as the new amino acid remains attached to its tRNA.

translation can begin. With the formation of the complete ribosome, the initiator tRNA is bound to the P site with the A site empty.

Eukaryotic initiation

Initiation in eukaryotes is similar, although it differs in two important ways. First, in eukaryotes, the initiating amino acid is methionine rather than *N*-formylmethionine. Second, the initiation complex is far more complicated than in prokaryotes, containing nine or more protein factors, many consisting of several subunits. Eukaryotic mRNAs also lack an RBS. The small subunit binds to the mRNA initially by binding to the 5′ cap of the mRNA.

Elongation adds successive amino acids

When the entire ribosome is assembled around the initiator tRNA and mRNA, the second charged tRNA can be brought to the ribosome and bind to the empty A site. This requires an **elongation factor** called **EF-Tu,** which binds to the charged tRNA and to GTP.

A peptide bond can then form between the amino acid of the initiator tRNA and the newly arrived charged tRNA in the A site. The geometry of this bond relative to the two charged tRNAs is critical to understanding the process. Remember that an amino acid is attached to a tRNA by its carboxyl terminus. The peptide bond is formed between the amino end of the incoming amino acid (in the A site) and the carboxyl end of the growing chain (in the P site) (figure 15.19).

The addition of successive amino acids is a series of events that occur in a cyclic fashion. Figure 15.20 shows the details of the elongation cycle.

1. **Matching tRNA anticodon with mRNA codon.** Each new charged tRNA comes to the ribosome bound to EF-Tu and GTP. The charged tRNA binds to the A site if its anticodon is complementary to the mRNA codon in the A site.

   After binding, GTP is hydrolyzed, and EF-Tu–GDP dissociates from the ribosome where it is recycled by another factor. This two-step binding and hydrolysis of GTP is thought to increase the accuracy of translation.

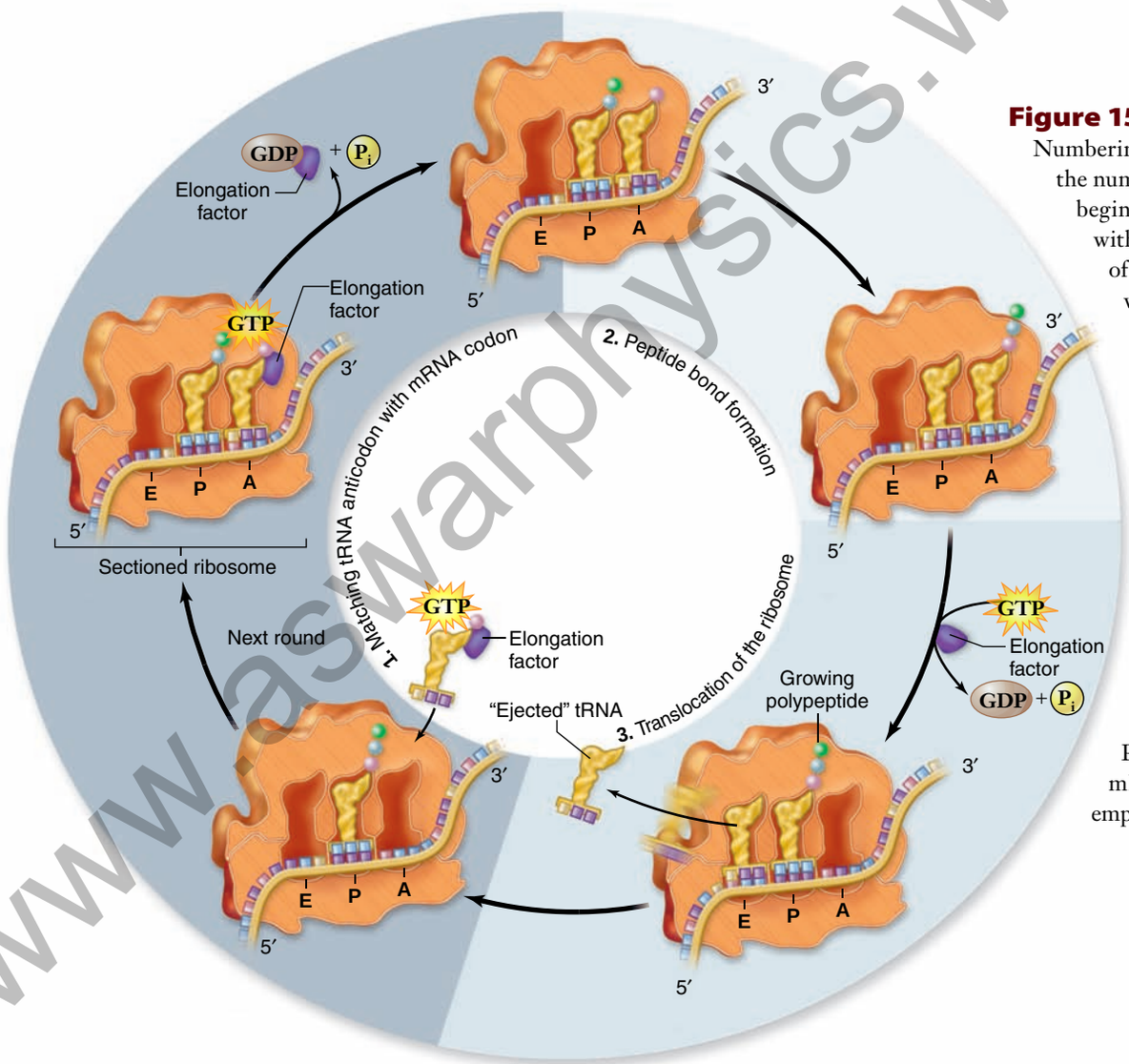


Figure 15.20 Elongation cycle. Numbering of the cycle corresponds to the numbering in the text. The cycle begins when a new charged tRNA with anticodon matching the codon of the mRNA in the A site arrives with EF-Tu. The EF-Tu hydrolyzes GTP and dissociates from the ribosome. A peptide bond is formed between the amino acid in the A site and the growing chain in the P site, transferring the growing chain to the A site, and leaving the tRNA in the P site empty. Ribosome translocation requires another elongation factor and GTP hydrolysis. This moves the tRNA in the A site into the P site, the next codon in the mRNA into the A site, and the empty tRNA into the E site.

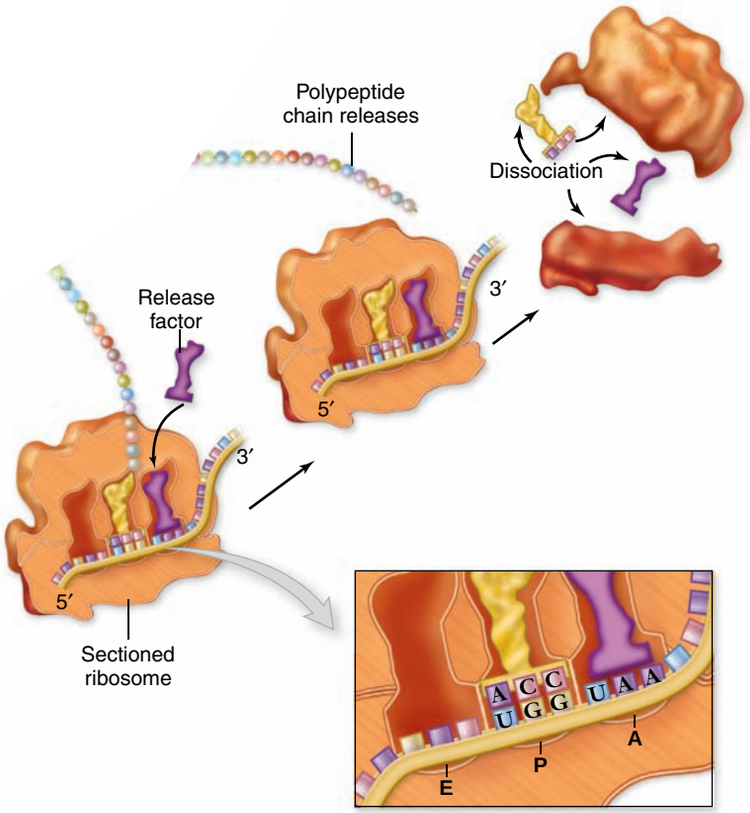


Figure 15.21 Termination of protein synthesis. There is no tRNA with an anticodon complementary to any of the three termination signal codons. When a ribosome encounters a termination codon, it stops translocating. A specific protein release factor facilitates the release of the polypeptide chain by breaking the covalent bond that links the polypeptide to the P site tRNA.

2. **Peptide bond formation.** Peptidyl transferase, located in the large subunit, catalyzes the formation of a peptide bond between the amino group of the amino acid in the A site and the carboxyl group of the growing chain. This also breaks the bond between the growing chain and the tRNA in the P site leaving it empty (no longer charged). The overall result of this is to transfer the growing chain to the tRNA in the A site.
3. **Translocation of the ribosome.** After the peptide bond has been formed, the ribosome moves relative to the mRNA and the tRNAs. The next codon in the mRNA shifts into the A site, and the tRNA with the growing chain moves to the P site. The uncharged tRNA formerly in the P site is now in the E site, and it will be ejected in the next cycle. This translocation step requires the accessory factor EF-G and the hydrolysis of another GTP.

This elongation cycle continues with each new amino acid added. The ribosome moves down the mRNA in a 5′-to-3′ direction, reading successive codons. The tRNAs move through the ribosome in the opposite direction, from the A site to the P site and finally the E site, before they are ejected as empty tRNAs, which can be charged with another amino acid and then used again.

Wobble pairing

As mentioned, there are fewer tRNAs than codons. This situation is easily rationalized because the pairing between the 3′ base of the codon and the 5′ base of the anticodon is less stringent than normal. In some tRNAs, the presence of modified bases with less accurate pairing in the 5′ position of the anticodon enhances this flexibility. This effect is referred to as

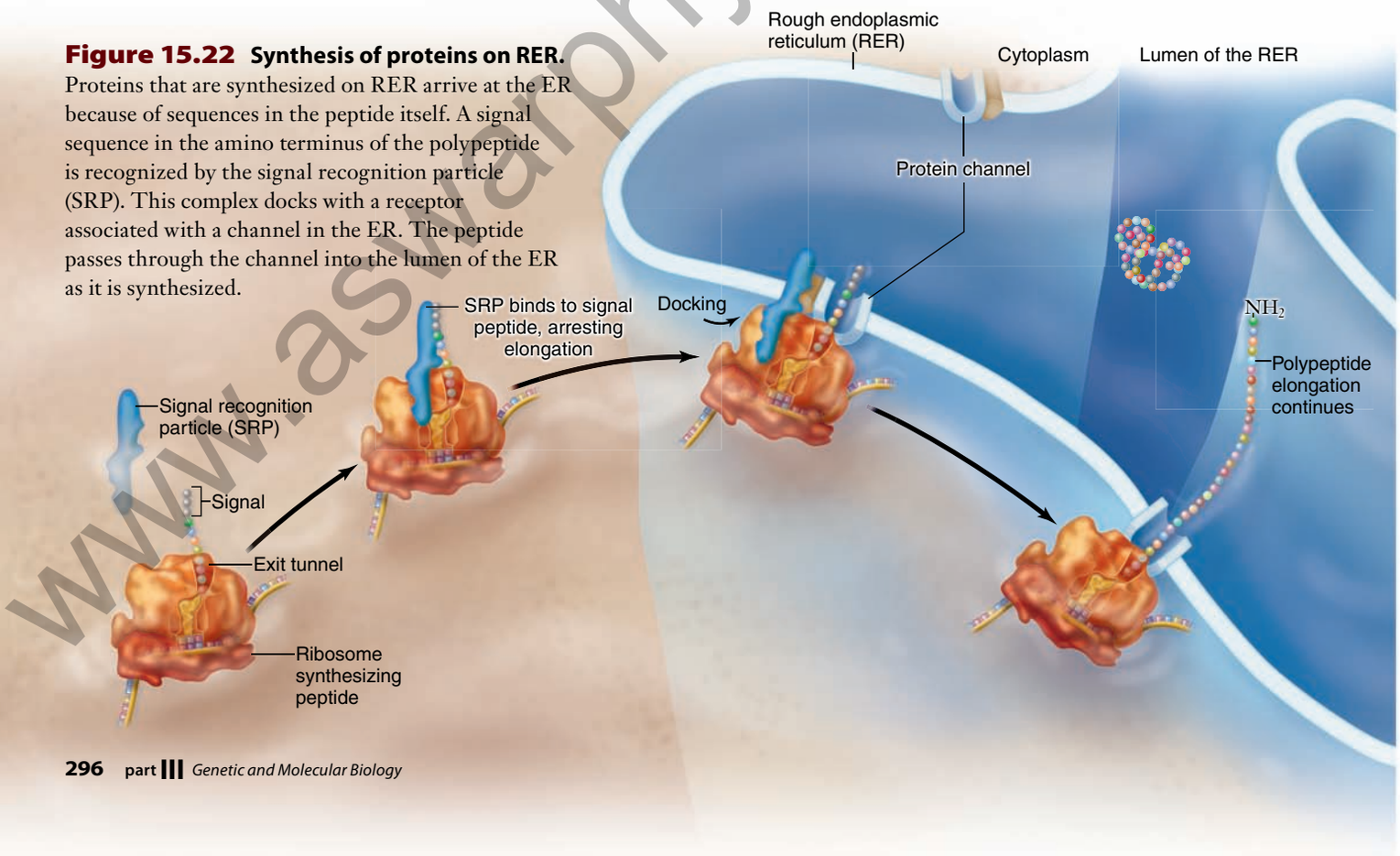


Figure 15.22 Synthesis of proteins on RER. Proteins that are synthesized on RER arrive at the ER because of sequences in the peptide itself. A signal sequence in the amino terminus of the polypeptide is recognized by the signal recognition particle (SRP). This complex docks with a receptor associated with a channel in the ER. The peptide passes through the channel into the lumen of the ER as it is synthesized.

wobble pairing because these tRNAs can "wobble" a bit on the mRNA, so that a single tRNA can "read" more than one codon in the mRNA.

Inquiry question

How is the wobble phenomenon related to the number of tRNAs and the degeneracy of the genetic code?

Termination requires accessory factors

Elongation continues in this fashion until a chain-terminating stop codon is reached (for example, UAA in figure 15.21). These stop codons do not bind to tRNA; instead, they are recognized by release factors, proteins that release the newly made polypeptide from the ribosome.

Proteins may be targeted to the ER

In eukaryotes, translation can occur either in the cytoplasm or on the RER. Proteins that are translated on the RER are targeted there based on their own initial amino acid sequence. The ribosomes found on the RER are actively translating and are not permanently bound to the ER.

A polypeptide that starts with a short series of amino acids called a **signal sequence** is specifically recognized and bound by a cytoplasmic complex of proteins called the *signal recognition particle (SRP)*. The complex of signal sequence and SRP is in turn recognized by a receptor protein in the ER membrane. The binding of the ER receptor to the signal sequence/SRP complex holds the ribosome engaged in translation of the protein on the ER membrane, a process called *docking* (figure 15.22).

As the protein is assembled, it passes through a channel formed by the docking complex and into the interior ER compartment, the cisternal space. This is the basis for the docking metaphor—the ribosome is not actually bound to the ER itself, but with the newly synthesized protein entering the ER, the ribosome is like a boat tied to a dock with a rope.

The basic mechanism of protein translocation across membranes by the SRP and its receptor and channel complex has been conserved across all three cell types: eukaryotes, bacteria, and archaea. Given that only eukaryotic cells have an endomembrane system, this universality may seem curious; however, bacteria and archaea both export proteins through their plasma membrane, and the mechanism used is similar to the way in which eukaryotes move proteins into the cisternal space of the ER.

Once within the ER cisternal space, or lumen, the newly synthesized protein can be modified by the addition of sugars (glycosylation) and transported by vesicles to the Golgi apparatus (see chapter 4). This is the beginning of the protein-trafficking pathway that can lead to other intracellular targets, to incorporation into the plasma membrane, or to release outside of the cell itself.

Learning Outcomes Review 15.7

Translation initiation involves the interaction of the small ribosomal subunit with mRNA and a charged initiator tRNA. The elongation cycle involves bringing in new charged tRNAs to the ribosome's A site, forming peptide bonds between amino acids, and translocating the ribosome along the mRNA chain. The tRNAs transit through the ribosome from A to P to E sites during the process. In eukaryotes, signal sequences of a newly forming polypeptide may target it and its ribosome to be moved to the RER. Polypeptides formed on the RER enter the cisternal space rather than being released into the cytoplasm.

- ***What stages of translation require energy?***

15.8 Summarizing Gene Expression

Because of the complexity of the process of gene expression, it is worth stepping back to summarize some key points:

- The process of gene expression converts information in the genotype into the phenotype.
- A copy of the gene in the form of mRNA is produced by transcription, and the mRNA is used to direct the synthesis of a protein by translation.
- Both transcription and translation can be broken down into initiation, elongation, and termination cycles that produce their respective polymers. (The same is true for DNA replication.)
- Eukaryotic gene expression is much more complex than that of prokaryotes.

The nature of eukaryotic genes with their intron and exon components greatly complicates the process of gene expression by requiring additional steps between transcription and translation. The production and processing of eukaryotic mRNAs also takes place in the nucleus, whereas translation takes place in the cytoplasm. This necessitates the transport of the mRNA through the nuclear pores to the cytoplasm before translation can take place. The entire eukaryotic process is summarized in figure 15.23.

A number of differences can be highlighted between gene expression in prokaryotes and in eukaryotes. Table 15.2 (on p. 298) summarizes these main points.

Learning Outcome Review 15.8

The greater complexity of eukaryotic gene expression is related to the functional organization of the cell, with DNA in the nucleus and ribosomes in the cytoplasm. The differences in gene expression between prokaryotes and eukaryotes is mainly in detail, but some differences have functional significance.

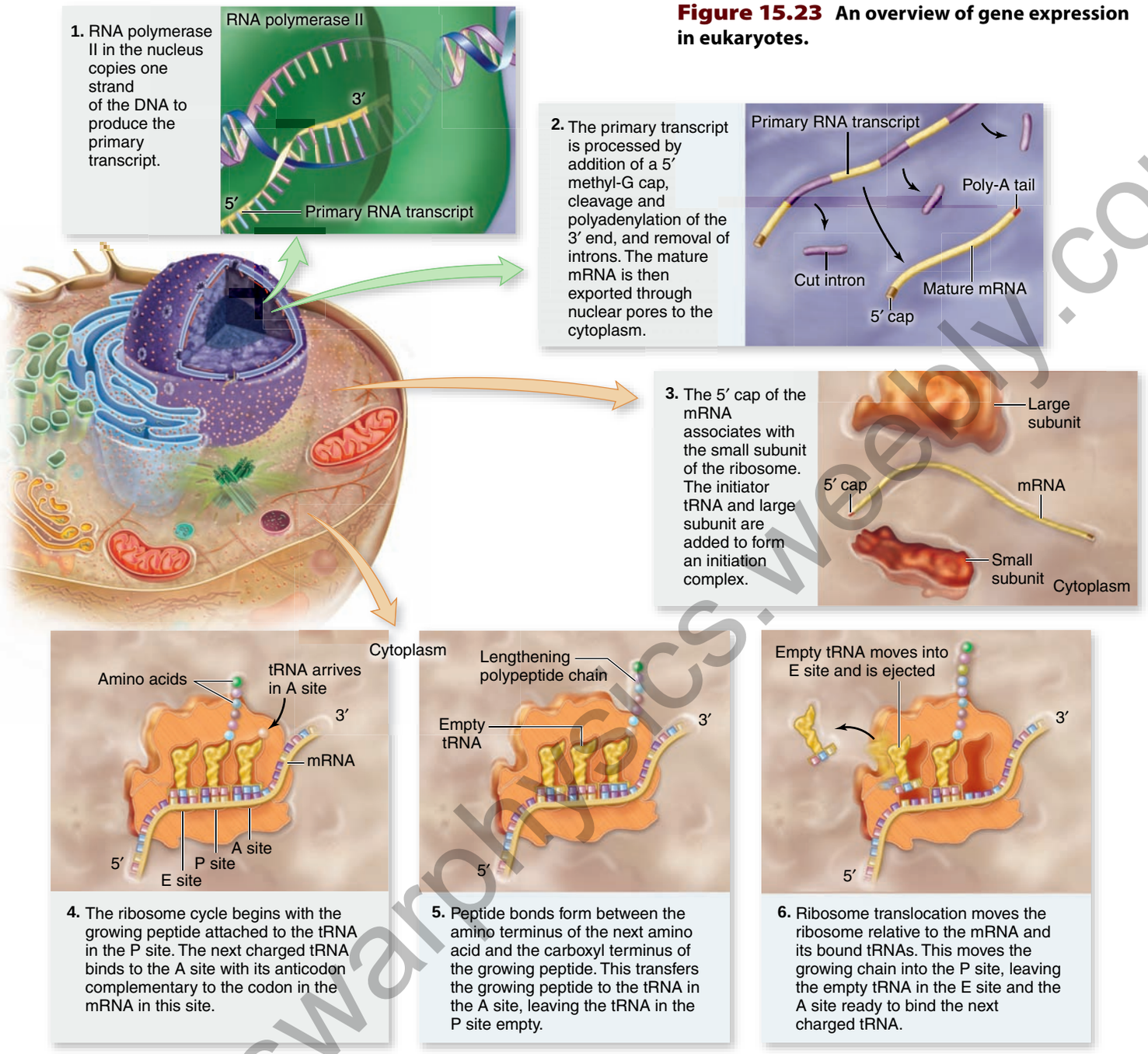


Figure 15.23 An overview of gene expression in eukaryotes.

