Chapter 17

From Gene to Protein

Lecture Outline

Overview: The Flow of Genetic Information

- The information content of genes is in the form of specific sequences of nucleotides along the DNA strands.
- The DNA inherited by an organism leads to specific traits by dictating the synthesis of proteins and of RNA molecules involved in protein synthesis.
    - Proteins are the link between genotype and phenotype.
- **Gene expression**, the process by which DNA directs protein synthesis, includes two stages called transcription and translation.

Concept 17.1 Genes specify proteins via transcription and translation

The study of metabolic defects provided evidence that genes specify proteins.

- In 1902, Archibald Garrod was the first to suggest that genes dictate phenotype through enzymes that catalyze specific chemical reactions in the cell.
    - Garrod suggested that the symptoms of an inherited disease reflect a person's inability to synthesize a particular enzyme and referred to such diseases as "inborn errors of metabolism."
    - He speculated that alkaptonuria, a hereditary disease, is caused by the absence of an enzyme that breaks down a specific substrate, alkapton.
- Research conducted several decades later supported Garrod's hypothesis.
- Progress in linking genes and enzymes rested on the growing understanding that cells synthesize and degrade most organic molecules in a series of steps, a metabolic pathway.
- In the 1930s, George Beadle and Boris Ephrussi speculated that each mutation affecting eye color in *Drosophila* blocks pigment synthesis at a specific step by preventing production of the enzyme that catalyzes that step.
    - Neither the chemical reactions nor the enzymes that catalyze them were known at the time.
- Beadle and Edward Tatum were able to establish the link between genes and enzymes by exploring the metabolism of a bread mold, *Neurospora crassa*.
    - Beadle and Tatum bombarded *Neurospora* with X-rays and screened the survivors for mutants that differed in their nutritional needs from wild-type mold.

- Wild-type *Neurospora* can grow on a *minimal medium* of agar, inorganic salts, glucose, and the vitamin biotin.
    - Beadle and Tatum identified mutants that could not survive on the minimal medium because they were unable to synthesize certain essential molecules from the minimal ingredients.
    - Most of these nutritional mutants were able to survive on a *complete growth medium* that includes all 20 amino acids and a few other nutrients.
- One type of mutant required only the addition of the amino acid arginine to the minimal growth medium.
    - Beadle and Tatum concluded that this mutant was defective in the biochemical pathway that synthesizes arginine.
- Srb and Horowitz identified three classes of arginine-deficient mutants, each lacking a key enzyme at a different step in the synthesis of arginine.
    - They demonstrated this by growing these mutant strains in media that provided different intermediate molecules.
- These results provided strong evidence for the *one gene–one enzyme hypothesis*, which states that the function of a gene is to dictate the production of a specific enzyme.
- Later research refined the one gene–one enzyme hypothesis.
- Not all proteins are enzymes. Keratin, the structural protein of hair, and insulin, a hormone, are proteins and gene products.
    - This altered the hypothesis to one gene–one protein.
- Later research demonstrated that many proteins are composed of several polypeptides, each of which has its own gene.
    - The hypothesis was restated as the *one gene–one polypeptide hypothesis*.
- This hypothesis is not entirely accurate, however.
    - Many eukaryotic genes code for a set of closely related polypeptides in a process called alternative splicing.
    - Some genes code for RNA molecules that play important roles in cells, although they are never translated into protein.

Transcription and translation are the two main processes linking gene to protein.

- Genes provide the instructions for making specific proteins.
- The bridge between DNA and protein synthesis is the nucleic acid RNA.
    - RNA is chemically similar to DNA, except that it contains ribose as its sugar and substitutes the nitrogenous base uracil for thymine.
    - An RNA molecule usually consists of a single strand.
- In DNA or RNA, the four nucleotide monomers act like the letters of the alphabet to communicate information.
- The specific sequence of hundreds to thousands of nucleotides in a gene carries the information for the primary structure of proteins, the linear order of the 20 possible amino acids.