TABLE 15.2 Differences Between Prokaryotic and Eukaryotic Gene Expression

Characteristic	Prokaryotes	Eukaryotes
Introns	No introns, although some archaeal genes possess them.	Most genes contain introns.
Number of genes in mRNA	Several genes may be transcribed into a single mRNA molecule. Often these have related functions and form an operon, which helps coordinate regulation of biochemical pathways.	Only one gene per mRNA molecule; regulation of pathways accomplished in other ways.
Site of transcription and translation	No membrane-bounded nucleus, transcription and translation are coupled.	Transcription in nucleus; mRNA is transported to the cytoplasm for translation.
Initiation of translation	Begins at AUG codon preceded by special sequence that binds the ribosome.	Begins at AUG codon preceded by the 5′ cap (methylated GTP) that binds the ribosome.
Modification of mRNA after transcription	None; translation begins before transcription is completed. Transcription and translation are coupled.	A number of modifications while the mRNA is in the nucleus: Introns are removed and exons are spliced together; a 5′ cap is added; a poly-A tail is added.

15.9 Mutation: Altered Genes

Learning Outcomes

1. *Describe the effects of different point mutations.*
2. *Explain the nature of triplet repeat expansion.*
3. *List the different chromosomal mutations and their effects.*

One way to analyze the function of genes is to find or to induce mutations in a gene to see how this affects its function. In terms of the organism, however, inducing mutations is usually negative; most mutations have deleterious effects on the phenotype of the organism. In chapter 13, you saw how a number of genetic diseases, such as sickle cell anemia, are due to single base changes. We now consider mutations from the perspective of how the DNA itself is altered. Mutational changes range from the alteration of a single base to the loss of genetic material (deletion) to the loss of an entire chromosome. The change of a single base can result in changing a single amino acid in a protein, and this in turn can lead to a debilitating clinical phenotype. This is illustrated for the case of sickle cell anemia in figure 15.24. In the sickle cell allele, a single A is changed to a T resulting in a glutamic acid being replaced with a valine. The substitution of nonpolar valine causes the beta chains to aggregate into polymers, and this alters the shape of the cells, leading to the disease state.

Point mutations affect a single site in the DNA

A mutation that alters a single base is termed a **point mutation.** The mutation can be either the substitution of one base for another, or the deletion or addition of a single base (or a small number of bases).

Base substitution

The substitution of one base pair for another in DNA is called a **base substitution mutation.** Because of the degenerate nature of the genetic code, base substitution may or may not alter the amino acid encoded. If the new codon from the base substitution still encodes the same amino acid, we say the mutation is *silent* (figure 15.25*b*). When base substitution changes an amino acid in a protein, it is also called a **missense mutation** as the "sense" of the codon produced after transcription of the mutant gene will be altered (figure 15.25*c*). These fall into two classes, *transitions* and *transversions*. A transition does not change the type of bases in the base pair, that is, a pyrimidine is substituted for a pyrimidine, or purine for purine. In contrast, a transversion does change the type of bases in a base pair, that is, pyrimidine to purine or the reverse. A variety of human genetic diseases, including sickle cell anemia, are caused by base substitutions.

Nonsense mutations

A special category of base substitution arises when a base is changed such that the transcribed codon is converted to a stop codon (see figure 15.25*d*). We call these **nonsense mutations** because the mutation does not make "sense" to the translation apparatus. The stop codon results in premature termination of translation and leads to a truncated protein. How short the resulting protein is depends on where a stop codon has been introduced in the gene.

Frameshift mutations

The addition or deletion of a single base has much more profound consequences than does the substitution of one base for another. These mutations are called *frameshift mutations* because they alter the reading frame in the mRNA downstream of the mutation. This class of mutations was used by Crick and Brenner, as described earlier in the chapter, to infer the nature of the genetic code.

Changing the reading frame early in a gene, and thus in its mRNA transcript, means that the majority of the protein will be altered. Frameshifts also can cause premature termination of translation because 3 in 64 codons are stop codons, which represents a high probability in the sequence that has been randomized by the frameshift.

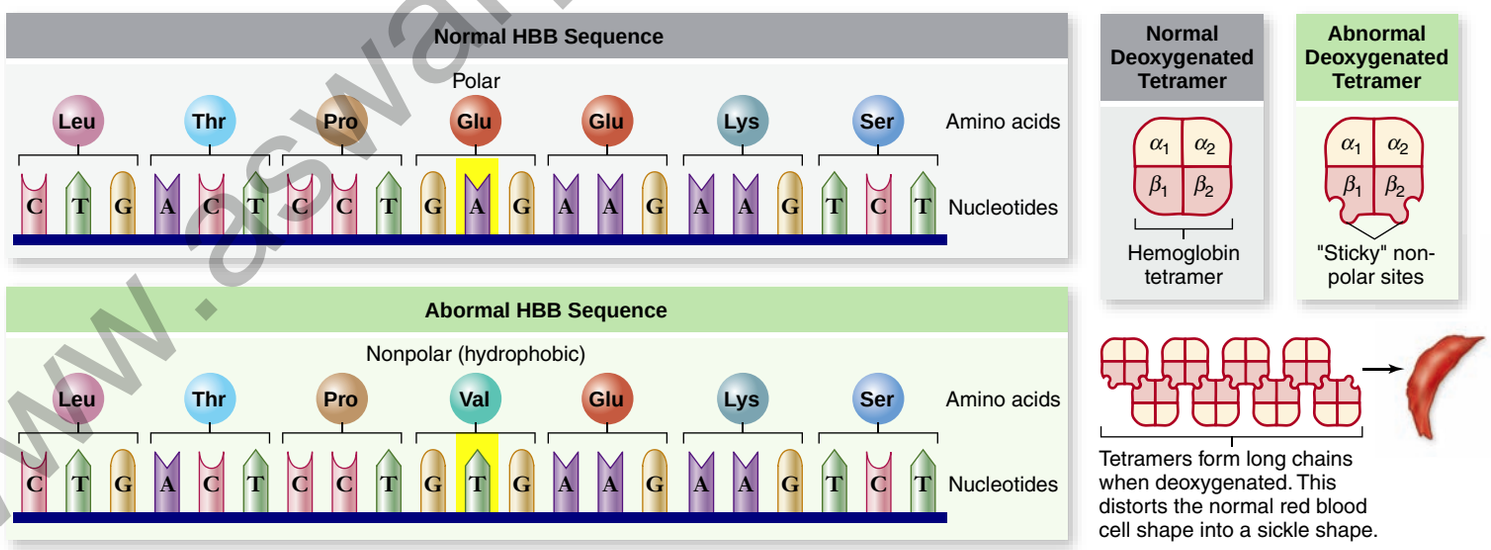
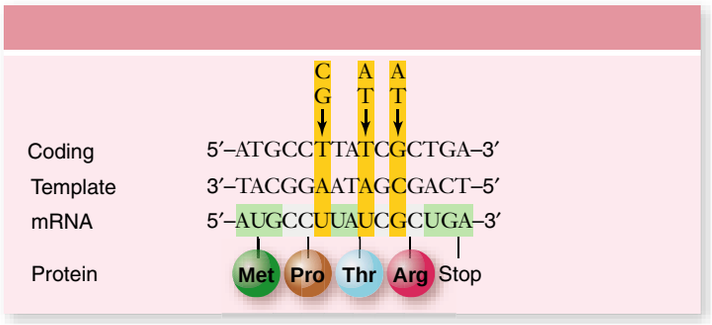
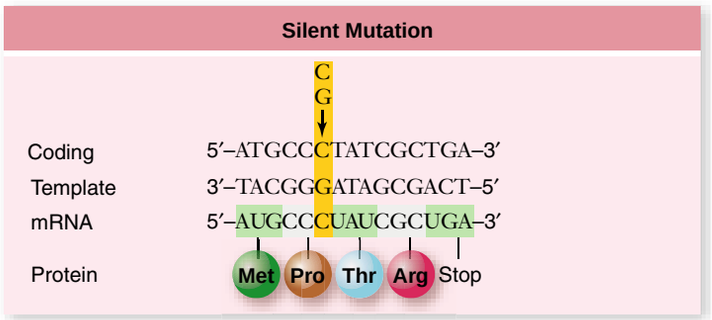


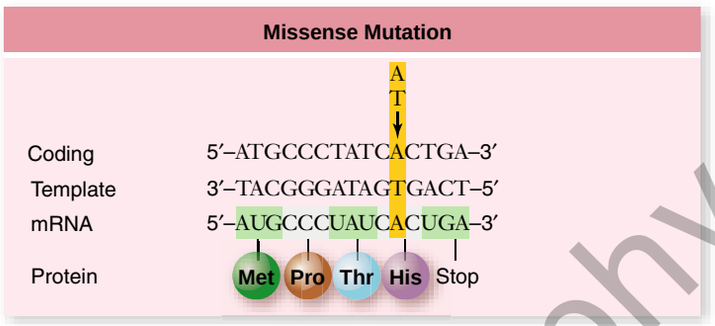
Figure 15.24 Sickle cell anemia is caused by an altered protein. Hemoglobin is composed of a tetramer of two α-globin and two β-globin chains. The sickle cell allele of the β-globin gene contains a single base change resulting in the substitution of Val for Glu. This creates a hydrophobic region on the surface of the protein that is "sticky" leading to their association into long chains that distort the shape of the red blood cells.



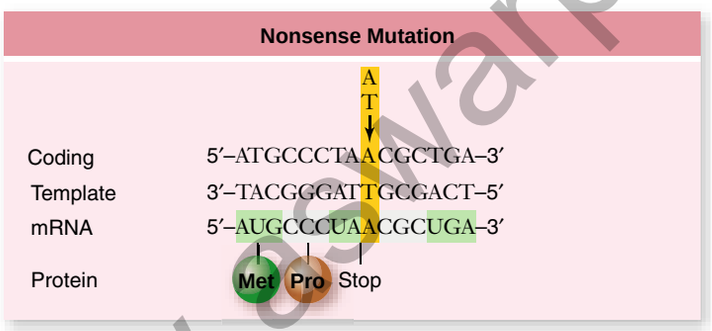
a.



b.



c.



d.

Figure 15.25 Types of mutations. ***a.*** A hypothetical gene is shown with encoded mRNA and protein. Arrows above the gene indicate sites of mutations described in the rest of the figure. ***b.*** Silent mutation. A change in the third position of a codon is often silent due to degeneracy in the genetic code. In this case T/A to C/G mutation does not change the amino acid encoded (proline). ***c.*** Missesense mutation. The G/C to A/T mutation changes the amino acid encoded from arginine to histidine. ***d.*** Nonsense mutation. The T/A to A/T mutation produces a UAA stop codon in the mRNA.

Triplet repeat expansion mutations

Given the long history of molecular genetics, and the relatively short time that molecular analysis has been possible on humans, it is surprising that a new kind of mutation was discovered in humans. However, one of the first genes isolated that was associated with human disease, the gene for *Huntington disease*, provided a new kind of mutation. The gene for Huntington contains a triplet sequence of DNA that is repeated, and this repeat unit is expanded in the disease allele relative to the normal allele. Since this initial discovery, at least 20 other human genetic diseases appear to be due to this mechanism. The prevalence of this kind of mutation is unknown, but at present humans and mice are the only organisms in which they have been observed, implying that they may be limited to vertebrates, or even mammals. No such mutation has ever been found in *Drosophila* for example.

The expansion of the triplet can occur in the coding region or in noncoding transcribed DNA. In the case of Huntington disease, the repeat unit is actually in the coding region of the gene where the triplet encodes glutamine, and expansion results in a polyglutamine region in the protein. A number of other neurodegenerative disorders also show this kind of mutation. In the case of fragile-X syndrome, an inherited form of mental retardation, the repeat is in noncoding DNA.

Chromosomal mutations change the structure of chromosomes

Point mutations affect a single site in a chromosome, but more extensive changes can alter the structure of the chromosome itself, resulting in **chromosomal mutations.** Many human cancers are associated with chromosomal abnormalities, so these are of great clinical relevance. We briefly consider possible alterations to chromosomal structure, all of which are summarized in figure 15.26.

Deletions

A **deletion** is the loss of a portion of a chromosome. Frameshifts can be caused by one or more small deletions, but much larger regions of a chromosome may also be lost. If too much information is lost, the deletion is usually fatal to the organism.

One human syndrome that is due to a deletion is *cri-du-chat*, which is French for "cry of the cat" after the noise made by children with this syndrome. Cri-du-chat syndrome is caused by a large deletion from the short arm of chromosome 5. It usually results in early death, although many affected individuals show a normal lifespan. It has a variety of effects, including respiratory problems.

Duplications

The **duplication** of a region of a chromosome may or may not lead to phenotypic consequences. Effects depend upon the location of the "breakpoints" where the duplication occurred. If the duplicated region does not lie within a gene, there may be no effect. If the duplication occurs next to the original region, it is termed a *tandem duplication*. These tandem duplications are important in the evolution of families of related genes, such as the globin family that encode the protein hemoglobin.

Inversions

An **inversion** results when a segment of a chromosome is broken in two places, reversed, and put back together. An inversion

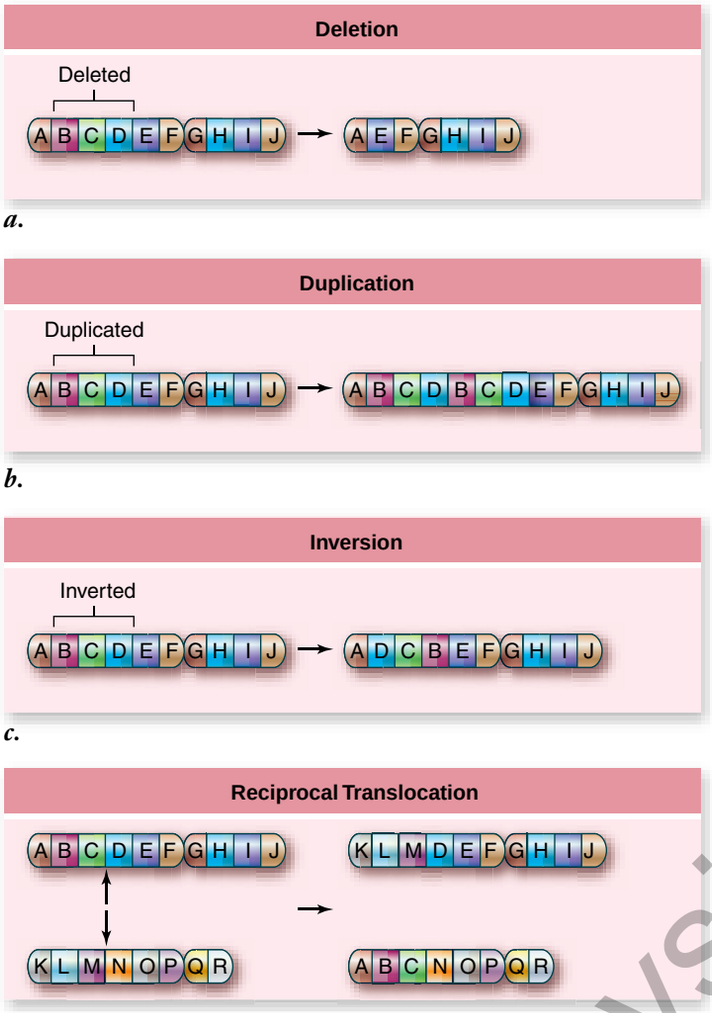


Figure 15.26 Chromosomal mutations. Larger-scale changes in chromosomes are also possible. Material can be deleted ***(a)***, duplicated ***(b)***, and inverted ***(c)***. Translocations occur when one chromosome is broken and becomes part of another chromosome. This often occurs where both chromosomes are broken and exchange material, an event called a reciprocal translocation ***(d)***.

may not have an effect on phenotype if the sites where the inversion occurs do not break within a gene. In fact, although humans all have the "same" genome, the order of genes in all individuals in a population is not precisely the same due to inversions that occur in different lineages.

Translocations

If a piece of one chromosome is broken off and joined to another chromosome, we call this a **translocation.** Translocations are complex because they can cause problems during meiosis, particularly when two different chromosomes try to pair with each other during meiosis I.

Translocations can also move genes from one chromosomal region to another in a manner that changes the expression of genes in the region involved. Two forms of leukemia have been shown to be associated with translocations that move oncogenes into regions of a chromosome where they are expressed inappropriately in blood cells.

Mutations are the starting point of evolution

If no changes occurred in genomes over time, then there could be no evolution. Too much change, however, is harmful to the individual with a greatly altered genome. Thus a delicate balance must exist between the amount of new variation that arises in a species and the health of individuals in the species. This topic is explored in more detail later in the book when we consider evolution and population genetics (chapter 20).

The larger scale alteration of chromosomes has also been important in evolution, although its role is poorly understood. It is clear that gene families arise by the duplication of an ancestral gene, followed by the functional divergence of the duplicated copies. It is also clear that even among closely related species, the number and arrangements of genes on chromosomes can differ. Large-scale rearrangements may have occurred.

Our view of the nature of genes has changed with new information

In this and the preceding chapters, we have seen multiple views of genes. Mendel used crosses to follow traits determined by what we now call genes. The behavior of these genes can be predicted based on the behavior of chromosomes during meiosis. Morgan and others learned to map the location of genes on chromosomes. These findings led to the view of genes as abstract entities that could be followed through generations and mapped to chromosomal locations like "beads on a string," with the beads being genes and the string the chromosome.

The original molecular analysis of genes led to the simple one-gene/one-polypeptide paradigm. This oversimplification was changed when geneticists observed the alternative splicing of eukaryotic genes, which can lead to multiple protein products from the same genetic information. Furthermore, some genes do not encode proteins at all, but only RNA, which can either be a part of the gene expression machinery (rRNA, tRNA, and other forms) or can itself act as an enzyme. Other stretches of DNA are important for regulating genes but are not expressed. All of these findings make a simple definition of genes difficult.

We are left with the rich complexity of the nature of genes, which defies simple definition. To truly understand the nature of genes we must consider both their molecular nature as well as their phenotypic expression. This brings us full circle, back to the relationship between genotype and phenotype, with a much greater appreciation for the complexity of this relationship.

Learning Outcomes Review 15.9

Point mutations (single-base changes, additions, or deletions) include missense mutations that cause substitution of one amino acid for another, nonsense mutations that halt transcription, and frameshift mutations that throw off the correct reading of codons. Triplet repeat expansion is the abnormal duplication of a codon with each round of cell division. Mutations affecting chromosomes include deletions, duplications, inversions, and translocations.

- ***Would an inversion or duplication always be expected to have a phenotype?***

15.1 The Nature of Genes

Garrod concluded that inherited disorders can involve specific enzymes.

Garrod found that alkaptonuria is due to an altered enzyme.

Beadle and Tatum showed that genes specify enzymes.

Neurospora mutants unable to synthesize arginine were found to lack specific enzymes. Beadle and Tatum advanced the "one gene/one polypeptide" hypothesis (figure 15.1).

The central dogma describes information flow in cells as DNA to RNA to protein (figure 15.2).

We call the DNA strand copied to mRNA the template (antisense) strand; the other the coding (sense) strand.

Transcription makes an RNA copy of DNA.

Translation uses information in RNA to synthesize proteins.

An adapter molecule, tRNA, is required to connect the information in mRNA into the sequence of amino acids.

RNA has multiple roles in gene expression.

15.2 The Genetic Code

The code is read in groups of three.

Crick and Brenner showed that the code is nonoverlapping and is read in groups of three. This finding established the concept of reading frame.

Nirenberg and others deciphered the code.

A codon consists of 3 nucleotides, so there are 64 possible codons. Three codons signal "stop," and one codon signals "start" and also encodes methionine. Thus 61 codons encode the 20 amino acids.

The code is degenerate but specific.

Many amino acids have more than one codon, but each codon specifies only a single amino acid.

The code is practically universal, but not quite.

In some mitochondrial and protist genomes, a STOP codon is read as an amino acid; otherwise the code is universal.

15.3 Prokaryotic Transcription

Prokaryotes have a single RNA polymerase.

Prokaryotic RNA polymerase exists in two forms: core polymerase, which can synthesize mRNA; and holoenzyme, core plus σ factor, which can accurately initiate synthesis (figure 15.6).

Initiation occurs at promoters.

Initiation requires a start site and a promoter. The promoter is upstream of the start site, and binding of RNA polymerase holoenzyme to its –35 region positions the polymerase properly.

Elongation adds successive nucleotides.

Transcription proceeds in the 5′-to-3′ direction. The transcription bubble contains RNA polymerase, the locally unwound DNA template, and the growing mRNA transcript (figure 15.7).

Termination occurs at specific sites.

Terminators consist of complementary sequences that form a double-stranded hairpin loop where the polymerase pauses (figure 15.8).

Prokaryotic transcription is coupled to translation.

Translation begins while mRNAs are still being transcribed.

15.4 Eukaryotic Transcription

Eukaryotes have three RNA polymerases.

RNA polymerase I transcribes rRNA; polymerase II transcribes mRNA and some snRNAs; polymerase III transcribes tRNA.

Each polymerase has its own promoter.

Initiation and termination differ from that in prokaryotes.

Unlike prokaryotic promoters, RNA polymerase II promoters require a host of transcription factors. Although termination sites exist, the end of the mRNA is modified after transcription.

Eukaryotic transcripts are modified (figure 15.11).

After transcription, a methyl-GTP cap is added to the 5′ end of the transcript. A poly-A tail is added to the 3′ end. Noncoding internal regions are also removed by splicing.

15.5 Eukaryotic pre-mRNA Splicing

Eukaryotic genes may contain interruptions.

Coding DNA (an exon) is interrupted by noncoding introns. These introns are removed by splicing (figure 15.13).

The spliceosome is the splicing organelle.

snRNPs recognize intron–exon junctions and recruit spliceosomes. The spliceosome ultimately joins the 3′ end of the first exon to the 5′ end of the next exon.

Splicing can produce multiple transcripts from the same gene.

15.6 The Structure of tRNA and Ribosomes

Aminoacyl-tRNA synthetases attach amino acids to tRNA.

The tRNA charging reaction attaches the carboxyl terminus of an amino acid to the 3′ end of the correct tRNA (figure 15.15).

The ribosome has multiple tRNA-binding sites (figure 15.16).

A charged tRNA first binds to the A site, then moves to the P site where its amino acid is bonded to the peptide chain, and finally, without its amino acid, moves to the E site from which it is released.

The ribosome has both decoding and enzymatic functions.

Ribosomes hold tRNAs and mRNA in position for a ribosomal enzyme to form peptide bonds.

15.7 The Process of Translation

Initiation requires accessory factors.

In prokaryotes, initiation-complex formation is aided by the ribosome-binding sequence (RBS) of mRNA, complementary to a small subunit. Eukaryotes use the 5′ cap for the same function.

Elongation adds successive amino acids (figure 15.20).

As the ribosome moves along the mRNA, new amino acids from charged tRNAs are added to the growing peptide (figure 15.19).

Termination requires accessory factors.

Stop codons are recognized by termination factors.

Proteins may be targeted to the ER.

In eukaryotes, proteins with a signal sequence in their amino terminus bind to the SRP, and this complex docks on the ER.

(15.8 Summary is omitted.)

15.9 Mutation: Altered Genes

Point mutations affect a single site in the DNA .

Base substitutions exchange one base for another, and frameshift mutations involve the addition or deletion of a base. Triplet repeat expansion mutations can cause genetic diseases.

Chromosomal mutations change the structure of chromosomes.

Chromosomal mutations include additions, deletions, inversions, or translocations.

Mutations are the starting point of evolution.

Our view of the nature of genes has changed with new information.

Review Questions

UNDERSTAND

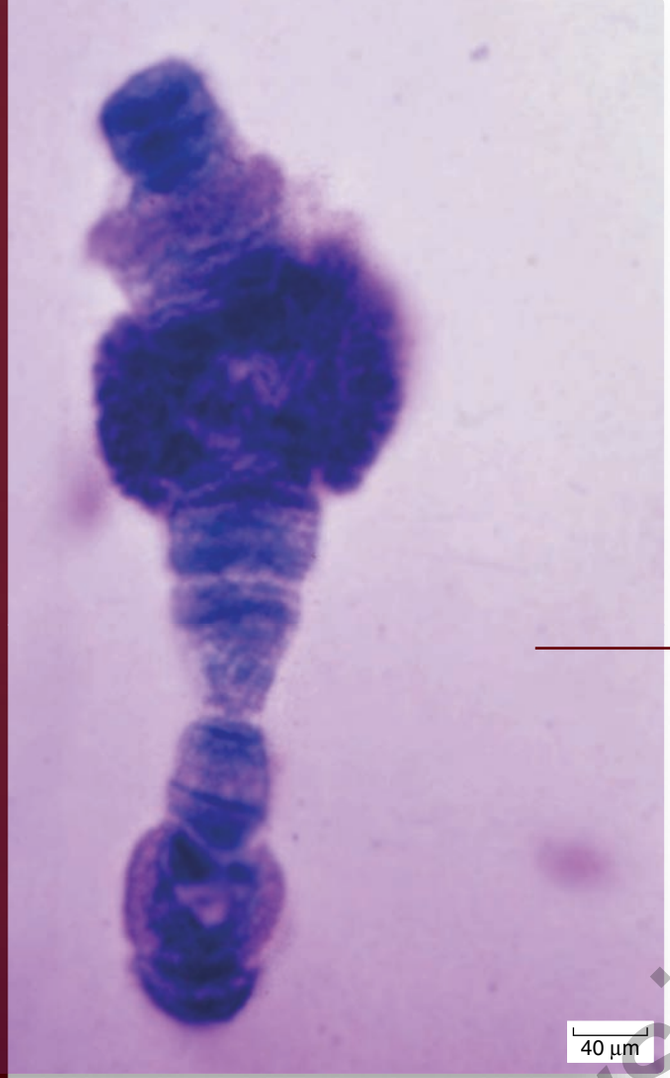
1. The experiments with nutritional mutants in *Neurospora* by Beadle and Tatum provided evidence that
   a. bread mold can be grown in a lab on minimal media.
   b. X-rays can damage DNA.
   c. cells need enzymes.
   d. genes specify enzymes.
2. What is the *central dogma* of molecular biology?
   a. DNA is the genetic material.
   b. Information passes from DNA directly to protein.
   c. Information passes from DNA to RNA to protein.
   d. One gene encodes only one polypeptide.
3. In the genetic code, one codon
   a. consists of three bases.
   b. specifies a single amino acid.
   c. specifies more than one amino acid.
   d. both a & b
4. Eukaryotic transcription differs from prokaryotic in that
   a. eukaryotes have only one RNA polymerase.
   b. eukaryotes have three RNA polymerases.
   c. prokaryotes have three RNA polymerases.
   d. both a & c
5. An anticodon would be found on which of the following types of RNA?
   a. snRNA (small nuclear RNA)
   b. mRNA (messenger RNA)
   c. tRNA (transfer RNA)
   d. rRNA (ribosomal RNA)
6. RNA polymerase binds to a _____ to initiate _____.
   a. mRNA; translation
   b. promoter; transcription
   c. primer; transcription
   d. transcription factor; translation
7. During translation, the codon in mRNA is actually "read" by
   a. the A site in the ribosome.
   b. the P site in the ribosome.
   c. the anticodon in a tRNA.
   d. the anticodon in an amino acid.

APPLY

1. Which of the following functions as a "stop" signal for a prokaryotic RNA polymerase?
   a. A specific sequence of bases called a terminator
   b. The Poly-A site
   c. Addition of a 5′ cap
   d. A region of the mRNA that can base-pair to form a hairpin
2. The splicing process
   a. occurs in prokaryotes.
   b. joins introns together.
   c. can produce multiple mRNAs from the same transcript.
   d. only joins exons for each gene in one way.
3. During translation, the ribosome must move along the mRNA. This movement
   a. requires the ribosome to come apart into subunits.
   b. requires an accessory factor and energy.
   c. does not require energy, but requires the ribosome to change conformation.
   d. requires accessory factors and uses energy from peptide bond formation.
4. In comparing transcription in prokaryotes and eukaryotes the mRNAs
   a. differ in that eukaryotic mRNAs often encode more than one protein.
   b. differ in that prokaryotic mRNAs often encode more than one protein.
   c. are similar in that both are always colinear with their genes.
   d. are similar in that neither is colinear with their genes.
5. A nonsense mutation
   a. results in large scale change to a chromosome.
   b. will lead to the premature termination of transcription.
   c. will lead to the premature termination of translation.
   d. is the same as a transversion.
6. An inversion will
   a. necessarily cause a mutant phenotype.
   b. only cause a mutant phenotype if the inversion breakpoints fall within a gene.
   c. halt transcription in the inverted region because the chromosome is now backwards.
   d. interfere with translation of genes in the inverted region.
7. What is the relationship between mutations and evolution?
   a. Mutations make genes better.
   b. Mutations can create new alleles.
   c. Mutations happened early in evolution, but not now.
   d. There is no relationship between evolution and genetic mutations.

SYNTHESIZE

1. A template strand of DNA has the following sequence:
   3′ – CGTTACCCGAGCCGTACGATTAGG – 5′
   Use the sequence information to determine
   a. the predicted sequence of the mRNA for this gene.
   b. the predicted amino acid sequence of the protein.
2. Frameshift mutations often result in truncated proteins. Explain this observation based on the genetic code.
3. Describe how each of the following mutations will affect the final protein product (protein begins with START codon). Name the type of mutation.
   Original template strand:
   3′ – CGTTACCCGAGCCGTACGATTAGG – 5′
   a. 3′ – CGTTACCCGAGCCGTAACGATTAGG – 5′
   b. 3′ – CGTTACCCGATCCGTACGATTAGG – 5′
   c. 3′ – CGTTACCCGAGCCGTTCGATTAGG – 5′
4. There are a number of features that are unique to bacteria, and others that are unique to eukaryotes. Could any of these features offer the possibility to control gene expression in a way that is unique to either eukaryotes or bacteria?



Chapter 16

Control of Gene Expression

Chapter Outline

16.1 Control of Gene Expression
16.2 Regulatory Proteins
16.3 Prokaryotic Regulation
16.4 Eukaryotic Regulation
16.5 Eukaryotic Chromatin Structure
16.6 Eukaryotic Posttranscriptional Regulation
16.7 Protein Degradation

Introduction

In a symphony, various instruments play their own parts at different times; the musical score determines which instruments play when. Similarly, in an organism, different genes are expressed at different times, with a "genetic score," written in regulatory regions of the DNA, determining which genes are active when. The picture shows the expanded "puff" of this Drosophila *chromosome, which represents genes that are being actively expressed. Gene expression and how it is controlled is our topic in this chapter.*

16.1 Control of Gene Expression

Learning Outcomes

1. *Identify the point at which control of gene expression usually occurs.*
2. *Describe the usual action of regulatory proteins.*
3. *List differences between control of gene expression in prokaryotes and that in eukaryotes.*

Control of gene expression is essential to all organisms. In prokaryotes, it allows the cell to take advantage of changing environmental conditions. In multicellular eukaryotes, it is critical for directing development and maintaining homeostasis.

Control usually occurs at the level of transcription initiation

You learned in the previous chapter that gene expression is the conversion of genotype to phenotype—the flow of information from DNA to produce functional proteins that control cellular activities. We could envision controlling this process at any point

along the way, and in fact, examples of control occur at most steps. The most logical place to control this process, however, is at the beginning: production of mRNA from DNA by transcription.

Transcription itself could be controlled at any step, but again, the beginning is the most logical place. Although cells do not always behave in ways that conform to human logic, control of the initiation of transcription is common.

RNA polymerase is key to transcription, and it must have access to the DNA helix and must be capable of binding to the gene's promoter for transcription to begin. **Regulatory proteins** act by modulating the ability of RNA polymerase to bind to the promoter. This idea of controlling the access of RNA polymerase to a promoter is common to both prokaryotes and eukaryotes, but the details differ greatly, as you will see.

These regulatory proteins bind to specific nucleotide sequences on the DNA that are usually only 10–15 nt in length. (Even a large regulatory protein has a "footprint," or binding area, of only about 20 nt.) Hundreds of these regulatory sequences have been characterized, and each provides a binding site for a specific protein that is able to recognize the sequence. Binding of the protein either *blocks* transcription by getting in the way of RNA polymerase or *stimulates* transcription by facilitating the binding of RNA polymerase to the promoter.

Control strategies in prokaryotes are geared to adjust to environmental changes

Control of gene expression is accomplished very differently in prokaryotes than it is in eukaryotes. Prokaryotic cells have been shaped by evolution to grow and divide as rapidly as possible, enabling them to exploit transient resources. Proteins in prokaryotes turn over rapidly, allowing these organisms to respond quickly to changes in their external environment by changing patterns of gene expression.

In prokaryotes, the primary function of gene control is to adjust the cell's activities to its immediate environment. Changes in gene expression alter which enzymes are present in response to the quantity and type of available nutrients and the amount of oxygen. Almost all of these changes are fully reversible, allowing the cell to adjust its enzyme levels up or down in response to environment changes.

Control strategies in eukaryotes are aimed at maintaining homeostasis

The cells of multicellular organisms, in contrast, have been shaped by evolution to be protected from transient changes in their immediate environment. Most of them experience fairly constant conditions. Indeed, *homeostasis*—the maintenance of a constant internal environment—is considered by many to be the hallmark of multicellular organisms. Cells in such organisms respond to signals in their immediate environment (such as growth factors and hormones) by altering gene expression, and in doing so they participate in regulating the body as a whole.

Some of these changes in gene expression compensate for changes in the physiological condition of the body. Others mediate the decisions that actually produce the body, ensuring that the correct genes are expressed in the right cells at the right time during development. Later chapters deal with the details, but for now we can simplify by saying that the growth and development of multicellular organisms entail a long series of biochemical reactions, each catalyzed by a specific enzyme. Once a particular developmental change has occurred, these enzymes cease to be active, lest they disrupt the events that must follow.

To produce this sequence of enzymes, genes are transcribed in a carefully prescribed order, each for a specified period of time, following a fixed genetic program that may even lead to programmed cell death **(apoptosis).** The one-time expression of the genes that guide a developmental program is fundamentally different from the reversible metabolic adjustments prokaryotic cells make in response to the environment. In all multicellular organisms, changes in gene expression within particular cells serve the needs of the whole organism, rather than the survival of individual cells.

Unicellular eukaryotes also use different control mechanisms from those of prokaryotes. All eukaryotes have a membrane-bounded nucleus, use similar mechanisms to condense DNA into chromosomes, and have the same gene expression machinery, all of which differ from those of prokaryotes.

Learning Outcomes Review 16.1

Gene expression is usually controlled at the level of transcription initiation. Regulatory proteins bind to specific DNA sequences and affect the binding of RNA polymerase to promoters. Individual protein may either prevent or stimulate transcription. In prokaryotes, regulation is focused on adjusting the cell's activities to the environment to ensure viability. In multicellular eukaryotes, regulation is geared to maintaining internal homeostasis, and even in unicellular forms, this control has mechanisms to deal with a bounded nucleus and multiple chromosomes.

- ***Would you expect the control of gene expression in a unicellular eukaryote like yeast to be more like that of humans or* E. coli?**

16.2 *Regulatory Proteins*

Learning Outcomes

1. ***Explain how proteins can interact with base pairs without unwinding the helix.***
2. ***Describe the common features of DNA-binding motifs.***

The ability of certain proteins to bind to *specific* DNA regulatory sequences provides the basic tool of gene regulation—the key ability that makes transcriptional control possible. To understand how cells control gene expression, it is first necessary to gain a clear picture of this molecular recognition process.

Proteins can interact with DNA through the major groove

In the past, molecular biologists thought that the DNA helix had to unwind before proteins could distinguish one DNA

sequence from another; only in this way, they reasoned, could regulatory proteins gain access to the hydrogen bonds between base-pairs. We now know it is unnecessary for the helix to unwind because proteins can bind to its outside surface, where the edges of the base-pairs are exposed.

Careful inspection of a DNA molecule reveals two helical grooves winding around the molecule, one deeper than the other. Within the deeper groove, called the **major groove,** the nucleotides' hydrogen bond donors and acceptors are accessible. The pattern created by these chemical groups is unique for each of the four possible base-pair arrangements, providing a ready way for a protein nestled in the groove to read the sequence of bases (figure 16.1).

DNA-binding domains interact with specific DNA sequences

Protein–DNA recognition is an area of active research; so far, the structures of over 30 regulatory proteins have been analyzed. Although each protein is unique in its fine details, the part of the protein that actually binds to the DNA is much less variable. Almost all of these proteins employ one of a small set of **DNA-binding motifs.** A motif, as described in chapter 3, is a form of three-dimensional substructure that is found in many proteins. These DNA-binding motifs share the property of interacting with specific sequences of bases, usually through the major groove of the DNA helix.

DNA-binding motifs are the key structure within the DNA-binding domain of these proteins. This domain is a functionally distinct part of the protein necessary to bind to DNA in a sequence-specific manner. Regulatory proteins also need to be able to interact with the transcription apparatus, which is accomplished by a different regulatory domain.

Note that two proteins that share the same DNA-binding domain do not necessarily bind to the same DNA sequence. The similarities in the DNA-binding motifs appear in their three-dimensional structure, and not in the specific contacts that they make with DNA.

Several common DNA-binding motifs are shared by many proteins

A limited number of common DNA-binding motifs are found in a wide variety of different proteins. Four of the best known are detailed in the following sections to give the sense of how DNA-binding proteins interact with DNA.

The helix-turn-helix motif

The most common DNA-binding motif is the **helix-turn-helix,** constructed from two α-helical segments of the protein linked by a short, nonhelical segment, the "turn" (figure 16.2*a*). As the first motif recognized, the helix-turn-helix motif has since been identified in hundreds of DNA-binding proteins.

A close look at the structure of a helix-turn-helix motif reveals how proteins containing such motifs interact with the major groove of DNA. The helical segments of the motif interact with one another, so that they are held at roughly right angles. When this motif is pressed against DNA, one of the helical

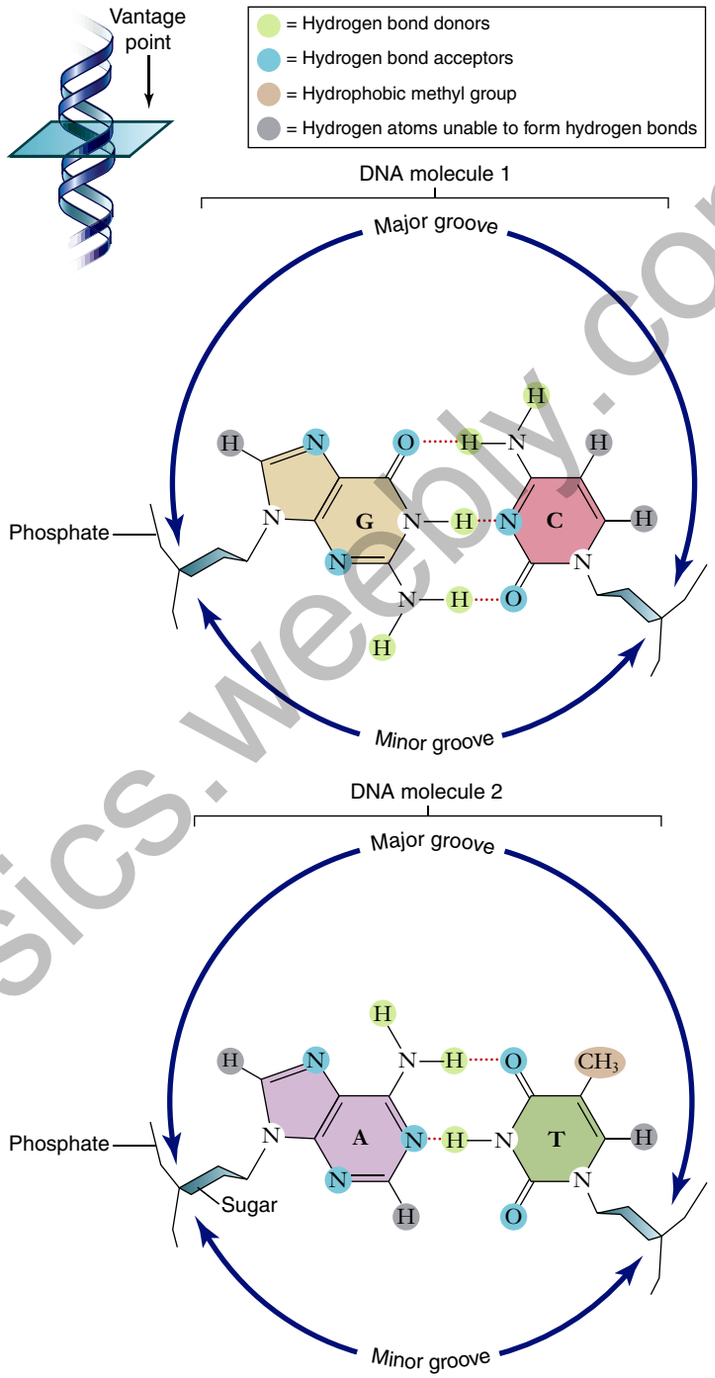


Figure 16.1 Reading the major groove of DNA. Looking down into the major groove of a DNA helix, we can see the edges of the bases protruding into the groove. Each of the four possible base-pair arrangements (two are shown here) extends a unique set of chemical groups into the groove, indicated in this diagram by differently colored circles. A regulatory protein can identify the base-pair arrangement by this characteristic signature.

segments (called the *recognition helix*) fits snugly in the major groove of the DNA molecule, and the other butts up against the outside of the DNA molecule, helping to ensure the proper positioning of the recognition helix.

Most DNA-regulatory sequences recognized by helix-turn-helix motifs occur in symmetrical pairs. Such sequences

are bound by proteins containing two helix-turn-helix motifs separated by 3.4 nanometers (nm), the distance required for one turn of the DNA helix (see figure 16.2*a*). Having *two* protein–DNA-binding sites doubles the zone of contact between protein and DNA and greatly strengthens the bond between them.

The homeodomain motif

A special class of helix-turn-helix motifs, the **homeodomain,** plays a critical role in development in a wide variety of eukaryotic organisms, including humans. These motifs were discovered when researchers began to characterize a set of homeotic mutations in *Drosophila* (mutations that cause one body part to be replaced by another). They found that the mutant genes encoded regulatory proteins. Normally these proteins would initiate key stages of development by binding to developmental switch-point genes. More than 50 of these regulatory proteins have been analyzed, and they all contain a nearly identical sequence of 60 amino acids, which was termed the *homeodomain*. The most conserved part of the homeodomain contains a recognition helix of a helix-turn-helix motif. The rest of the homeodomain forms the other two helices of this motif.