- Getting from DNA to protein requires two major stages: transcription and translation.
- During **transcription**, a DNA strand provides a template for the synthesis of a complementary RNA strand.
    - Just as a DNA strand provides a template for the synthesis of each new complementary strand during DNA replication, it provides a template for assembling a sequence of RNA nucleotides.
- Transcription of a protein-coding gene produces a **messenger RNA (mRNA)** molecule.
- **Translation** is the synthesis of a polypeptide, using the information in mRNA.
    - During **translation**, there is a change of language.
- The sites of translation are the **ribosomes**, complex particles that facilitate the orderly assembly of amino acids into polypeptide chains.
- Transcription and translation occur in all organisms, from all three domains of life.
- The basic mechanics of transcription and translation are similar in eukaryotes and bacteria.
- Because bacteria lack nuclei, their DNA is not segregated from ribosomes and other protein-synthesizing equipment.
    - This allows the coupling of transcription and translation.
    - Ribosomes attach to the leading end of an mRNA molecule while transcription is still in progress.
- In a eukaryotic cell, transcription occurs in the nucleus, and translation occurs at ribosomes in the cytoplasm.
- The transcription of a protein-coding eukaryotic gene results in *pre-mRNA.*
- The initial RNA transcript of any gene is called a **primary transcript**, while further processing yields the finished mRNA.
- To summarize: Genes program protein synthesis via genetic messages in the form of messenger RNA.
- The molecular chain of command in a cell has a directional flow of genetic information: DNA → RNA → protein.
    - Francis Crick dubbed this concept the *central dogma* in 1956.
    - Although some RNA molecules can act as templates for DNA, this is a rare exception that does not invalidate the idea that, in general, genetic information flows from DNA to RNA to protein.

In the genetic code, nucleotide triplets specify amino acids.

- If the genetic code consisted of a single nucleotide or even pairs of nucleotides per amino acid, there would not be enough combinations (4 and 16, respectively) to code for all 20 amino acids.
- Triplets of nucleotide bases are the smallest units of uniform length that can code for all the amino acids.
    - With a **triplet code**, three consecutive bases specify an amino acid, creating 4^3 (64) possible code words.

- The genetic instructions for a polypeptide chain are written in DNA as a series of nonoverlapping three-nucleotide words, which are then translated into an amino acid chain.
- During transcription, one DNA strand, the **template strand**, provides a template for ordering the sequence of nucleotide bases in an mRNA transcript.
    - A given DNA strand can be the template strand for some genes along a DNA molecule, while for other genes in other regions, the complementary strand may function as the template.
- The complementary mRNA molecule is synthesized according to base-pairing rules, except that uracil is the complementary base to adenine.
- Like a new strand of DNA, the mRNA molecule is synthesized in an antiparallel direction to the template strand of DNA.
- The mRNA base triplets are called **codons**. They are written in the 5′→3′ direction.
    - The term *codon* is also used for the DNA base triplets along the *nontemplate* strand.
    - These codons are complementary to the template strand and thus identical in sequence to the mRNA except that they have T instead of U.
    - For this reason, the nontemplate DNA strand is called the "coding strand."
- During translation, the sequence of codons along an mRNA molecule is translated into a sequence of amino acids making up the polypeptide chain.
    - During translation, the codons are read in the 5′→3′ direction along the mRNA.
    - Each codon specifies which one of the 20 amino acids will be incorporated at the corresponding position along a polypeptide.
- Because codons are nucleotide triplets, the number of nucleotides making up a genetic message must be three times the number of amino acids making up the protein product.
    - It takes at least 300 nucleotides to code for a polypeptide that is 100 amino acids long.
- The task of matching each codon to its amino acid counterpart began in the early 1960s.
- Marshall Nirenberg determined the first match: UUU codes for the amino acid phenylalanine.
    - Nirenberg created an artificial mRNA molecule entirely of uracil and added it to a test-tube mixture of amino acids, ribosomes, and other components for protein synthesis.
    - This "poly-U" translated into a polypeptide containing a single amino acid, phenylalanine, in a long chain.
- AAA, GGG, and CCC were paired with amino acids in the same way.
- More elaborate techniques were required to decode mixed triplets such as AUA and CGA, but by the mid-1960s, the entire code was deciphered.
- Sixty-one of 64 triplets code for amino acids.
    - The codon AUG not only codes for the amino acid methionine but also indicates the "start" or initiation of translation.
    - Three codons do not indicate amino acids but are "stop" signals marking the termination of translation.
- There is redundancy in the genetic code but no ambiguity.

- Several codons may specify the same amino acid (redundancy), but no codon specifies more than one amino acid (no ambiguity).
- The redundancy in the code is not random. In many cases, codons that are synonyms for a particular amino acid differ only in the third nucleotide of the triplet.