The zinc finger motif

A different kind of DNA-binding motif uses one or more zinc atoms to coordinate its binding to DNA. Called **zinc fingers**, these motifs exist in several forms. In one form, a zinc atom links an α-helical segment to a β-sheet segment (see chapter 3) so that the helical segment fits into the major groove of DNA.

This sort of motif often occurs in clusters, the β sheets spacing the helical segments so that each helix contacts the major groove. The effect is like a hand wrapped around the DNA with the fingers lying in the major groove. The more zinc fingers in the cluster, the stronger the protein binds to the DNA.

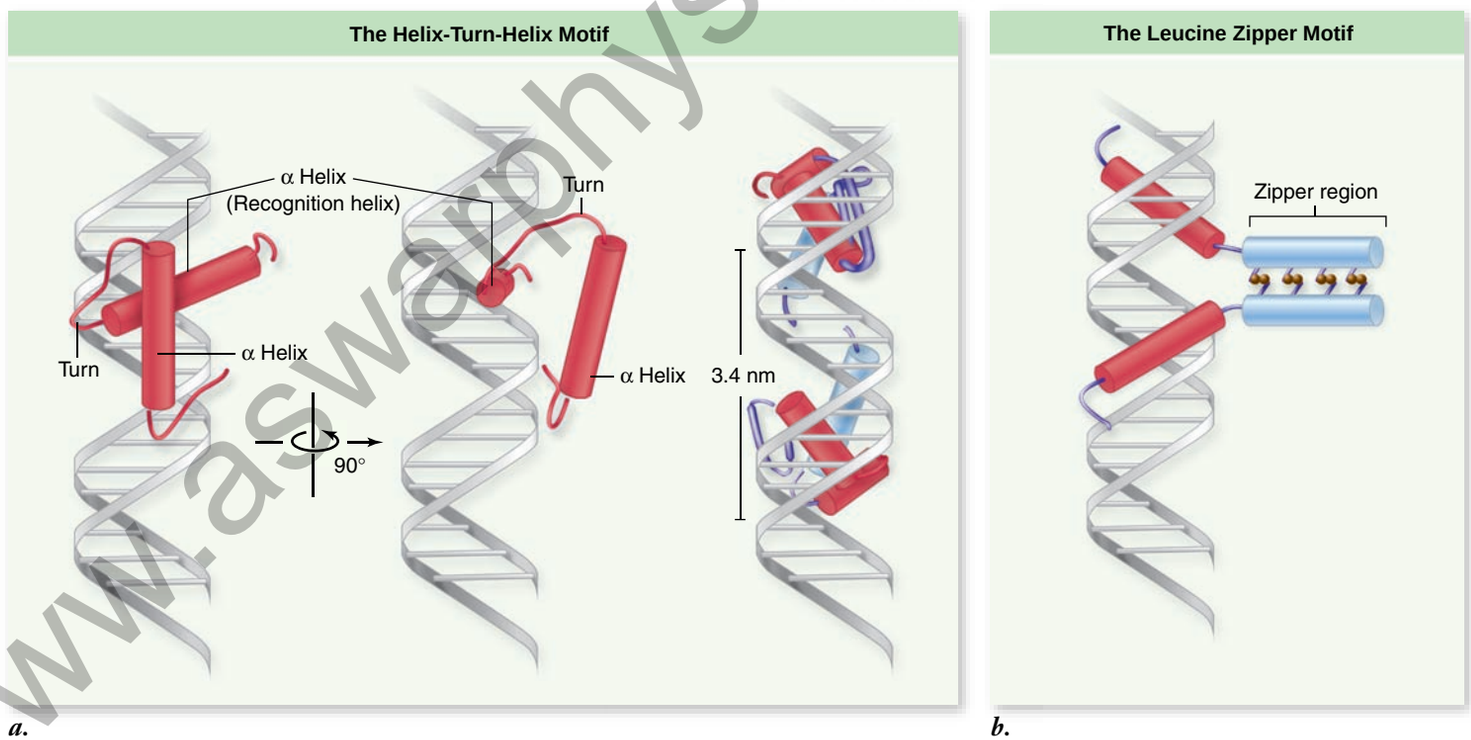
The leucine zipper motif

In yet another DNA-binding motif, two different protein subunits cooperate to create a single DNA-binding site. This motif is created where a region on one subunit containing several hydrophobic amino acids (usually leucines) interacts with a similar region on the other subunit. This interaction holds the two subunits together at those regions, while the rest of the subunits remain separated. Called a **leucine zipper,** this structure has the shape of a Y, with the two arms of the Y being helical regions that fit into the major groove of DNA (figure 16.2*b*). Because the two subunits can contribute quite different helical regions to the motif, leucine zippers allow for great flexibility in controlling gene expression.

Learning Outcomes Review 16.2

A DNA helix exhibits a major groove and a minor groove; regulatory proteins interact with DNA by accessing bases along the major groove. These proteins all contain DNA-binding motifs, and they often include one or two α-helical segments. These motifs form the active part of the DNA-binding domain, and another domain of the protein interacts with the transcription apparatus.

- ***What would be the effect of a mutation in a helix-turn-helix protein that altered the spacing of the two helices?***



a. ***b.***

Figure 16.2 Major DNA-binding motifs. Two different DNA-binding motifs are pictured interacting with DNA. ***a.*** The helix-turn-helix motif binds to DNA using one α helix, the recognition helix, to interact with the major groove. The other helix positions the recognition helix. Proteins with this motif are usually dimers, with two identical subunits, each containing the DNA-binding motif. The two copies of the motif *(red)* are separated by 3.4 nm, precisely the spacing of one turn of the DNA helix. ***b.*** The leucine zipper acts to hold two subunits in a multisubunit protein together, thereby allowing α-helical regions to interact with DNA.

16.3 Prokaryotic Regulation

Learning Outcomes

1. *Compare how control by induction differs from control by repression.*
2. *Explain control of gene expression in the* lac *operon.*
3. *Explain control of gene expression in the* trp *operon.*

The details of regulation can be revealed by examining mechanisms used by prokaryotes to control the initiation of transcription. Prokaryotes and eukaryotes share some common themes, but they have some profound differences as well. Later on we discuss eukaryotic systems and concentrate on how they differ from the simpler prokaryotic systems.

Control of transcription can be either positive or negative

Control at the level of transcription initiation can be either positive or negative. **Positive control** increases the frequency of initiation, and **negative control** decreases the frequency of initiation. Each of these forms of control are mediated by regulatory proteins, but the proteins have opposite effects.

Negative control by repressors

Negative control is mediated by proteins called **repressors.** Repressors are proteins that bind to regulatory sites on DNA called **operators** to prevent or decrease the initiation of transcription. They act as a kind of roadblock to prevent the polymerase from initiating effectively.

Repressors do not act alone; each responds to specific effector molecules. Effector binding can alter the conformation of the repressor to either enhance or abolish its binding to DNA. These repressor proteins are allosteric proteins with an active site that binds DNA and a regulatory site that binds effectors. Effector binding at the regulatory site changes the ability of the repressor to bind DNA (see chapter 6 for more details on allosteric proteins).

Positive control by activators

Positive control is mediated by another class of regulatory, allosteric proteins called *activators* that can bind to DNA and stimulate the initiation of transcription. These activators enhance the binding of RNA polymerase to the promoter to increase the level of transcription initiation.

Activators are the logical and physical opposites of repressors. Effector molecules can either enhance or decrease activator binding.

Prokaryotes adjust gene expression in response to environmental conditions

Changes in the environments that bacteria and archaea encounter often result in changes in gene expression. In general, genes encoding proteins involved in catabolic pathways (breaking down molecules) respond oppositely from genes encoding proteins involved in anabolic pathways (building up molecules). In the discussion that follows, we describe enzymes in the catabolic pathway that transports and utilizes the sugar lactose. Later we describe the anabolic pathway that synthesizes the amino acid tryptophan.

As mentioned in the preceding chapter, prokaryotic genes are often organized into operons, multiple genes that are part of a single transcription unit having a single promoter. Genes that are involved in the same metabolic pathway are often organized in this fashion. The proteins necessary for the utilization of lactose are encoded by the ***lac* operon,** and the proteins necessary for the synthesis of tryptophan are encoded by the ***trp* operon.**

Inquiry question

What advantage might a bacterium get by linking into a single operon several genes, all of the products of which contribute to a single biochemical pathway?

Induction and repression

If a bacterium encounters lactose, it begins to make the enzymes necessary to utilize lactose. When lactose is not present, however, there is no need to make these proteins. Thus, we say that the synthesis of the proteins is *induced* by the presence of lactose. **Induction** therefore occurs when enzymes for a certain pathway are produced in response to a substrate.

When tryptophan is available in the environment, a bacterium will not synthesize the enzymes necessary to make tryptophan. If tryptophan ceases to be available, then the bacterium begins to make these enzymes. **Repression** occurs when bacteria capable of making biosynthetic enzymes do not produce them. In the case of both induction and repression, the bacterium is adjusting to produce the enzymes that are optimal for its immediate environment.

Negative control

Knowing that gene expression is probably controlled at the level of initiation of transcription does not tell us whether that control is positive or negative. On the surface, repression may appear to be negative and induction positive; but in the case of both the *lac* and *trp* operons, control is negative by a repressor protein. The key is that the effector proteins have opposite effects on the repressor in induction with those seen in repression.

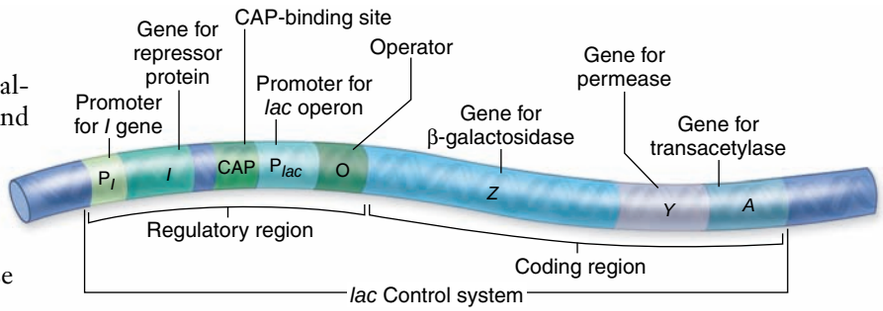


Figure 16.3 The *lac* region of the *Escherichia coli* chromosome. The *lac* operon consists of a promoter, an operator, and three genes (*lac Z*, *Y*, and *A*) that encode proteins required for the metabolism of lactose. In addition, there is a binding site for the catabolite activator protein (CAP), which affects RNA polymerase binding to the promoter. The *I* gene encodes the repressor protein, which can bind the operator and block transcription of the *lac* operon.

For either mechanism to work, the molecule in the environment, such as lactose or tryptophan, must produce the proper effect on the gene being regulated. In the case of *lac* induction, the presence of lactose must *prevent* a repressor protein from binding to its regulatory sequence. In the case of *trp* repression, by contrast, the presence of tryptophan must *cause* a repressor protein to bind to its regulatory sequence.

These responses are opposite because the needs of the cell are opposite in anabolic versus catabolic pathways. Each pathway is examined in detail in the following sections to show how protein–DNA interactions allow the cell to respond to environmental conditions.

The *lac* operon is negatively regulated by the *lac* repressor

The control of gene expression in the *lac* operon was elucidated by the pioneering work of Jaques Monod and François Jacob.

The *lac* operon consists of the genes that encode functions necessary to utilize lactose: β-galactosidase *(lacZ)*, lactose permease *(lacY)*, and lactose transacetylase *(lacA)*, plus the regulatory regions necessary to control the expression of these genes (figure 16.3). In addition, the gene for the *lac* repressor *(lacI)* is linked to the rest of the *lac* operon and is thus considered part of the operon although it has its own promoter. The arrangement of the control regions upstream of the coding region is typical of most prokaryotic operons, although the linked repressor is not.

Action of the repressor

Initiation of transcription of the *lac* operon is controlled by the *lac* repressor. The repressor binds to the operator, which is adjacent to the promoter (figure 16.4*a*). This binding prevents RNA polymerase from binding to the promoter. This DNA binding is sensitive to the presence of lactose: The repressor binds DNA in the absence of lactose, but not in the presence of lactose.

Interaction of repressor and inducer

In the absence of lactose, the *lac* repressor binds to the operator, and the operon is repressed (see figure 16.4*a*). The effector that controls the DNA binding of the repressor is a metabolite of

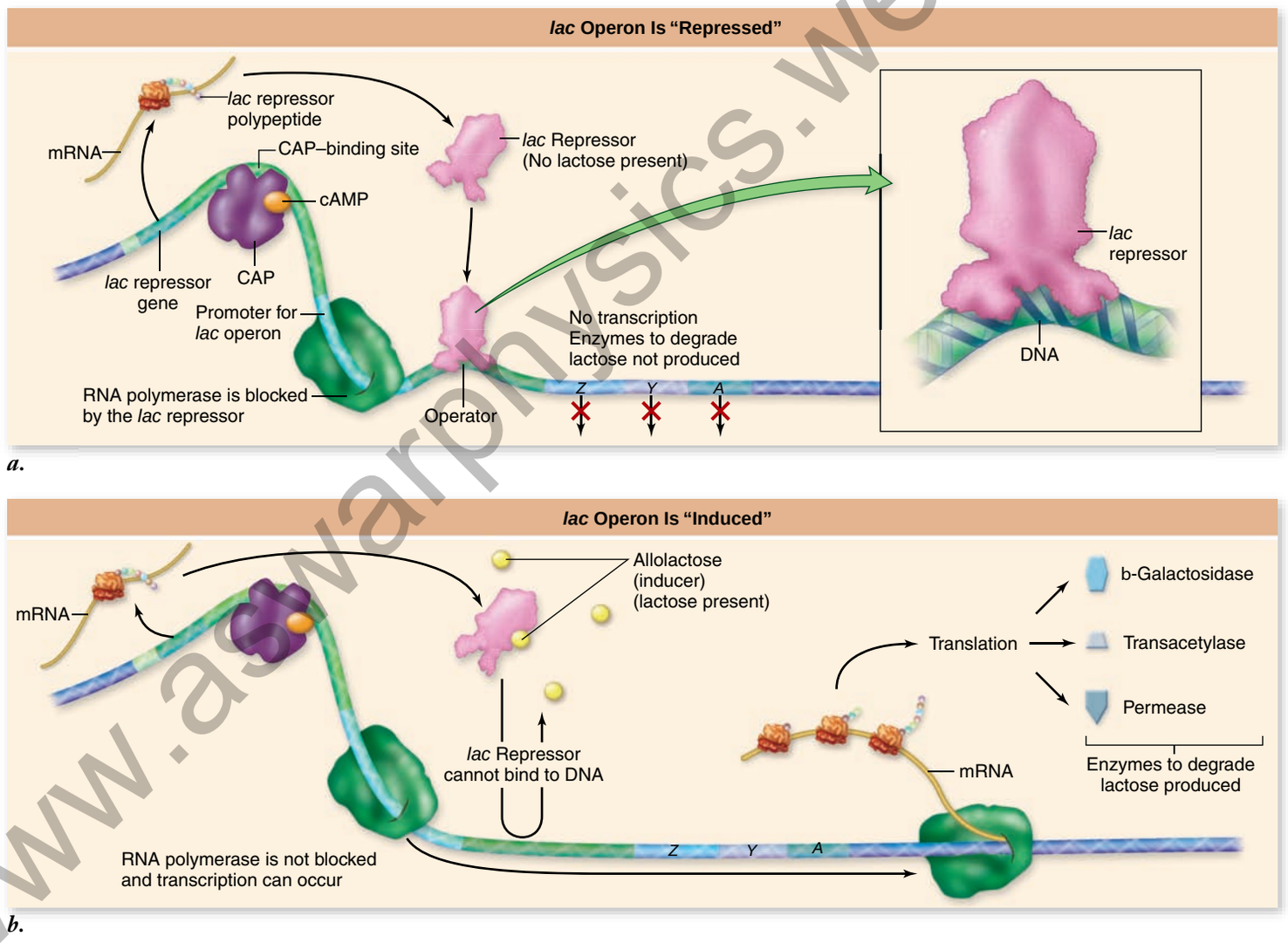


Figure 16.4 Induction of the *lac* operon. ***a.*** In the absence of lactose the *lac* repressor binds to DNA at the operator site, thus preventing transcription of the operon. When the repressor protein is bound to the operator site, the *lac* operon is shut down (repressed). ***b.*** The *lac* operon is transcribed (induced) when CAP is bound and when the repressor is not bound. Allolactose binding to the repressor alters the repressor's shape so it cannot bind to the operator site and block RNA polymerase activity.

lactose, allolactose, which is produced when lactose is available. Allolactose binds to the repressor, altering its conformation so that it no longer can bind to the operator (figure 16.4*b*). The operon is now induced. Since allolactose allows induction of the operon, it is usually called the inducer.

As the level of lactose falls, allolactose will no longer be available to bind to the repressor, allowing the repressor to bind to DNA again. Thus this system of negative control by the *lac* repressor and its inducer, allolactose, allows the cell to respond to changing levels of lactose in the environment.

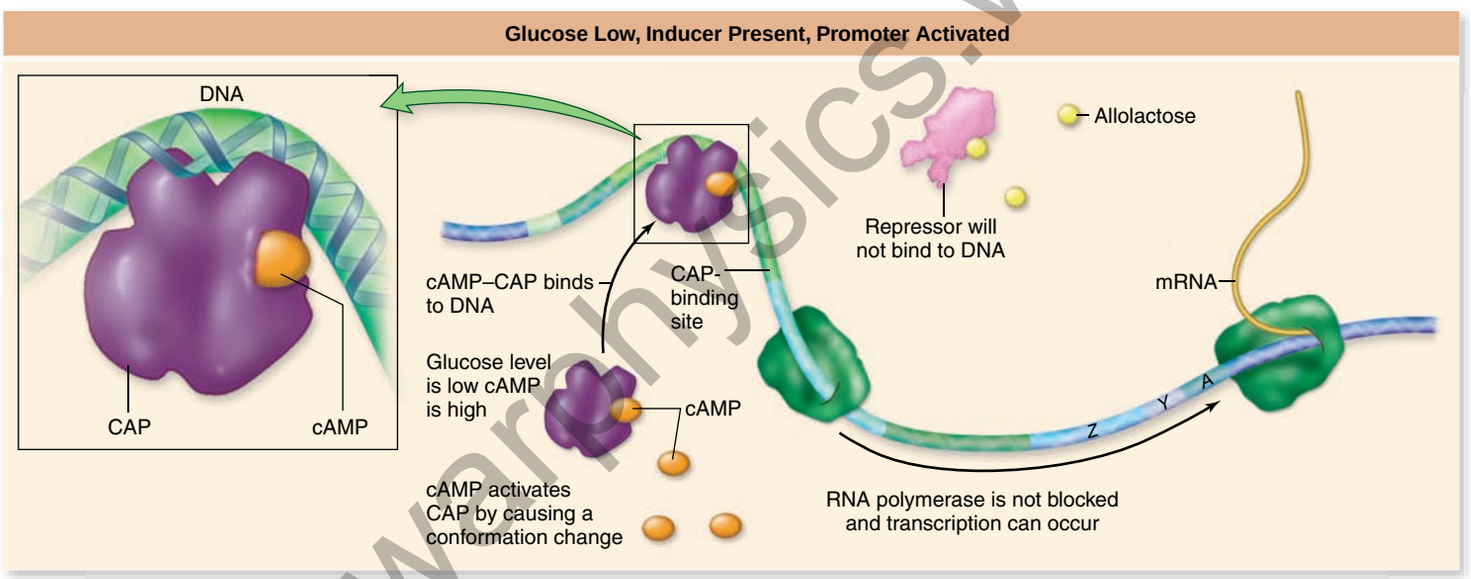
Even in the absence of lactose, the *lac* operon is expressed at a very low level. When lactose becomes available, it is transported into the cell and enough allolactose is produced that induction of the operon can occur.

The presence of glucose prevents induction of the *lac* operon

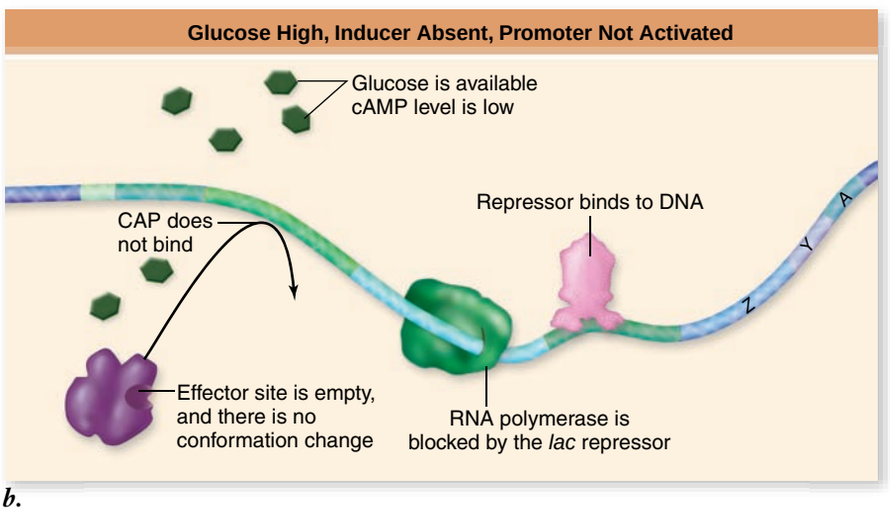
Glucose repression is the preferential use of glucose in the presence of other sugars such as lactose. If bacteria are grown in the presence of both glucose and lactose, the *lac* operon is not induced. When the glucose is used up, the *lac* operon is induced, allowing lactose to be used as an energy source.

Despite the name *glucose repression*, this mechanism involves an activator protein that can stimulate transcription from multiple catabolic operons, including the *lac* operon. This activator, **catabolite activator protein (CAP),** is an allosteric protein with cAMP as an effector. This protein is also called **cAMP response protein (CRP)** because it binds cAMP, but we will use the name CAP to emphasize its role as a positive regulator. CAP alone does not bind to DNA, but binding of the effector cAMP to CAP changes its conformation such that it can bind to DNA (figure 16.5). The level of cAMP in cells is reduced in the presence of glucose so that no stimulation of transcription from CAP-responsive operons takes place.

The CAP–cAMP system was long thought to be the sole mechanism of glucose repression. But more recent research has indicated that the presence of glucose inhibits the transport of lactose into the cell. This deprives the cell of the *lac* operon inducer, allolactose, allowing the repressor to bind to the operator. This mechanism, called **inducer exclusion,** is now thought to be the main form of glucose repression of the *lac* operon.



a.



b.

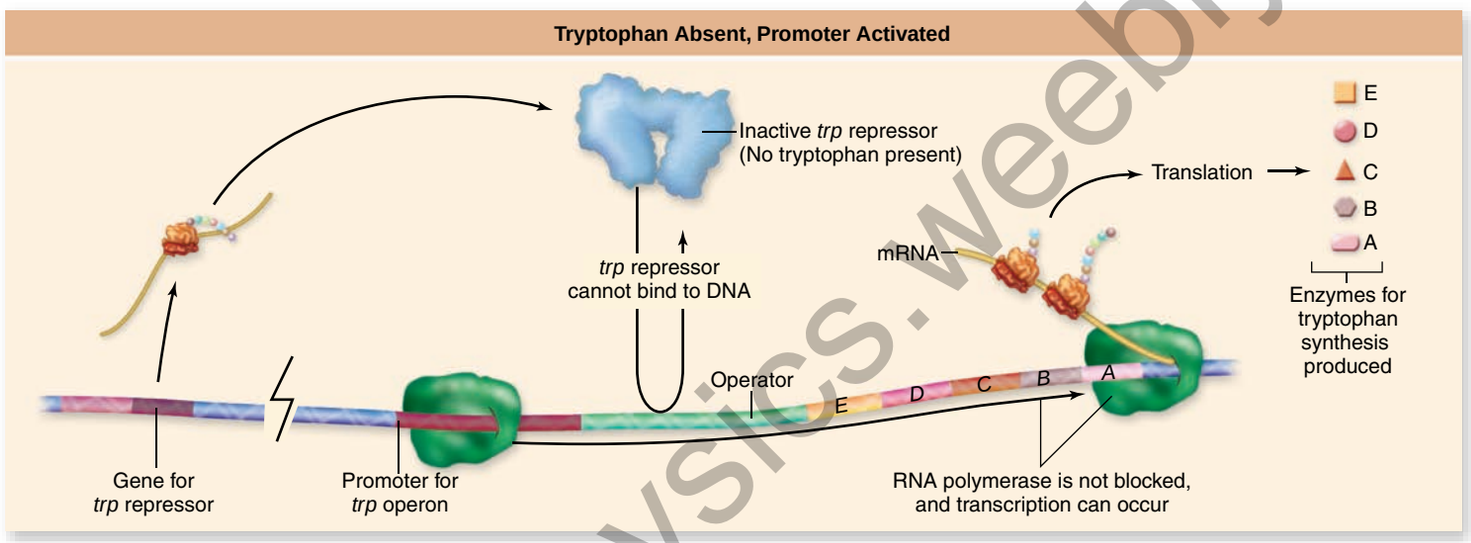
Figure 16.5 Effect of glucose on the *lac* operon. Expression of the *lac* operon is controlled by a negative regulator (repressor) and a positive regulator (CAP). The action of CAP is sensitive to glucose levels. ***a.*** For CAP to bind to DNA, it must bind to cAMP. When glucose levels are low, cAMP is abundant and binds to CAP. The CAP–cAMP complex causes the DNA to bend around it. This brings CAP into contact with RNA polymerase (not shown) making polymerase binding to the promoter more efficient. ***b.*** High glucose levels produce two effects: cAMP is scarce so CAP is unable to activate the promoter, and the transport of lactose is blocked (inducer exclusion).

Given that inducer exclusion occurs, the role of CAP in the absence of glucose seems superfluous. But in fact, the action of CAP–cAMP allows maximal expression of the operon in the absence of glucose. The positive control of CAP–cAMP is necessary because the promoter of the *lac* operon alone is not efficient in binding RNA polymerase. This inefficiency is overcome by the action of the positive control of the CAP–cAMP activator (see figure 16.5).

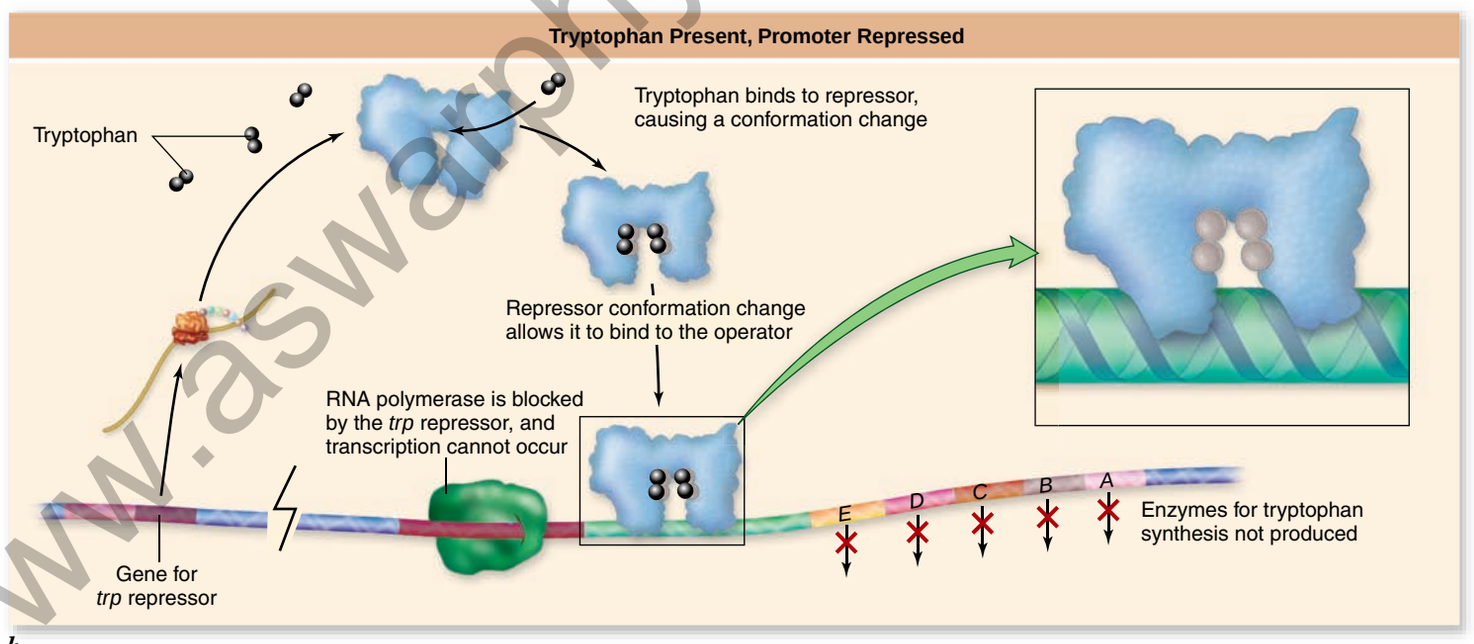
The *trp* operon is controlled by the *trp* repressor

Like the *lac* operon, the *trp* operon consists of a series of genes that encode enzymes involved in the same biochemical pathway. In the case of the *trp* operon these enzymes are necessary for synthesizing tryptophan. The regulatory region that controls transcription of these genes is located upstream of the genes. The *trp* operon is controlled by a repressor encoded by a gene located outside the *trp* operon. The *trp* operon is continuously expressed in the absence of tryptophan and is not expressed in the presence of tryptophan.

The *trp* repressor is a helix-turn-helix protein that binds to the operator site located adjacent to the *trp* promoter (figure 16.6). This repressor behaves in a manner opposite to the *lac* repressor. In the absence of tryptophan, the *trp* repressor does not bind to its operator, allowing expression of the operon, and production of the enzymes necessary to make tryptophan.



a.



b.

Figure 16.6 How the *trp* operon is controlled. The tryptophan operon encodes the enzymes necessary to synthesize tryptophan. ***a.*** The tryptophan repressor alone cannot bind to DNA. The promoter is free to function, and RNA polymerase transcribes the operon. ***b.*** When tryptophan is present, it binds to the repressor, altering its conformation so it now binds DNA. The tryptophan–repressor complex binds tightly to the operator, preventing RNA polymerase from initiating transcription.

When levels of tryptophan rise, then tryptophan (the *corepressor*) binds to the repressor and alters its conformation, allowing it to bind to its operator. Binding of the repressor–corepressor complex to the operator prevents RNA polymerase from binding to the promoter. The actual change in repressor structure due to tryptophan binding is an alteration of the orientation of a pair of helix-turn-helix motifs that allows their recognition helices to fit into adjacent major grooves of the DNA (figure 16.7).

When tryptophan is present and bound to the repressor and this complex is bound to the operator, the operon is said to be *repressed*. As tryptophan levels fall, the repressor alone cannot bind to the operator, allowing expression of the operon. In this state, the operon is said to be **derepressed,** distinguishing this state from induction (see figure 16.6).

The key to understanding how both induction and repression can be due to negative regulation is knowledge of the behavior of repressor proteins and their effectors. In induction, the repressor alone can bind to DNA, and the inducer prevents DNA binding. In the case of repression, the repressor only binds DNA when bound to the corepressor. Induction and repression are excellent examples of how interactions of molecules can affect their structures, and how molecular structure is critical to function.

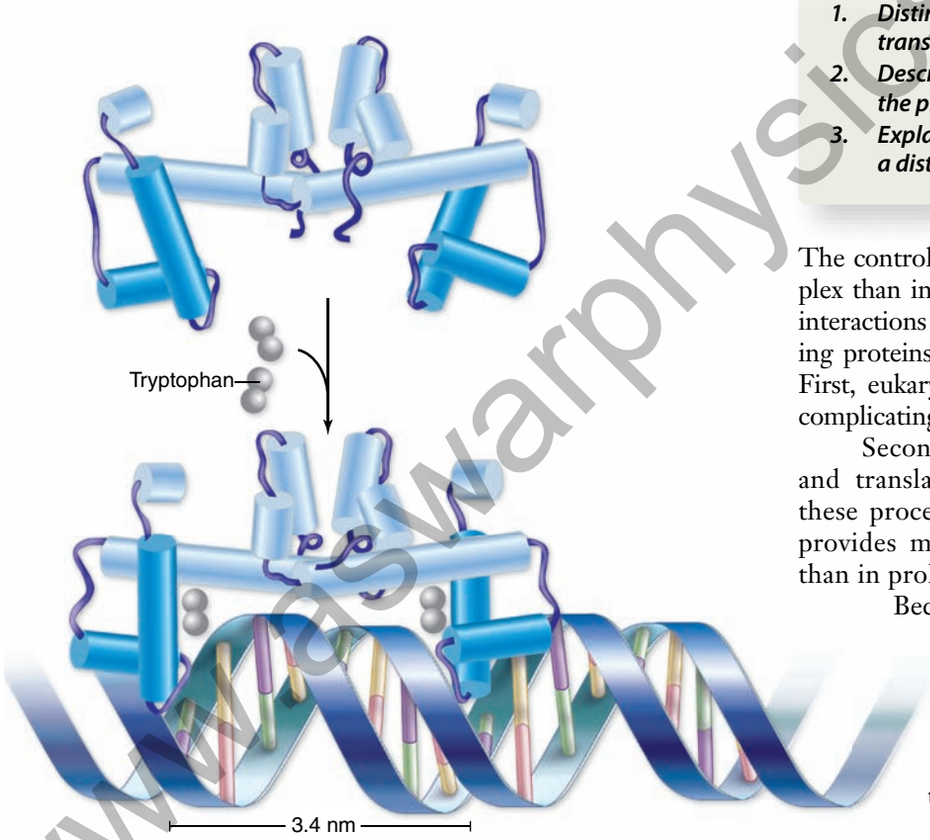


Figure 16.7 How the tryptophan repressor works. The binding of tryptophan to the repressor increases the distance between the two recognition helices in the repressor, allowing the repressor to fit snugly into two adjacent portions of the major groove in DNA.

Learning Outcomes Review 16.3

Induction occurs when expression of genes in a pathway is turned on in response to a substrate; repression occurs when expression is prevented in response to a substrate. The *lac* operon is negatively controlled by a repressor protein that binds to DNA, thus preventing transcription. When lactose is present, the operon is turned on; allolactose binds to the repressor, which then no longer binds to DNA. This operon is also positively regulated by an activator protein. The *trp* operon is negatively controlled by a repressor protein that must be bound to tryptophan in order to bind to DNA. In the absence of tryptophan, the repressor cannot bind DNA, and the operon is derepressed.

- ***What would be the effect on regulation of the* trp *operon of a mutation in the* trp *repressor that can still bind to* trp*, but no longer bind to DNA?***

16.4 Eukaryotic Regulation

Learning Outcomes

1. ***Distinguish between the role of general and specific transcription factors.***
2. ***Describe events necessary for Pol II to bind to the promoter.***
3. ***Explain how transcription factors can have an effect from a distance in the DNA.***

The control of transcription in eukaryotes is much more complex than in prokaryotes. The basic concepts of protein–DNA interactions are still valid, but the nature and number of interacting proteins is much greater due to some obvious differences. First, eukaryotes have their DNA organized into chromatin, complicating protein–DNA interactions considerably.

Second, eukaryotic transcription occurs in the nucleus, and translation occurs in the cytoplasm; in prokaryotes, these processes are spatially and temporally coupled. This provides more opportunities for regulation in eukaryotes than in prokaryotes.

Because of these differences, the amount of DNA involved in regulating eukaryotic genes is much greater. The need for a fine degree of flexible control is especially important for multicellular eukaryotes, with their complex developmental programs and multiple tissue types. General themes, however, emerge from this complexity.

Transcription factors can be either general or specific

In the preceding chapter we introduced the concept of transcription factors. Eukaryotic transcription requires a variety of

these protein factors, which fall into two categories: *general transcription factors* and *specific transcription factors*. General factors are necessary for the assembly of a transcription apparatus and recruitment of RNA polymerase II to a promoter. Specific factors increase the level of transcription in certain cell types or in response to signals.

General transcription factors

Transcription of RNA polymerase II templates (the majority being genes that encode protein products) requires more than just RNA polymerase II to initiate transcription. A host of **general transcription factors** are also necessary to establish productive initiation. These factors are required for transcription to occur, but they do not increase the rate above this basal rate.

General transcription factors are named with letter designations that follow the abbreviation TFII, for "transcription factor RNA polymerase II." The most important of these factors, TFIID, contains the TATA-binding protein that recognizes the TATA box sequence found in many eukaryotic promoters (figure 16.8).

Binding of TFIID is followed by binding of TFIIE, TFIIF, TFIIA, TFIIB, and TFIIH and a host of accessory factors called *transcription-associated factors*, TAFs. The *initiation complex* that results (figure 16.9) is clearly much more complex than the bacterial RNA polymerase holoenzyme binding to a promoter. And there is yet another level of complexity: The initiation complex, although capable of initiating synthesis at a basal level, does not achieve transcription at a high level without the participation of other, specific factors.

Specific transcription factors

Specific transcription factors act in a tissue- or time-dependent manner to stimulate higher levels of transcription than the basal level. The number and diversity of these factors are overwhelming. Some sense can be made of this proliferation of factors by concentrating on the DNA-binding motif, as opposed to the specific factors.

A key common theme that emerges from the study of these factors is that specific transcription factors, called *activators*, have a domain organization. Each factor consists of a DNA-binding domain and a separate activating domain that interacts with the transcription apparatus, and these domains are essentially independent in the protein. If the DNA-binding domains are "swapped" between different factors the binding specificity for the factors is switched without affecting their ability to activate transcription.

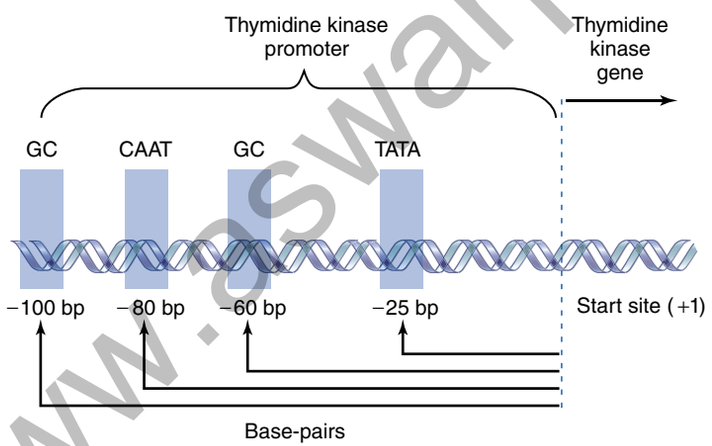


Figure 16.8 A eukaryotic promoter. This promoter is for the gene encoding the enzyme thymidine kinase. Formation of the transcription initiation complex begins with a general transcription factor binding to the TATA box. There are three other DNA sequences that direct the binding of other specific transcription factors.

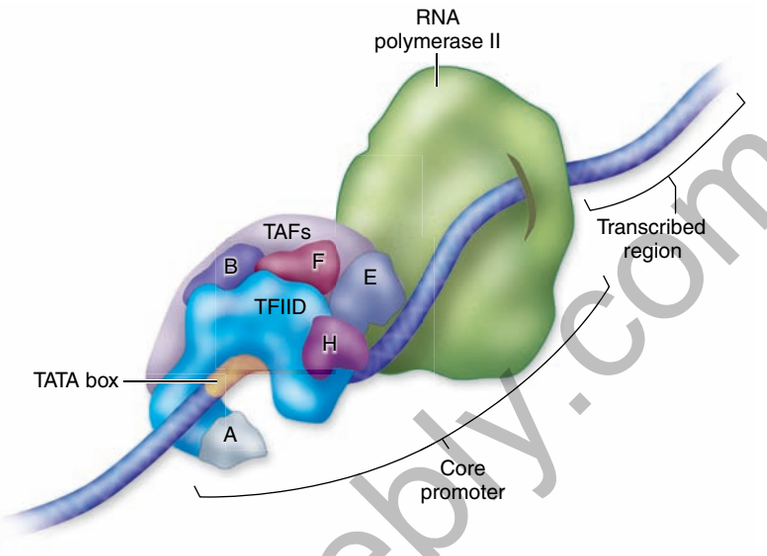


Figure 16.9 Formation of a eukaryotic initiation complex. The general transcription factor, TFIID, binds to the TATA box and is joined by the other general factors, TFIIE, TFIIF, TFIIA, TFIIB, and TFIIH. This complex is added to by a number of transcription-associated factors (TAFs) that together recruit the RNA pol II molecule to the core promoter.

Promoters and enhancers are binding sites for transcription factors

Promoters, as mentioned in the preceding chapter, form the binding sites for general transcription factors. These factors then mediate the binding of RNA polymerase II to the promoter (and also the binding of RNA polymerases I and III to their specific promoters). In contrast, the holoenzyme portion of the RNA polymerase of prokaryotes can directly recognize a promoter and bind to it.