- To extract the message from the genetic code requires specifying the correct starting point.
  - The starting point establishes the **reading frame**; subsequent codons are read in groups of three nucleotides.
- The cell's protein-synthesizing machinery reads the message as a series of nonoverlapping three-letter words.

The genetic code must have evolved very early in the history of life.

- The genetic code is nearly universal, shared by organisms from the simplest bacteria to the most complex plants and animals.
- In laboratory experiments, genes can be transcribed and translated after they are transplanted from one species to another.
  - This has permitted bacteria to be programmed to synthesize certain human proteins after insertion of the appropriate human genes.
  - Such applications have produced many exciting developments in biotechnology.
- Exceptions to the universality of the genetic code exist in certain unicellular eukaryotes and in the organelle genes of some species.
- The evolutionary significance of the *near* universality of the genetic code is clear: A language shared by all living things arose very early in the history of life—early enough to be present in the common ancestors of all modern organisms.
- A shared genetic vocabulary is a reminder of the kinship that bonds all life on Earth.

Concept 17.2 Transcription is the DNA-directed synthesis of RNA: *a closer look*

- Messenger RNA, the carrier of information from DNA to the cell's protein-synthesizing machinery, is transcribed from the template strand of a gene.
- **RNA polymerase** separates the DNA strands at the appropriate point and joins RNA nucleotides complementary to the DNA template strand.
  - Like DNA polymerases, RNA polymerases can assemble a polynucleotide only in its 5′→3′ direction.
  - Unlike DNA polymerases, RNA polymerases are able to start a chain from scratch; they don't need a primer.
- Specific sequences of nucleotides along the DNA mark where gene transcription begins and ends.
  - RNA polymerase attaches and initiates transcription at the **promoter**.
  - In bacteria, the sequence that signals the end of transcription is called the **terminator**.
- Molecular biologists refer to the direction of transcription as "downstream" and the other direction as "upstream."
- These terms also describe the positions of nucleotide sequences within the DNA or RNA.

- Thus the promoter sequence in DNA is said to be upstream from the terminator.

- The stretch of DNA that is transcribed into an RNA molecule is called a **transcription unit**.
- Bacteria have a single type of RNA polymerase that synthesizes all RNA molecules.
- In contrast, eukaryotes have three RNA polymerases (I, II, and III) in their nuclei.
  - RNA polymerase II is used for mRNA synthesis.

Transcription can be separated into three stages: initiation, elongation, and termination of the RNA chain.

- The presence of a promoter sequence determines which strand of the DNA helix is the template.
  - Within the promoter is the transcription **start point** for the transcription of a gene.
- RNA polymerase binds in a precise location and orientation on the promoter, determining where transcription starts and which of the two strands of the DNA helix is the template.
- In bacteria, RNA polymerase can recognize and bind directly to the promoter region.
- In eukaryotes, proteins called **transcription factors** mediate the binding of RNA polymerase and the initiation of transcription.
  - Only after transcription factors are attached to the promoter does RNA polymerase II bind to it.
- The complex of transcription factors and RNA polymerase II bound to a promoter is called a **transcription initiation complex**.
  - A crucial promoter DNA sequence is called a **TATA box**.
- Once the appropriate transcription factors are attached to the promoter DNA and the polymerase is bound in the correct orientation, the enzyme unwinds the two DNA strands and starts transcribing the template strand.
- As RNA polymerase moves along the DNA, it untwists the double helix, 10 to 20 nucleotides at a time.
  - The enzyme adds nucleotides to the 3′ end of the growing strand.
- Behind the point of RNA synthesis, the double helix re-forms and the RNA molecule peels away.
  - Transcription progresses at a rate of 40 nucleotides per second in eukaryotes.
- A single gene can be transcribed simultaneously by several RNA polymerases at a time, with a growing strand of RNA trailing off from each polymerase.
  - The length of each new strand reflects how far along the template the enzyme has traveled from the start point.
- Many polymerase molecules simultaneously transcribing a single gene increases the amount of mRNA transcribed from it and helps the cell make the encoded protein in large amounts.
- The mechanism of termination differs between bacteria and eukaryotes.
- In bacteria, transcription proceeds through a terminator sequence in the DNA.
  - The transcribed terminator (an RNA sequence) functions as the termination signal, causing the polymerase to detach from the DNA and release the transcript, which is translated as mRNA without further modification.

- In eukaryotes, RNA polymerase II transcribes a polyadenylation signal sequence, which codes for a polyadenylation signal (AAUAAA) in the pre-mRNA.
    - At a point about 10 to 35 nucleotides past this sequence, the RNA transcript is cut from the polymerase.
    - This releases the pre-mRNA, which then undergoes processing.

Concept 17.3 Eukaryotic cells modify RNA after transcription

- Enzymes in the eukaryotic nucleus modify the pre-mRNA before the genetic messages are dispatched to the cytoplasm.
- During **RNA processing**, both ends of the primary transcript are altered.
    - In most cases, certain interior parts of the molecule are cut out and the remaining parts are spliced together.
    - These modifications help form an mRNA molecule that is ready for translation.
- At the 5′ end of the pre-mRNA molecule, a modified form of guanine is added, the **5′ cap**.
- At the 3′ end, an enzyme adds 50 to 250 adenine nucleotides, the **poly-A tail**.
- These modifications have several important functions.
    - They facilitate the export of mRNA from the nucleus.
    - They help protect mRNA from hydrolytic enzymes.
    - They help the ribosomes attach to the 5′ end of the mRNA.
- The parts of the mRNA that will not be translated into protein are referred to as UTRs (untranslated regions).
- The most remarkable stage of RNA processing occurs during the removal of a large portion of the RNA molecule in a cut-and-paste job of **RNA splicing**.
    - The average length of a transcription unit along a human DNA molecule is about 27,000 nucleotide pairs.
    - However, it takes only 1,200 nucleotides to code for an average-sized protein of 400 amino acids.
- Most eukaryotic genes and their RNA transcripts have long noncoding stretches of nucleotides.
    - Noncoding segments of nucleotides called intervening regions, or **introns**, lie between coding regions.
    - The regions called **exons** are eventually expressed, usually by being translated into amino acid sequences.
- RNA splicing removes introns and joins exons to create an mRNA molecule with a continuous coding sequence.
- The signal for RNA splicing is a short nucleotide sequence at each end of an intron.
    - Particles called *small nuclear ribonucleoproteins (snRNPs)* recognize the splice sites.
- snRNPs are located in the cell nucleus and are composed of RNA and protein molecules.
    - The RNA in an snRNP particle is called a *small nuclear RNA molecule (snRNA).*

- Each RNA molecule is about 150 nucleotides long.

- Several different snRNPs join with a variety of proteins to form a larger assembly called a **spliceosome**, which is about the size of a ribosome.
- The spliceosome interacts with certain sites along an intron, releasing the introns and joining together the two exons that flanked the introns.
    - snRNAs catalyze these processes as well as participating in spliceosome assembly and splice site recognition.

Ribozymes are RNA molecules that function as enzymes.

- The idea of a catalytic role for snRNA arose from the discovery of **ribozymes**, RNA molecules that function as enzymes.
- In some organisms, RNA splicing occurs without proteins or additional RNA molecules: The intron RNA functions as a ribozyme and catalyzes its own excision.
    - For example, in the protozoan *Tetrahymena*, self-splicing occurs in the production of ribosomal RNA (rRNA), a component of the organism's ribosomes.
    - The pre-rRNA removes its own introns.
- The discovery of ribozymes rendered obsolete the idea that all biological catalysts are proteins.
- Three properties of RNA allow some RNA molecules to function as ribozymes.
    1. Because RNA is single-stranded, a region of the RNA molecule may base-pair with a complementary region elsewhere in the same molecule, giving the RNA a specific three-dimensional structure that is key to its ability to catalyze reactions.
    2. Some of the bases in RNA contain functional groups that participate in catalysis.
    3. The ability of RNA to hydrogen-bond with RNA or DNA adds specificity to its catalytic activity.

Introns may play a regulatory role in the cell.

- Specific functions have not been identified for most introns, but some contain sequences that regulate gene expression, and many affect gene products.
- Due to the presence of introns in genes, a single gene can encode more than one kind of polypeptide.
- Many genes give rise to two or more different polypeptides, depending on which segments are treated as exons during RNA processing; this is called **alternative RNA splicing**.
    - Sex differences in fruit flies are due to differences in how males and females splice the RNA transcribed from certain genes.
- Results from the Human Genome Project suggest that alternative splicing explains why humans can get along with a relatively small number of genes.
    - Because of alternative splicing, the number of different protein products an organism can produce is much greater than its number of genes.
- Proteins often have a modular architecture with discrete structural and functional regions called **domains**.
    - One domain of an enzyme may include the active site, while another might allow the protein to bind to a cellular membrane.