Enhancers were originally defined as DNA sequences necessary for high levels of transcription that can act independently of position or orientation. At first, this concept seemed counterintuitive, especially since molecular biologists had been conditioned by prokaryotic systems to expect control regions to be immediately upstream of the coding region. It turns out that enhancers are the binding site of

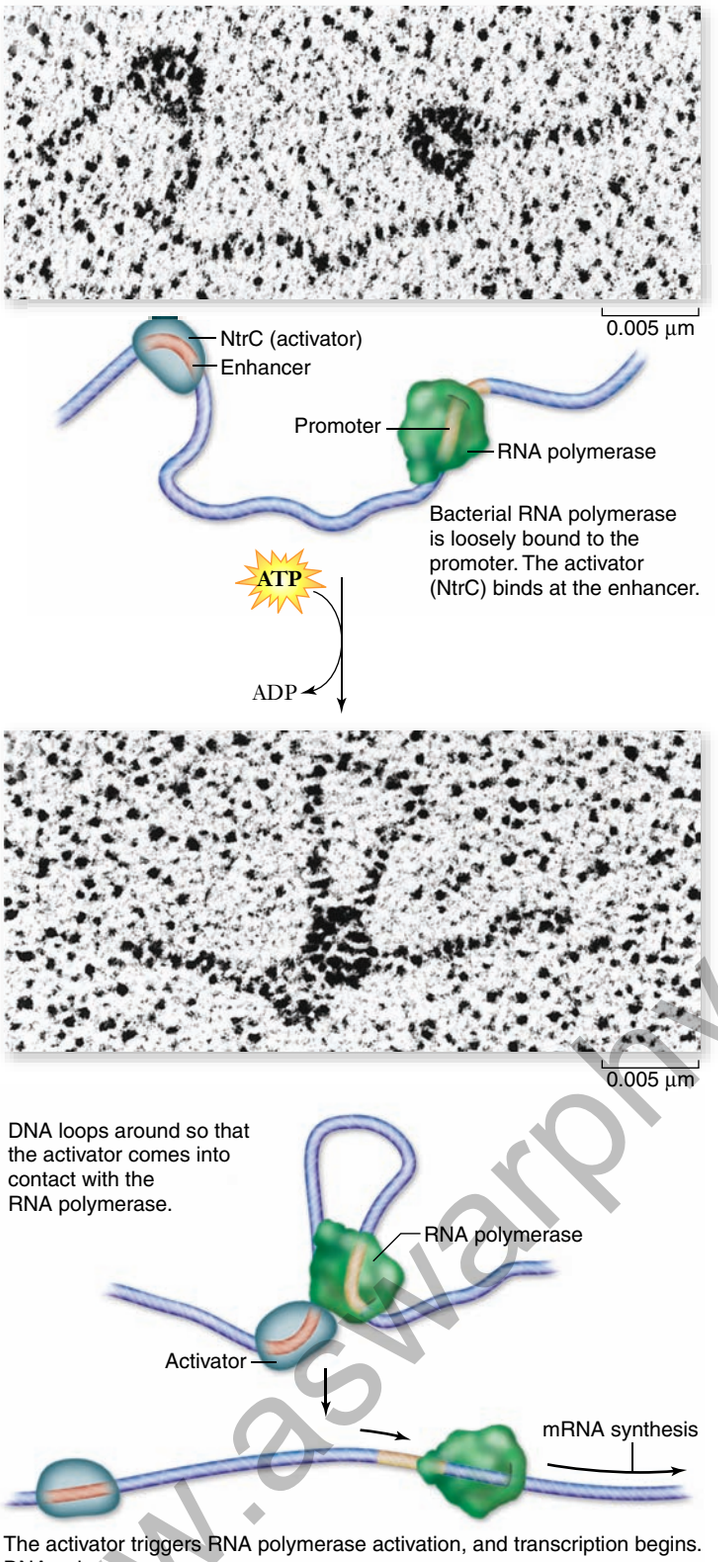


Figure 16.10 DNA looping caused by proteins. When the bacterial activator NtrC binds to an enhancer, it causes the DNA to loop over to a distant site where RNA polymerase is bound, thereby activating transcription. Although such enhancers are rare in prokaryotes, they are common in eukaryotes.

the specific transcription factors. The ability of enhancers to act over large distances was at first puzzling, but investigators now think this action is accomplished by DNA bending to form a loop, positioning the enhancer closer to the promoter.

Although more important in eukaryotic systems, this looping was first demonstrated using prokaryotic DNA-binding proteins (figure 16.10). The important point is that the linear distance separating two sites on the chromosome does not have to translate to great physical distance, because the flexibility of DNA allows bending and looping. An activator bound to an enhancer can thus be brought into contact with the transcription factors bound to a distant promoter (figure 16.11).

Coactivators and mediators link transcription factors to RNA polymerase II

Other factors specifically mediate the action of transcription factors. These *coactivators* and *mediators* are also necessary for activation of transcription by the transcription factor. They act by binding the transcription factor and then binding to another part of the transcription apparatus. Mediators are essential to the function of some transcription factors, but not all transcription

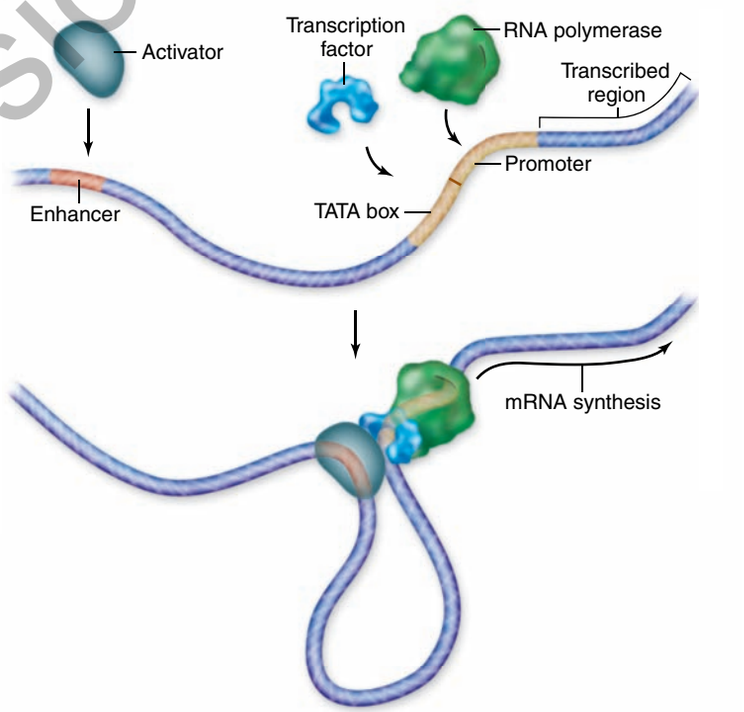


Figure 16.11 How enhancers work. The enhancer site is located far away from the gene being regulated. Binding of an activator *(gray)* to the enhancer allows the activator to interact with the transcription factors *(blue)* associated with RNA polymerase, stimulating transcription.

factors require them. The number of coactivators is much smaller than the number of transcription factors because the same coactivator can be used with multiple transcription factors.

The transcription complex brings things together

Although a few general principles apply to a broad range of situations, nearly every eukaryotic gene—or group of genes with coordinated regulation—represents a unique case. Virtually all genes that are transcribed by RNA polymerase II need the same suite of general factors to assemble an initiation complex, but the assembly of this complex and its ultimate level of transcription depend on specific transcription factors that in combination make up the **transcription complex** (figure 16.12).

The makeup of eukaryotic promoters, therefore, is either very simple, if we consider only what is needed for the initiation complex, or very complicated, if we consider all factors that may bind in a complex and affect transcription. This kind of combinatorial gene regulation leads to great flexibility because it can respond to the many signals a cell may receive affecting transcription, allowing integration of these signals.

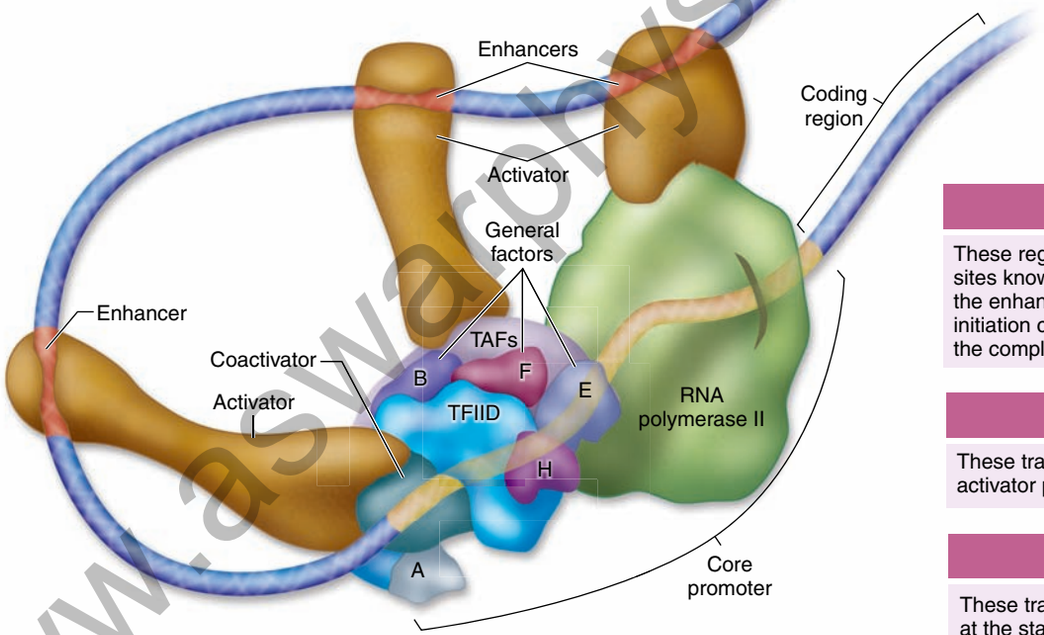
Inquiry question

How do eukaryotes coordinate the activation of many genes whose transcription must occur at the same time?

Learning Outcomes Review 16.4

In eukaryotes, initiation requires general transcription factors that bind to the promoter and recruit RNA polymerase II to form an initiation complex. General factors produce the basal level of transcription. Specific transcription factors, which bind to enhancer sequences, can increase the level of transcription. Enhancers can act at a distance because DNA can loop, bringing an enhancer and a promoter closer together. Additional coactivators and mediators link certain specific transcription factors to RNA polymerase II.

- ***What would be the effect of a mutation that results in the loss of a general transcription factor versus the loss of a specific factor?***



Activators

These regulatory proteins bind to DNA at distant sites known as enhancers. When DNA folds so that the enhancer is brought into proximity with the initiation complex, the activator proteins interact with the complex to increase the rate of transcription.

Coactivators

These transcription factors transmit signals from activator proteins to the general factors.

General Factors

These transcription factors position RNA polymerase at the start of a protein-coding sequence and then release the polymerase to initiate transcription.

Figure 16.12 Interactions of various factors within the transcription complex. All specific transcription factors bind to enhancer sequences that may be distant from the promoter. These proteins can then interact with the initiation complex by DNA looping to bring the factors into proximity with the initiation complex. As detailed in the text, some transcription factors, called activators, can directly interact with the RNA polymerase II or the initiation complex, whereas others require additional coactivators.

16.5 Eukaryotic Chromatin Structure

Learning Outcomes

1. *Describe how chromatin structure can affect gene expression.*
2. *Explain the function of chromatin remodeling complexes.*

Eukaryotes have the additional gene expression hurdle of possessing DNA that is packaged into chromatin. The packaging of DNA first into nucleosomes and then into higher order chromatin structures is now thought to be directly related to the control of gene expression.

Chromatin structure at its lowest level is the organization of DNA and histone proteins into *nucleosomes* (see chapter 10). These nucleosomes may block binding of transcription factors and RNA polymerase II at the promoter.

The higher order organization of chromatin, which is not completely understood, appears to depend on the state of the histones in nucleosomes. Histones can be modified to result in a greater condensation of chromatin, making promoters even less accessible for protein–DNA interactions. A chromatin remodeling complex exists that can make DNA more accessible.

Both DNA and histone proteins can be modified

Chemical *methylation* of the DNA was once thought to play a major role in gene regulation in vertebrate cells. The addition of a methyl group to cytosine creates 5-methylcytosine, but this change has no effect on its base-pairing with guanine (figure 16.13). Similarly, the addition of a methyl group to uracil produces thymine, which clearly does not affect base-pairing with adenine.

Many inactive mammalian genes are methylated, and it was tempting to conclude that methylation caused the inactivation. But methylation is now viewed as having a less direct role, blocking the accidental transcription of "turned-off" genes. Vertebrate cells apparently possess a protein that binds to clusters of 5-methylcytosine, preventing transcriptional activators from gaining access to the DNA. DNA methylation in vertebrates thus ensures that once a gene is turned off, it stays off.

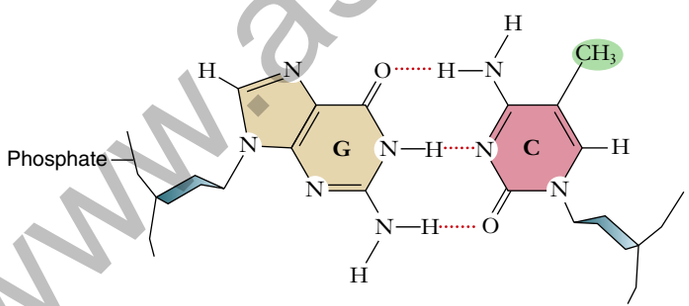


Figure 16.13 DNA methylation. Cytosine is methylated, creating 5-methylcytosine. Because the methyl group *(green)* is positioned to the side, it does not interfere with the hydrogen bonds of a G—C base-pair, but it can be recognized by proteins.

The histone proteins that form the core of the nucleosome (chapter 10) can also be modified. This modification is correlated with active versus inactive regions of chromatin, similar to the methylation of DNA just described. Histones can also be methylated, and this alteration is generally found in inactive regions of chromatin. Finally, histones can be modified by the addition of an acetyl group, and this addition is correlated with active regions of chromatin.

Some transcription activators alter chromatin structure

The control of eukaryotic transcription requires the presence of many different factors to activate transcription. Some activators seem to interact directly with the initiation complex or with coactivators that themselves interact with the initiation complex, as described earlier. Other cases are not so clear. The emerging consensus is that some coactivators have been shown to be histone acetylases. In these cases, it appears that transcription is increased by removing higher order chromatin structure that would prevent transcription (figure 16.14). Some corepressors have been shown to be histone deacetylases as well.

These observations have led to the suggestion that a "histone code" might exist, analogous to the genetic code.

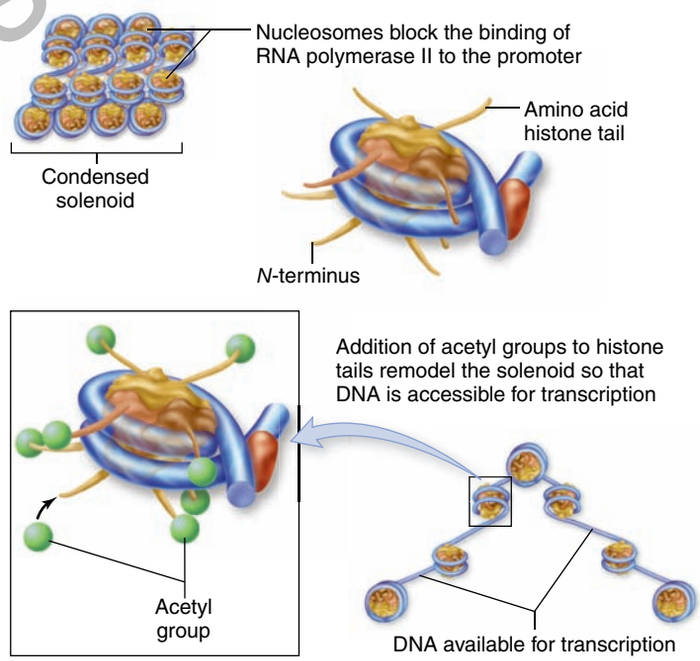


Figure 16.14 Histone modification affects chromatin structure. DNA in eukaryotes is organized first into nucleosomes and then into higher order chromatin structures. The histones that make up the nucleosome core have amino tails that protrude. These amino tails can be modified by the addition of acetyl groups. The acetylation alters the structure of chromatin, making it accessible to the transcription apparatus.

This histone code is postulated to underlie the control of chromatin structure and, thus, of access of the transcription machinery to DNA.

Chromatin-remodeling complexes also change chromatin structure

The outline of how alterations to chromatin structure can regulate gene expression are beginning to emerge. A key discovery is the existence of so-called **chromatin-remodeling complexes.** These large complexes of proteins include enzymes that modify histones and DNA and that also change chromatin structure itself.

One class of these remodeling factors, ATP-dependent chromatin remodeling factors, function as molecular motors that affect DNA and histones. These ATP-dependent remodeling factors use energy from ATP to alter the relationships between histones and DNA. They can catalyze four different changes in histone/DNA binding (figure 16.15): 1) nucleosome sliding along DNA, which changes the position of a nucleosome on the DNA; 2) create a remodeled state where DNA is more accessible; 3) removal of nucleosomes from DNA; and 4) replacement of histones with variant histones. These functions all act to make DNA more accessible to regulatory proteins that in turn, affect gene expression.

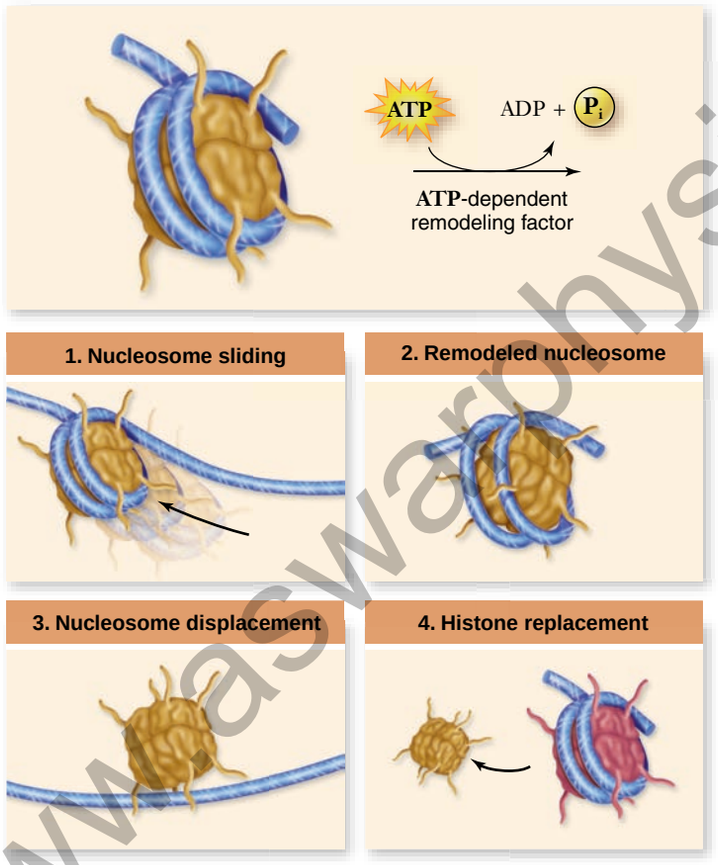


Figure 16.15 Function of ATP-dependent remodeling factors. ATP dependent remodeling factors use the energy from ATP to alter chromatin structure. They can *(1)* slide nucleosomes along DNA to reveal binding sites for proteins; *(2)* create a remodeled state of chromatin where the DNA is more accessible; *(3)* completely remove nucleosomes from DNA; and *(4)* replace histones in nucleosomes with variant histones.

Learning Outcomes Review 16.5

Eukaryotic DNA is packaged into chromatin, adding another structural challenge to transcription. Changes in chromatin structure correlate with modification of DNA and histones, and access to DNA by transcriptional regulators requires changes in chromatin structure. Some transcriptional activators modify histones by acetylation. Large chromatin-remodeling complexes include enzymes that alter the structure of chromatin, making DNA more accessible to regulatory proteins.

- ***Genes that are turned on in all cells are called "housekeeping" genes. Explain the idea behind this name.***

16.6 Eukaryotic Posttranscriptional Regulation

Learning Outcomes

1. *Explain how small RNAs can affect gene expression.*
2. *Differentiate between the different kinds of posttranscriptional regulation.*

The separation of transcription in the nucleus and translation in the cytoplasm in eukaryotes provides possible points of regulation that do not exist in prokaryotes. For many years we thought of this as "alternative" forms of regulation, but it now appears that they play a much more central role than previously suspected. In this section we will consider several of these mechanisms to control gene expression beginning with the exciting new area of regulation by small RNAs.

Small RNAs act after transcription to control gene expression

The study of development has lead to a number of important insights into the regulation of gene expression. A striking example is the discovery of small RNAs that affect gene expression. A mutant isolated in the worm *C. elegans* called *lin-4* was known to alter developmental timing, a so-called heterochronic mutant. Genetic studies had shown that this gene regulated another gene, *lin-14*. When the *lin-4* gene was isolated by Ambros, Lee, and Feinbaum in 1992, they showed that it did not encode a protein product. Instead, the *lin-4* gene encoded only two small RNA molecules, one of 22 nt and one of 61 nt. Further the 22 nt RNA was derived from the longer 61 nt RNA. Further work showed that this small RNA was complementary to a region in another heterochronic gene, *lin-14*. A model was developed where the *lin-4* RNA acted as a translational repressor of the *lin-14* mRNA

(figure 16.16). Although it was not called that at the time, this was the first identified micro RNA, or miRNA.

A completely different line of inquiry involved the use of double-stranded RNAs to turn off gene expression. This has been shown to act via another class of small RNA called small interfering RNAs, or siRNAs. These may be experimenatally introduced, derived from invading viruses, or even encoded in the genome. The use of siRNA to control gene expression revealed the existence of cellular mechanisms for the control of gene expression via small RNAs.

Since its discovery, gene silencing by small RNAs has been a source of great interest for both its experimental uses, and as an explanation for posttranslational control of gene expression. As these small RNAs have been studied in a variety of systems, it has led to a proliferation of terms to describe them. Recent research has uncovered a wealth of new types of small RNAs, but we will confine ourselves to the two classes of miRNA and siRNA as these are well established and illustrate the RNA silencing machinery.

SCIENTIFIC THINKING

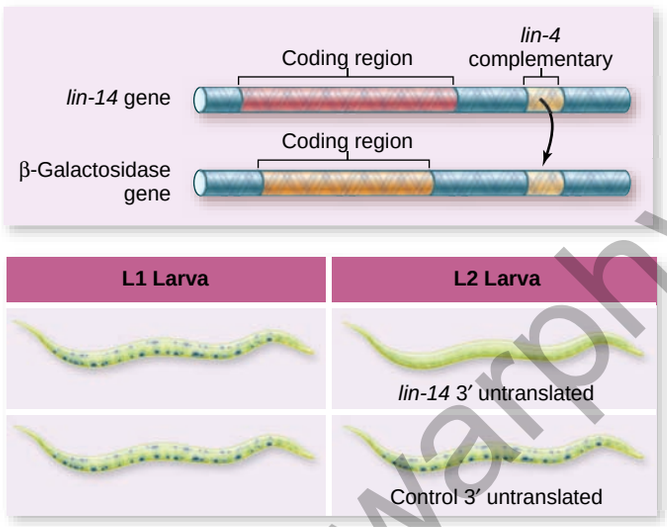
Hypothesis: *The region of the* lin-14 *gene complementary to the* lin-4 *miRNA controls* lin-14 *expression.*

Prediction: *If the* lin-4 *complementary region of the* lin-14 *gene is spliced into a reporter gene, then this reporter gene should show regulation similar to* lin-14.