- The presence of introns in a gene may facilitate the evolution of new and potentially useful proteins in a process known as *exon shuffling*.
- Introns increase the probability of potentially beneficial crossing over between the exons of alleles of a gene by providing more terrain for crossovers without interrupting coding sequences.
    - This might lead to a protein with a new combination of exons, a novel structure, and a novel function.
- Occasionally, exons may be exchanged between completely different (nonallelic) genes.
- Either way, exon shuffling can lead to new proteins through novel combinations of functions.

Concept 17.4 Translation is the RNA-directed synthesis of a polypeptide: *a closer look*

- In the process of translation, a cell translates a genetic message to build a polypeptide.

The tRNA molecule is a translator.

- The interpreter is **transfer RNA (tRNA)**, which transfers amino acids from the cytoplasmic pool to a growing polypeptide in a ribosome.
    - A cell has all 20 amino acids available in its cytoplasm, either by synthesizing them from scratch or by taking them up from the surrounding solution.
- Each type of tRNA molecule translates a particular mRNA codon into a particular amino acid.
    - Each tRNA molecule bears a specific amino acid at one end.
    - At the other end of the tRNA is a nucleotide triplet called an **anticodon**, which base-pairs with a complementary codon on mRNA.
- Codon by codon, the genetic message is translated as tRNAs deposit amino acids in the order prescribed, and the ribosome joins the amino acids into a chain.
    - tRNA is a translator because it reads a nucleic acid word (the mRNA codon) and interprets it as a protein word (the amino acid).
- Like other types of RNA, tRNA molecules are transcribed from DNA templates.
- In both bacterial and eukaryotic cells, each tRNA is used repeatedly, picking up its designated amino acid in the cytosol, depositing the amino acid at the ribosome, and returning to the cytosol to pick up another copy of that amino acid.
- A tRNA molecule consists of a single strand of about 80 nucleotides folded back on itself to form a three-dimensional structure.
    - tRNA includes a loop containing the anticodon and an attachment site at the 3′ end for an amino acid.
- Each amino acid is joined to the correct tRNA by **aminoacyl-tRNA synthetase**.
    - The 20 different synthetases match the 20 different amino acids.
    - Each has active sites for only a specific tRNA–amino acid combination.
- The synthetase catalyzes a covalent bond between them in a process driven by ATP hydrolysis.

- The result is an aminoacyl-tRNA or charged amino acid.

- The second recognition process involves a correct match between the tRNA anticodon and an mRNA codon.
- If each anticodon had to be a perfect match to each codon, we would expect to find 61 types of tRNA, but the actual number is about 45, because the anticodons of some tRNAs recognize more than one codon.
- Such versatility is possible because the rules for base pairing between the third base of the codon and the anticodon are relaxed. This flexible base pairing is called **wobble**.
  - Wobble explains why the synonymous codons for a given amino acid most often differ in their third base, not in their other bases.

The ribosome is the site of translation.

- Ribosomes facilitate the specific coupling of the tRNA anticodons with mRNA codons during protein synthesis.
  - A ribosome consists of a large and a small subunit, each made up of proteins and **ribosomal RNA (rRNA)**.
- In eukaryotes, the subunits are made in the nucleolus.
  - rRNA genes are transcribed and the RNA is processed and assembled with proteins imported from the cytoplasm.
  - The subunits are exported via nuclear pores to the cytoplasm.
- In both bacteria and eukaryotic cells, large and small subunits join to form a functional ribosome only when they attach to an mRNA molecule.
  - Because most cells contain thousands of ribosomes, rRNA is the most abundant type of cellular RNA.
- Though similar in structure and function, bacterial and eukaryotic ribosomes have enough differences that certain antibiotic drugs (like tetracycline and streptomycin) can inactivate bacterial ribosomes without inhibiting eukaryotic ribosomes.
- Each ribosome has a binding site for mRNA and three binding sites for tRNA molecules.
  - The **P site** (**p**eptidyl tRNA-binding site) holds the tRNA carrying the growing polypeptide chain.
  - The **A site** (**a**minoacyl tRNA-binding site) holds the tRNA carrying the next amino acid to be added to the chain.
  - Discharged tRNAs leave the ribosome at the **E** (**e**xit) **site.**
- The ribosome holds the tRNA and mRNA in close proximity and positions the new amino acid for addition to the carboxyl end of the growing polypeptide.
  - It then catalyzes the formation of the peptide bond.
- As the polypeptide becomes longer, it passes through an *exit tunnel* in the ribosome's large unit and is released to the cytosol through the exit tunnel.
- Evidence supports the hypothesis that rRNA, not protein, is responsible for the ribosome's structure and function.
  - Proteins on the exterior support the shape changes of the rRNA molecules as they carry out catalysis during translation.

- rRNA is the main constituent at the interface between the two subunits and of the A and P sites, and it is the catalyst for peptide bond formation.
- A ribosome can be regarded as one colossal ribozyme.

The process of translation builds a polypeptide.

- Translation can be divided into three stages: initiation, elongation, and termination.
  - All three phases require protein "factors" that aid in the translation process.
  - Both initiation and chain elongation require energy provided by the hydrolysis of GTP.
- **Initiation** brings together mRNA, a tRNA with the first amino acid, and the two ribosomal subunits.
- First, a small ribosomal subunit binds with mRNA and a special initiator tRNA, which carries methionine.
  - In bacteria, the binding occurs at a specific RNA sequence, just upstream of the start codon, AUG.
  - In eukaryotes, the small subunit, with the initiator tRNA already bound, binds to the 5′ cap of the mRNA and then moves, or *scans*, downstream along the mRNA until it reaches the start codon AUG, which signals the start of translation.
  - This establishes the codon reading frame for the mRNA.
- The union of mRNA, initiator tRNA, and a small ribosomal subunit is followed by the attachment of a large ribosomal subunit, forming the *translation initiation complex*.
  - Proteins called *initiation factors* bring all these components together.
  - Energy in the form of a GTP molecule is invested in the formation of the initiation complex.
- In the elongation stage of translation, amino acids are added one by one to the previous amino acid at the C-terminus of the growing chain.
  - Each addition involves the participation of protein *elongation factors* and occurs in three-step cycles as each amino acid is added to the preceding one.
- Energy expenditure occurs in the first and third steps.
  - Codon recognition requires hydrolysis of one molecule of GTP, which increases the accuracy and efficiency of this step.
  - One more GTP is hydrolyzed to provide energy for the translocation step.
- The mRNA is moved through the ribosome in one direction only, 5′ end first; this is equivalent to the ribosome moving 5′ $\rightarrow$ 3′ on the mRNA.
  - The ribosome and the mRNA move relative to each other, unidirectionally, codon by codon.
- The elongation cycle takes less than a tenth of a second in bacteria and is repeated as each amino acid is added to the chain until the polypeptide is completed.
- **Termination** occurs when one of the three stop codons reaches the A site of the ribosome.
  - A *release factor* binds to the stop codon and causes hydrolysis of the bond between the polypeptide and its tRNA in the P site.

- This frees the polypeptide, which is released through the exit tunnel of the ribosome's large subunit.

- The translation complex disassembles.
  - Breakdown of the translation assembly requires the hydrolysis of two more GTP molecules.
- A ribosome requires less than a minute to translate an average-sized mRNA into a polypeptide.
- A single mRNA may be used to make many copies of a polypeptide simultaneously as multiple ribosomes, **polyribosomes** or **polysomes**, trail along the same mRNA.
  - Polyribosomes can be found in bacterial and eukaryotic cells.

Folding and modification of a protein follows translation.

- During and after synthesis, a polypeptide spontaneously coils and folds to its three-dimensional shape.
  - The primary structure, the order of amino acids, determines the secondary and tertiary structure.
- A chaperone protein (chaperonin) helps the polypeptide fold correctly.
- In addition, proteins may require *post-translational modifications* before doing their particular job.
- These modifications may require additions such as sugars, lipids, or phosphate groups to amino acids.
  - Enzymes may remove one or more amino acids from the leading (amino) end of the polypeptide chain.
- In some cases, a single polypeptide chain may be enzymatically cleaved into two or more pieces.
  - In other cases, two or more polypeptides may join to form a protein with quaternary structure.