Test: *Recombinant DNA is used to make two versions of a reporter gene (β-galactosidase). In transgenic worms (*C. elegans*), expression of the reporter gene produces a blue color.*

1. The β-galactosidase gene with the lin-14 *3′ untranslated region containing the* lin-4 *complementary region*

2. The β-galactosidase gene with a control 3′ untranslated region lacking the lin-4 *complementary region*



Result:

1. Transgenic worms with reporter gene plus lin-14 *3′ untranslated region show expression in L1 but not L2 stage larvae. This is the pattern expected for the* lin-14 *gene, which is controlled by* lin-4.

2. Transgenic worms with reporter gene plus control 3′ untranslated region do not show expression pattern expected for control by lin-4.

Conclusion: *The 3′ untranslated region from* lin-14 *is sufficient to turn off gene expression in L2 larvae.*

Further Experiments: *What expression pattern would you predict for these constructs in a mutant that lacks* lin-4 *function?*

Figure 16.16 Control of *lin-14* gene expression. The *lin-14* gene is controlled by the *lin-4* gene. This is mediated by a region of the 3′ untranslated region of the *lin-14* mRNA that is complementary to *lin-4* miRNA.

miRNA genes

The discovery of the role of miRNAs in gene expression initially appeared to be confined to nematodes as the *lin-4* gene did not have any obvious homologs in other systems. Seven years later, a second gene, *let-7*, was discovered in the same pathway in *C. elegans*. The *let-7* gene also encoded a 22 nt RNA that could influence translation. In this case, homologs for *let-7* were immediately found in both *Drosophila* and in humans.

As an increasing number of miRNAs were discovered in different organisms, miRNA gene discovery has turned to computer searching and high throughput methods such as microarrays and new high throughput sequencing. A database devoted to miRNAs currently lists 695 known human miRNA sequences.

Genes for miRNA are found in a variety of locations including the introns of expressed genes, and they are often clustered with multiple miRNAs in a single transcription unit. They are also found in regions of the genome that were previously considered transcriptionally silent. This finding is particularly exciting because other work looking at transcription across animal genomes has found that much we thought was transcriptionally silent, is actually not.

miRNA biogenesis and function

The production of a functional miRNA begins in the nucleus, and ends in the cytoplasm with a ~22 nt RNA that functions to repress gene expression (figure 16.17). The initial transcript of an miRNA gene is by RNA polymerase II producing a transcript called the Pri-miRNA. The region of this transcript containing the miRNA can fold back on itself and base-pair to form a stem and loop structure. This is cleaved in the nucleus by a nuclease called Drosha that trims the miRNA to just the stem and loop structure, which is now called the pre-miRNA. This pre-miRNA is exported from the nucleus through a nuclear pore bound to the protein exportin 5. Once in the cytoplasm the pre-miRNA is further cleaved by another nuclease called Dicer to produce a short double-stranded RNA containing the miRNA. The miRNA is loaded into a complex of proteins called an RNA induced silencing complex, or RISC. The RISC includes the RNA-binding protein Argonaute (Ago), which interacts with the miRNA. The complementary strand is either removed by a nuclease, or is removed during the loading process.

At this point, the RISC is targeted to repress the expression of other genes based on sequence complementarity to the miRNA. The complementary region is usually in the 3′ untranslated region of genes, and the result can be cleavage of the mRNA or inhibition of translation. It appears that in animals,

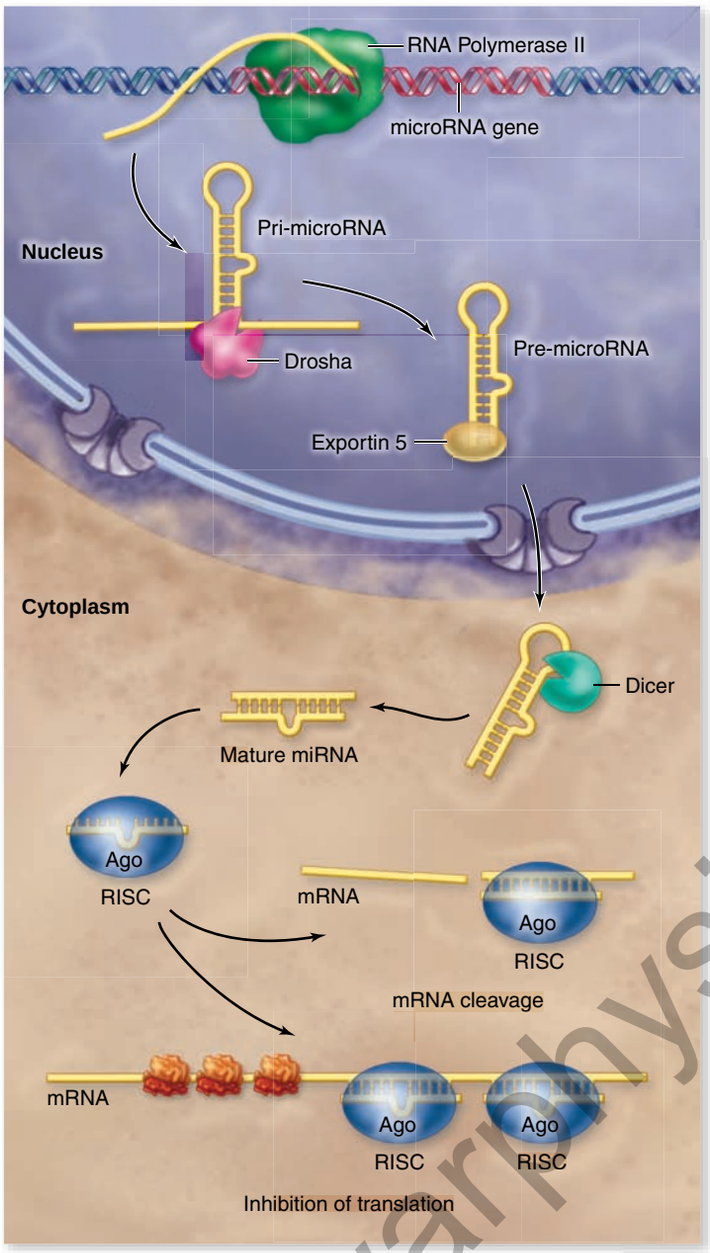


Figure 16.17 Biogenesis and function of miRNA. Genes for miRNAs are transcribed by RNA polymerase II to produce a Pri-miRNA. This is processed by the Drosha nuclease to produce the Pre-miRNA, which is exported from the nucleus bound to export factor Exportin 5. Once in the cytoplasm, the pre-miRNA is processed by Dicer nuclease to produce the mature miRNA. The miRNA is loaded into a RISC, which can act to either cleave target mRNAs, or to inhibit translation of target mRNAs.

the inhibition of translation is more common than the cleavage of the mRNA, although the precise mechanism of this inhibition is still unclear. In plants, the cleavage of the mRNA by the RISC is common and seems to be related to the more precise complementarity found between plant miRNAs and their targets compared to animal systems.

RNA interference

Small RNA mediated gene silencing has been known for a number of years. There has been some confusion created by observations in different systems leading to multiple names for similar phenomenon. Thus RNA interference, cosuppression, and posttranscriptional gene silencing all act through similar biochemical mechanisms. The term RNA interference is currently the most common and involves the production of siRNAs.

The production of siRNAs is similar to miRNAs except that they arise from a long double-stranded RNA (figure 16.18). This can be either a very long region of self complementarity, or from two complementary RNAs. These long double-stranded RNAs are processed by Dicer to yield multiple siRNAs that are loaded into an Ago containing RISC. The siRNAs usually have near-perfect complementarity to their target mRNAs, and the result is cleavage of the mRNA by the siRNA containing RISC.

The source of the double-stranded RNA to produce siRNAs can be either from the cell, or from outside the cell. From the cell itself, there are genes that produce RNAs with long regions of self-complementarity that fold back to produce a substrate for Dicer in the cytoplasm. They can also arise from repeated regions of the genome that contain transposable elements. Exogenous double-stranded RNAs can be introduced

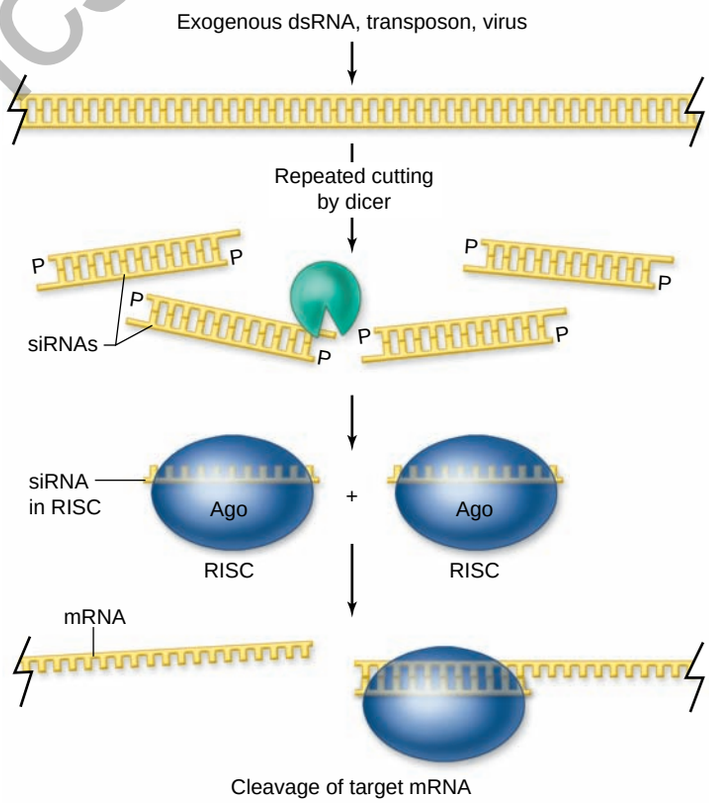


Figure 16.18 Biogenesis and function of siRNA. SiRNAs can arise from a variety of sources that all produce long double-stranded regions of RNA. The double-stranded RNA is processed by Dicer nuclease to produce a number of siRNAs that are each loaded onto their own RISC. The RISC then cleaves target mRNA.

experimentally, or by infection with a virus. The last origin for double-stranded RNAs may point to the evolution of the RNA silencing machinery as a form of antiviral defense.

Distinguishing miRNAs and siRNAs

The biogenesis of both miRNA and siRNA involves cleavage by Dicer, and incorporation into a RISC complex. The main thing that distinguishes these two types of molecules is their targets: miRNAs tend to repress genes different from their origin, while endogenous siRNAs tend to repress the genes they were derived from. Additionally siRNAs are used experimentally to turn off the expression of genes. This takes advantage of the cellular machinery to turn off a gene based with a double-stranded RNA corresponding to the gene of interest.

There are other differences between the two classes of small RNA. When multiple species are examined, miRNAs tend to be evolutionarily conserved while siRNAs do not. While the biogenesis is similar in terms of the nucleases involved, the actual structure of the double-stranded RNAs is not the same. The transcript of miRNA genes form stem-loop structures containing the miRNA while the double-stranded RNAs generating siRNAs may be bimolecular, or very long stem-loops. These longer double-stranded regions lead to multiple siRNAs while there is only a single miRNA generated from a pre-miRNA.

Small RNAs can mediate heterochromatin formation

RNA silencing pathways have also been implicated in the formation of heterochromatin in fission yeast, plants, and *Drosophila*. In fission yeast, centromeric heterochromatin formation is driven by siRNAs produced by the action of the Dicer nuclease. This heterochromatin formation also involves modification of histone proteins, and thus connects RNA interference with chromatin remodeling complexes in this system. It is not yet clear how widespread this is.

In *Drosophila*, there is genetic evidence for the involvement of the RNA interference machinery in the formation of heterochromatin. This is particularly clear in the germ line where a specific class of small RNA appears to be involved in silencing transposons during spermatogenesis and oogenesis. There is also evidence that a similar mechanism may act in vertebrates.

Plants are an interesting case in that they have a variety of small RNA species. The RNA interference pathway is more complex than in animals with multiple forms of Dicer nuclease proteins and Argonaute RNA binding proteins. One class of endogenous siRNA can lead to heterochromatin formation by DNA methylation and histone modification.

Alternative splicing can produce multiple proteins from one gene

As noted in the preceding chapter, splicing of pre-mRNA is one of the processes leading to mature mRNA. Many of these splicing events may produce different mRNAs from a single primary transcript by alternative splicing. This mechanism allows another level of control of gene expression.

Alternative splicing can change the splicing events that occur during different stages of development or in different tissues. An example of developmental differences is found in *Drosophila*, in which sex determination is the result of a complex series of alternative splicing events that differ in males and females.

An excellent example of tissue-specific alternative splicing in action is found in two different human organs: the thyroid gland and the hypothalamus. The thyroid gland is responsible for producing hormones that control processes such as metabolic rate. The hypothalamus, located in the brain, collects information from the body (for example, salt balance) and releases hormones that in turn regulate the release of hormones from other glands, such as the pituitary gland. (You'll learn more about these glands in chapter 46.)

These two organs produce two distinct hormones: *calcitonin* and *CGRP* (calcitonin gene-related peptide) as part of their function. Calcitonin controls calcium uptake and the balance of calcium in tissues such as bones and teeth. CGRP is involved in a number of neural and endocrine functions. Although these two hormones are used for very different physiological purposes, they are produced from the same transcript (figure 16.19).

The synthesis of one product versus another is determined by tissue-specific factors that regulate the processing of the primary transcript. In the case of calcitonin and CGRP, pre-mRNA splicing is controlled by different factors that are present in the thyroid and in the hypothalamus.

RNA editing alters mRNA after transcription

In some cases, the editing of mature mRNA transcripts can produce an altered mRNA that is not truly encoded in the genome—an unexpected possibility. RNA editing was first discovered as the insertion of uracil residues into some RNA transcripts in protozoa, and it was thought to be an anomaly.

RNA editing of a different sort has since been found in mammalian species, including humans. In this case, the editing involves chemical modification of a base to change its base-pairing properties, usually by deamination. For example, both deamination of cytosine to uracil and deamination of adenine to inosine have been observed (inosine pairs as G would during translation).

Apolipoprotein B

The human protein apolipoprotein B is involved in the transport of cholesterol and triglycerides. The gene that encodes this protein, *apoB*, is large and complex, consisting of 29 exons scattered across almost 50 kilobases (kb) of DNA.

The protein exists in two isoforms: a full-length APOB100 form and a truncated APOB48 form. The truncated form is due to an alteration of the mRNA that changes a codon for glutamine to one that is a stop codon. Furthermore, this editing occurs in a tissue-specific manner; the edited form appears only in the intestine, whereas the liver makes only the full-length form. The full-length APOB100 form is part of the low-density lipoprotein (LDL) particle that carries cholesterol. High levels of serum LDL are thought

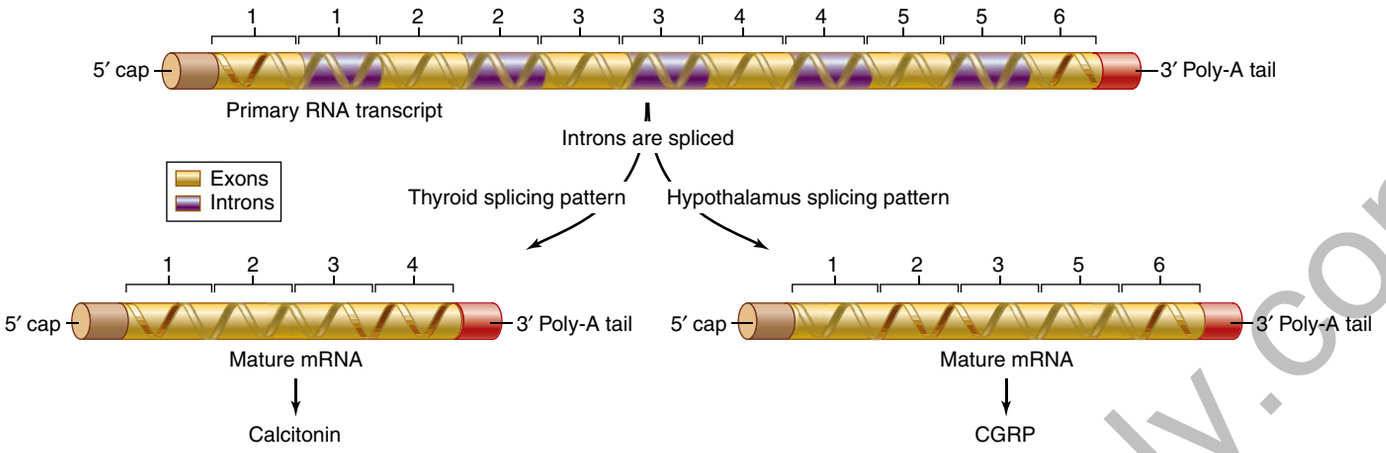


Figure 16.19 Alternative splicing. Many primary transcripts can be spliced in different ways to give rise to multiple mRNAs. In this example, in the thyroid the primary transcript is spliced to contain four exons encoding the protein calcitonin. In the hypothalamus the fourth exon, which contains the poly-A site used in the thyroid, is skipped and two additional exons are added to encode the protein calcitonin-gene-related peptide (CGRP).

to be a major predictor of atherosclerosis in humans. It does not appear that editing has any effect on the levels of the intestine-specific transcript.

The 5-HT serotonin receptor

RNA editing has also been observed in some brain receptors for opiates in humans. One of these receptors, the serotonin (5-HT) receptor, is edited at multiple sites to produce a total of 12 different isoforms of the protein.

It is unclear how widespread these forms of RNA editing are, but they are further evidence that the information encoded within genes is not the end of the story for protein production.

mRNA must be transported out of the nucleus for translation

Processed mRNA transcripts exit the nucleus through the nuclear pores (described in chapter 4). The passage of a transcript across the nuclear membrane is an active process that requires the transcript to be recognized by receptors lining the interior of the pores. Specific portions of the transcript, such as the poly-A tail, appear to play a role in this recognition.

There is little hard evidence that gene expression is regulated at this point, although it could be. On average, about 10% of primary transcripts consists of exons that will make up mRNA sequences, but only about 5% of the total mRNA produced as primary transcript ever reaches the cytoplasm. This observation suggests that about half of the exons in primary transcripts never leave the nucleus, but it is unclear whether the disappearance of this mRNA is selective.

Initiation of translation can be controlled

The translation of a processed mRNA transcript by ribosomes in the cytoplasm involves a complex of proteins called *translation factors*. In at least some cases, gene expression is regulated by modification of one or more of these factors. In other instances, **translation repressor proteins** shut down translation by binding to the beginning of the transcript, so that it cannot attach to the ribosome.

In humans, the production of ferritin (an iron-storing protein) is normally shut off by a translation repressor protein called aconitase. Aconitase binds to a 30-nt sequence at the beginning of the ferritin mRNA, forming a stable loop to which ribosomes cannot bind. When iron enters the cell, the binding of iron to aconitase causes the aconitase to dissociate from the ferritin mRNA, freeing the mRNA to be translated and increasing ferritin production 100-fold.

The degradation of mRNA is controlled

Another aspect that affects gene expression is the stability of mRNA transcripts in the cell cytoplasm. Unlike prokaryotic mRNA transcripts, which typically have a half-life of about 3 min, eukaryotic mRNA transcripts are very stable. For example, β-globin gene transcripts have a half-life of over 10 hr, an eternity in the fast-moving metabolic life of a cell.

The transcripts encoding regulatory proteins and growth factors, however, are usually much less stable, with half-lives of less than 1 hr. What makes these particular transcripts so unstable? In many cases, they contain specific sequences near their 3′ ends that make them targets for enzymes that degrade mRNA. A sequence of A and U nucleotides near the 3′ poly-A tail of a transcript promotes removal of the tail, which destabilizes the mRNA.

Loss of the poly-A tail leads to rapid degradation by 3′ to 5′ RNA exonucleases. Another consequence of this loss is the stimulation of decapping enzymes that remove the 5′ cap leading to degradation by 5′ to 3′ RNA exonucleases.

Other mRNA transcripts contain sequences near their 3′ ends that are recognition sites for endonucleases, which cause these transcripts to be digested quickly. The short half-lives of the mRNA transcripts of many regulatory genes are critical to the function of those genes because they enable the levels of regulatory proteins in the cell to be altered rapidly.

A review of various methods of posttranscriptional control of gene expression is provided in figure 16.20.

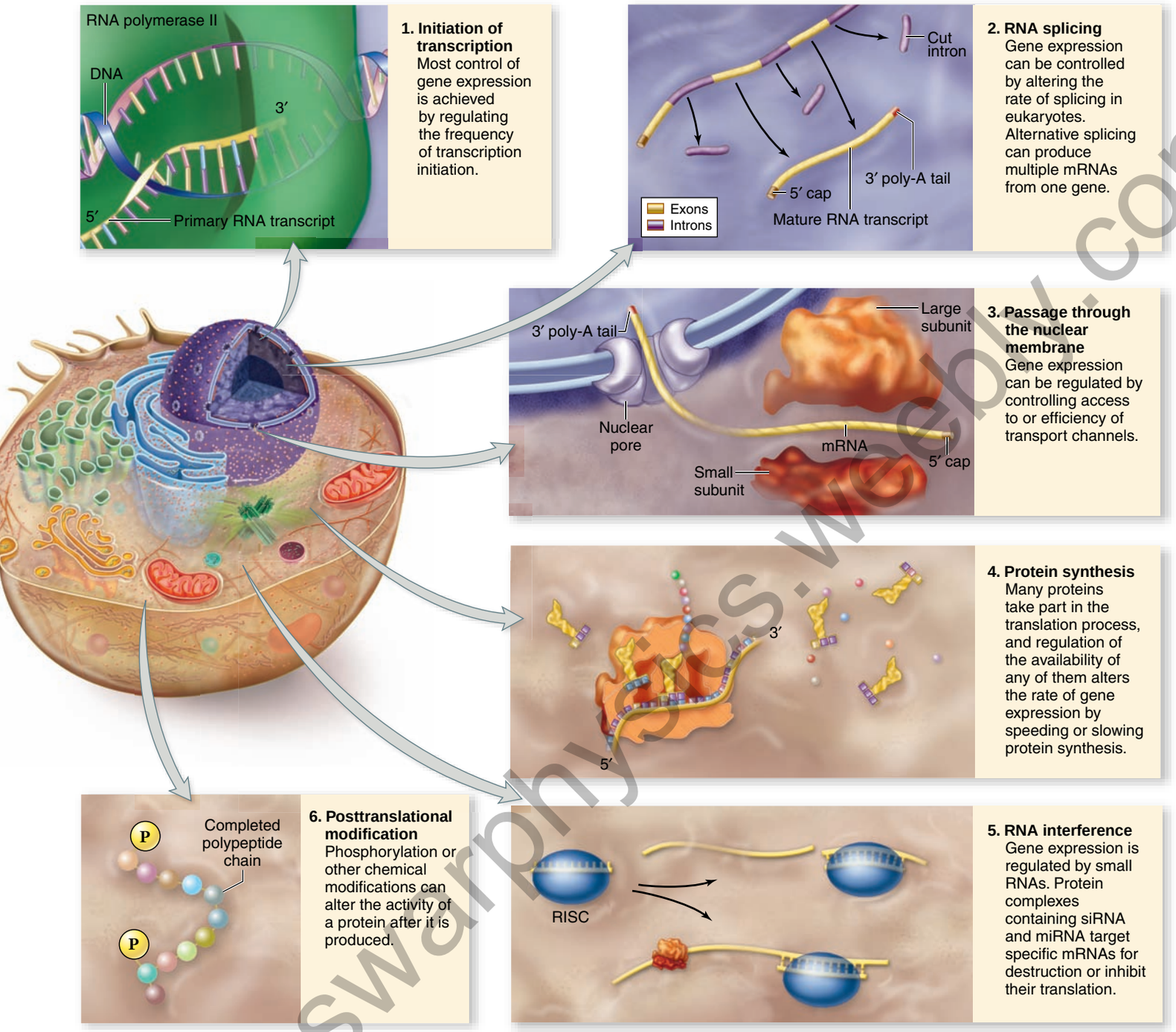


Figure 16.20 **Mechanisms for control of gene expression in eukaryotes.**

Learning Outcomes Review 16.6

Small RNAs control gene expression by either selective degradation of mRNA, inhibition of translation, or alteration of chromatin structure. Multiple mRNAs can be formed from a single gene via alternative splicing, which can be tissue- and developmentally specific. The sequence of an mRNA transcript can also be altered by RNA editing.

- *How could the phenomenon of RNA interference be used in drug design?*

16.7 *Protein Degradation*

Learning Outcomes

1. *Describe the role of ubiquitin in the degradation of proteins.*
2. *Explain the function of the proteasome.*

If all of the proteins produced by a cell during its lifetime remained in the cell, serious problems would arise. Protein labeling studies in the 1970s indicated that eukaryotic cells turn over proteins in a controlled manner. That is, proteins are continually being synthesized and degraded. Although this protein turnover is not as rapid as in prokaryotes, it indicates that a system regulating protein turnover is important.

Proteins can become altered chemically, rendering them nonfunctional; in addition, the need for any particular protein may be transient. Proteins also do not always fold correctly, or they may become improperly folded over time. These changes can lead to loss of function or other chemical behaviors, such as aggregating into insoluble complexes. In fact, a number of neurodegenerative diseases, such as Alzheimer dementia, Parkinson disease, and mad cow disease, are related to proteins that aggregate, forming characteristic plaques in brain cells. Thus, in addition to normal turnover of proteins, cells need a mechanism to get rid of old, unused, and incorrectly folded proteins.

Enzymes called **proteases** can degrade proteins by breaking peptide bonds, converting a protein into its constituent amino acids. Although there is an obvious need for these enzymes, they clearly cannot be floating around in the cytoplasm active at all times.

One way that eukaryotic cells handle such problems is to confine destructive enzymes to a specific cellular compartment. You may recall from chapter 4 that lysosomes are vesicles that contain digestive enzymes, including proteases. Lysosomes are used to remove proteins and old or nonfunctional organelles, but this system is not specific for particular proteins. Cells need another regulated pathway to remove proteins that are old or unused, but leave the rest of cellular proteins intact.

Addition of ubiquitin marks proteins for destruction

Eukaryotic cells solve this problem by marking proteins for destruction, then selectively degrading them. The mark that cells use is the attachment of a **ubiquitin** molecule. Ubiquitin, so named because it is found in essentially all eukaryotic cells (that is, it is ubiquitous), is a 76–amino-acid protein that can exist as an isolated molecule or in longer chains that are attached to other proteins.

The longer chains are added to proteins in a stepwise fashion by an enzyme called *ubiquitin ligase* (figure 16.21). This reaction requires ATP and other proteins, and it takes place in a multistep, regulated process. Proteins that have a ubiquitin chain attached are called *polyubiquitinated*, and this state is a signal to the cell to destroy this protein.

Two basic categories of proteins become ubiquitinated: those that need to be removed because they are improperly folded or nonfunctional, and those that are produced and degraded in a controlled fashion by the cell. An example of the latter are the cyclin proteins that help to drive the cell cycle (chapter 10). When these proteins have fulfilled their role in active division of the cell, they become polyubiquitinated and are removed. In this way, a cell can control entry into cell division or maintain a nondividing state.

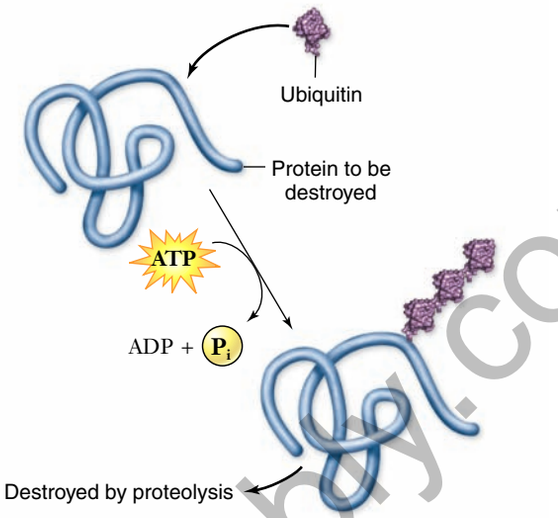


Figure 16.21 Ubiquitination of proteins. Proteins that are to be degraded are marked with ubiquitin. The enzyme ubiquitin ligase uses ATP to add ubiquitin to a protein. When a series of these have been added, the polyubiquitinated protein is destroyed.

The proteasome degrades polyubiquitinated proteins

The cellular organelle that degrades proteins marked with ubiquitin is the **proteasome,** a large cylindrical complex that proteins enter at one end and exit the other as amino acids or peptide fragments (figure 16.22).

The proteasome complex contains a central region that has protease activity and regulatory components at each end. Although not membrane-bounded, this organelle can be thought of as a form of compartmentalization on a very small scale. By using a two-step process, first to mark proteins for destruction, then to process them through a large complex, proteins to be degraded are isolated from the rest of the cytoplasm.

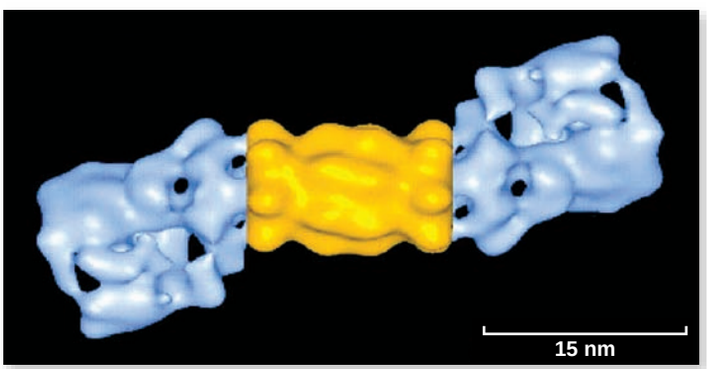


Figure 16.22 The *Drosophila* proteasome. The central complex contains the proteolytic activity, and the flanking regions act as regulators. Proteins enter one end of the cylinder and are cleaved to peptide fragments that exit the other end.

The process of ubiquitination followed by degradation by the proteasome is called the *ubiquitin–proteasome pathway*. It can be thought of as a cycle in that the ubiquitin added to proteins is not itself destroyed in the proteasome. As the proteins are degraded, the ubiquitin chain itself is simply cleaved back into ubiquitin units that can then be reused (figure 16.23).

Learning Outcomes Review 16.7

Control of protein degradation in eukaryotes involves addition of the protein ubiquitin, which marks the protein for destruction. The proteasome, a cylindrical complex with protease activity in its center, recognizes ubiquitinated proteins and breaks them down, much like a shredder destroys documents. Ubiquitin is recycled unchanged.

- ***If the ubiquitination process was not tightly controlled, what effect would this have on a cell?***

Inquiry question

What are two reasons a cell would polyubiquitinate a polypeptide?

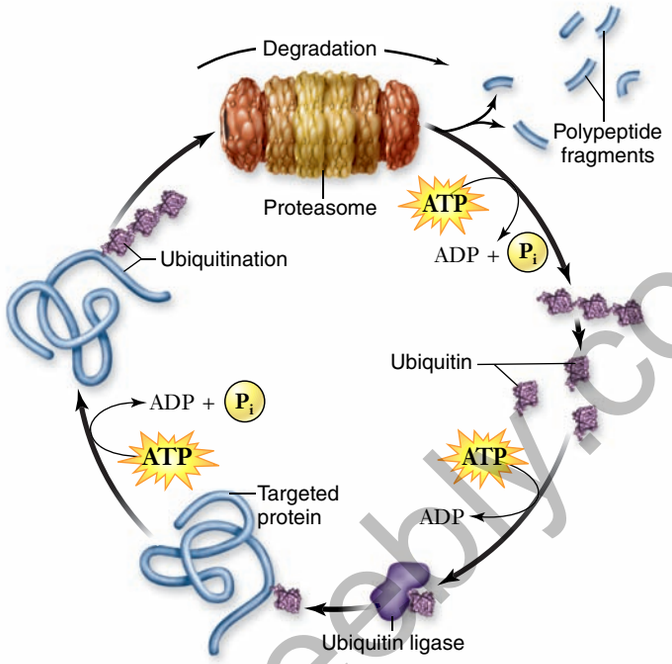
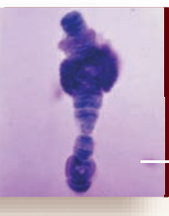


Figure 16.23 Degradation by the ubiquitin–proteasome pathway. Proteins are first ubiquitinated, then enter the proteasome to be degraded. In the proteasome, the polyubiquitin is removed and then is later "deubiquitinated" to produce single ubiquitin molecules that can be reused.



Chapter Review

16.1 Control of Gene Expression

Control usually occurs at the level of transcription initiation.

Transcription is controlled by regulatory proteins that modulate the ability of RNA polymerase to bind to the promoter. These may either block transcription or stimulate it.

Control strategies in prokaryotes are geared to adjust to environmental changes.

Control strategies in eukaryotes are aimed at maintaining homeostasis.

16.2 Regulatory Proteins

Proteins can interact with DNA through the major groove.

A DNA double helix exhibits a major groove and a minor groove; bases in the major groove are accessible to regulatory proteins.

DNA-binding domains interact with specific DNA sequences.

A region of the regulatory protein that can bind to the DNA is termed a DNA-binding motif (see figure 16.2).

Several common DNA-binding motifs are shared by many proteins.

Common motifs include the helix-turn-helix motif, the homeodomain motif, the zinc finger motif, and the leucine zipper.

16.3 Prokaryotic Regulation

Control of transcription can be either positive or negative.

Negative control is mediated by proteins called repressors that interfere with transcription. Positive control is mediated by a class of regulatory proteins called activators that stimulate transcription.

Prokaryotes adjust gene expression in response to environmental conditions.

The *lac* operon is induced in the presence of lactose; that is, the enzymes to utilize lactose are only produced when lactose is present. The *trp* operon is repressed; that is, the enzymes needed to produce tryptophan are turned off when tryptophan is present.

The* lac *operon is negatively regulated by the* lac *repressor.

The *lac* operon is induced when the effector (allolactose) binds to the repressor, altering its conformation such that it no longer binds DNA (see figure 16.4).

The presence of glucose prevents induction of the* lac *operon

Maximal expression of the *lac* operon requires positive control by catabolite activator protein (CAP) complexed with cAMP. When glucose is low, cAMP is high. Glucose repression involves both inducer exclusion, in which lactose is prevented from entering the cell, and the control of CAP function by the level of glucose.

The* trp *operon is controlled by the* trp *repressor.

The *trp* operon is repressed when tryptophan, acting as a corepressor, binds to the repressor, altering its conformation such that it can bind to DNA and turn off the operon. This prevents expression in the presence of excess *trp*.

16.4 Eukaryotic Regulation

Transcription factors can be either general or specific.

General transcription factors are needed to assemble the transcription apparatus and recruit RNA polymerase II at the promoter. Specific factors act in a tissue- or time-dependent manner to stimulate higher rates of transcription.

Promoters and enhancers are binding sites for transcription factors.

General factors bind to the promoter to recruit RNA polymerase. Specific factors bind to enhancers, which may be distant from the promoter but can be brought closer by DNA looping.

Coactivators and mediators link transcription factors to RNA polymerase II (see figure 16.12).

Some, but not all, transcription factors require a mediator. The number of coactivators is small because a single coactivator can be used with multiple transcription factors.

The transcription complex brings things together.

16.5 Eukaryotic Chromatin Structure

In eukaryotes, DNA is wrapped around proteins called histones, forming nucleosomes. These may block binding of transcription factors to promoters and enhancers.

Both DNA and histone proteins can be modified.

Methylation of DNA bases, primarily cytosine, correlates with genes that have been "turned off." Methylation is associated with inactive regions of chromatin.

Some transcription activators alter chromatin structure.

Acetylation of histones results in active regions of chromatin.

Chromatin-remodeling complexes also change chromatin structure.

Chromatin-remodeling complexes contain enzymes that move, reposition, and transfer nucleosomes.

16.6 Eukaryotic Posttranscriptional Regulation

Small RNAs act after transcription to control gene expression.

RNA interference is mediated by siRNAs formed by cleavage of double-stranded RNA by the Dicer nuclease. The siRNA is bound to a protein, Argonaute, in an RNA Induced Silencing Complex (RISC). The RISC can cleave mRNA or inhibit translation. Another class of small RNA, miRNA, is formed by the action of two nucleases, Drosha and Dicer, on RNA stem-and-loop structures. These also form a RISC that can either degrade mRNA or stop translation.

Small RNAs can mediate heterochromatin formation.

In fission yeast, *Drosophila*, and plants, RNA interference pathways lead to the formation of heterochromatin.

Alternative splicing can produce multiple proteins from one gene.

In response to tissue-specific factors, alternative splicing of pre-mRNA from one gene can result in multiple proteins.

RNA editing alters mRNA after transcription.

mRNA must be transported out of the nucleus for translation.

Initiation of translation can be controlled.

Translation factors may be modified to control initiation; translation repressor proteins can bind to the beginning of a transcript so that it cannot attach to the ribosome.

The degradation of mRNA is controlled.

An mRNA transcript is relatively stable, but it may carry targets for enzymes that degrade it more quickly as needed by the cell.

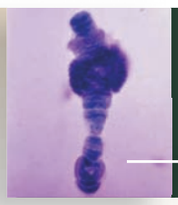
16.7 Protein Degradation

Addition of ubiquitin marks proteins for destruction.

In eukaryotes, proteins targeted for destruction have ubiquitin added to them as a marker.

The proteasome degrades polyubiquitinated proteins.

A cell organelle—the cylindrical proteasome—degrades ubiquitinated proteins that pass through it.



Review Questions

UNDERSTAND

1. In prokaryotes, control of gene expression usually occurs at the
   a. splicing of pre-mRNA into mature mRNA.
   b. initiation of translation.
   c. initiation of transcription.
   d. all of the above
2. Regulatory proteins interact with DNA by
   a. unwinding the helix and changing the pattern of base-pairing.
   b. binding to the sugar–phosphate backbone of the double helix.
   c. unwinding the helix and disrupting base-pairing.
   d. binding to the major groove of the double helix and interacting with base-pairs.
3. In *E. coli*, induction in the *lac* operon and repression in the *trp* operon are both examples of
   a. negative control by a repressor.
   b. positive control by a repressor.
   c. negative control by an activator.
   d. positive control by a repressor.
4. The *lac* operon is controlled by two main proteins. These proteins
   a. both act in a negative fashion.
   b. both act in a positive fashion.
   c. act in the opposite fashion, one negative and one positive.
   d. act at the level of translation.
5. In eukaryotes, binding of RNA polymerase to a promoter requires the action of
   a. specific transcription factors.
   b. general transcription factors.
   c. repressor proteins.
   d. inducer proteins.
6. In eukaryotes, the regulation of gene expression occurs
   a. only at the level of transcription.
   b. only at the level of translation.
   c. at the level of transcription initiation, or posttranscriptionally.
   d. only posttranscriptionally.

7. In the *trp* operon, the repressor binds to DNA
   a. in the absence of *trp*.
   b. in the presence of *trp*.
   c. in either the presence or absence of *trp*.
   d. only when *trp* is needed in the cell.

APPLY

1. The *lac* repressor, the *trp* repressor and CAP are all
   a. negative regulators of transcription.
   b. positive regulators of transcription.
   c. allosteric proteins that bind to DNA and an effector.
   d. proteins that can bind DNA or other proteins.
2. Specific transcription factors in eukaryotes interact with enhancers, which may be a long distance from the promoter. These transcription factors then
   a. alter the structure of the DNA between enhancer and promoter.
   b. do not interact with the transcription apparatus.
   c. can interact with the transcription apparatus via DNA looping.
   d. can interact with the transcription apparatus by removing the intervening DNA.
3. Repression in the *trp* operon and induction in the *lac* operon are both mechanisms that
   a. would only be possible with positive regulation.
   b. allow the cell to control the level of enzymes to fit environmental conditions.
   c. would only be possible with negative regulation.
   d. cause the cell to make the enzymes from these two operons all the time.
4. Regulation by small RNAs and alternative splicing are similar in that both
   a. act after transcription.
   b. act via RNA/protein complexes.
   c. regulate the transcription machinery.
   d. both a and b
5. Eukaryotic mRNAs differ from prokaryotic mRNAs in that they
   a. usually contain more than one gene.
   b. are colinear with the genes that encode them.
   c. are not colinear with the genes that encode them.
   d. both a and c
6. In the cell cycle, cyclin proteins are produced in concert with the cycle. This likely involves
   a. control of initiation of transcription of cyclin genes, and ubiquitination of cyclin proteins.
   b. alternative splicing of cyclin genes to produce different cyclin proteins.
   c. RNA editing to produce the different cyclin proteins.
   d. transcription/translation coupling.
7. A mechanism of control in *E. coli* not discussed in this chapter involves pausing of ribosomes allowing a transcription terminator to form in the mRNA. In eukaryotic fission yeast, this mechanism should
   a. be common since they are unicellular.
   b. not be common since they are unicellular.
   c. not occur as transcription occurs in the nucleus and translation in the cytoplasm.
   d. not occur due to possibility of alternative splicing.

SYNTHESIZE

1. You have isolated a series of mutants affecting regulation of the *lac* operon. All of these are constitutive, that is, they express the *lac* operon all the time. You also have both mutant and wild-type alleles for each mutant in all combinations, and on F′ plasmids, which can be introduced into cells to make the cell diploid for the relevant genes. How would you use these tools to determine which mutants affect DNA binding sites on DNA, and which affect proteins that bind to DNA?
2. Examples of positive and negative control of transcription can be found in the regulation of expression of the bacterial operons *lac* and *trp*. Use these two operon systems to describe the difference between positive and negative regulation.
3. What forms of eukaryotic control of gene expression are unique to eukaryotes? Could prokaryotes use the mechanisms, or are they due to differences in these cell types?
4. The number and type of proteins found in a cell can be influenced by genetic mutation and regulation of gene expression. Discuss how these two processes differ.

ONLINE RESOURCE

www.ravenbiology.com

Understand, Apply, and Synthesize—enhance your study with animations that bring concepts to life and practice tests to assess your understanding. Your instructor may also recommend the interactive eBook, individualized learning tools, and more.