Signal peptides target some eukaryotic polypeptides to specific destinations in the cell.

- Two populations of ribosomes, free and bound, are active participants in protein synthesis.
  - Free ribosomes are suspended in the cytosol and synthesize proteins that reside and function in the cytosol.
  - Bound ribosomes are attached to the cytosolic side of the endoplasmic reticulum (ER) or to the nuclear envelope.
  - Bound ribosomes make proteins of the endomembrane system as well as proteins secreted from the cell.
- Although bound and free ribosomes are identical in structure, their location depends on the type of protein they are synthesizing.
- Translation in all ribosomes begins in the cytosol, but a polypeptide destined for the endomembrane system or for export has a specific **signal peptide** region at or near the leading end.
  - The signal peptide consists of a sequence of about 20 amino acids.

- A **signal recognition particle (SRP)** binds to the signal peptide and attaches it and its ribosome to a receptor protein in the ER membrane.
    - The SRP, which consists of a protein–RNA complex, functions as an escort to bring the ribosome to a receptor protein built into the ER membrane.
    - The receptor is part of a multiprotein translocation complex.
- Protein synthesis resumes, with the growing polypeptide snaking across the membrane into the ER lumen via a protein pore.
    - An enzyme usually cleaves the signal polypeptide.
- Secretory proteins are released into solution within the ER lumen, but membrane proteins remain partially embedded in the ER membrane.
- Other kinds of signal peptides are used to target polypeptides to mitochondria, chloroplasts, the nucleus, and other organelles that are not part of the endomembrane system.
    - In these cases, translation is completed in the cytosol before the polypeptide is imported into the organelle.
    - While the mechanisms of translocation vary, each of these polypeptides has a "zip code" that ensures its delivery to the correct cellular location.
- Bacteria also employ signal sequences to target proteins to the plasma membrane for secretion.

Concept 17.5 Mutations of one or a few nucleotides can affect protein structure and function

- **Mutations** are changes in the genetic material of a cell (or virus).
- Mutations are the ultimate source of new genes.
- Mutations include large-scale mutations, in which long segments of DNA are affected (for example, translocations, duplications, and inversions), as well as **point mutations**, chemical changes in just one nucleotide pair of a gene.
- If a point mutation occurs in a gamete or in a cell that produces gametes, it may be transmitted to future generations.
- If the mutation has an adverse effect on the phenotype of an organism, the mutant condition is referred to as a genetic disorder or hereditary disease.
    - For example, sickle-cell disease is caused by a mutation of a single nucleotide pair in the gene that encodes the β-globin polypeptide of hemoglobin.
    - A change in a single nucleotide in the DNA's template strand leads to an abnormal protein.

A nucleotide-pair substitution is the replacement of one nucleotide and its partner with another pair of nucleotides.

- Some **nucleotide-pair substitutions** have no effect on protein function, due to the redundancy of the genetic code.
    - In a **silent mutation**, a change in a nucleotide pair transforms one codon into another that is translated into the same amino acid.
- **Missense mutations** change one amino acid for another with little effect on protein function.

- In some cases, the mutation switches one amino acid for another with similar properties.
- Other mutations occur in a region where the exact amino acid sequence is not essential for function.

- Other nucleotide-pair substitutions cause a major change in a protein.
  - Changes in amino acids at crucial sites, especially active sites, are likely to affect function.
  - These substitutions are usually detrimental but sometimes lead to an improved protein or one with novel capabilities.
- **Nonsense mutations** change an amino acid codon into a stop codon, causing premature termination of translation and nearly always leading to a nonfunctional protein.

Insertions and deletions are additions or losses of nucleotide pairs in a gene.

- **Insertions** and **deletions** are more likely than substitutions to have a disastrous effect on the resulting protein.
- Unless insertion or deletion mutations occur in multiples of 3, they cause a **frameshift mutation**.
  - All the nucleotides downstream of the deletion or insertion will be improperly grouped into codons.
  - The result will be extensive missense, ending sooner or later in nonsense—premature termination.

Mutations can occur during DNA replication, DNA repair, or DNA recombination.

- Errors during DNA replication or recombination can lead to nucleotide-pair substitutions, insertions, or deletions.
  - In many cases, the error will be corrected.
  - In other cases, the incorrect base will be used as a template in the next round of replication. Such mutations are called *spontaneous mutations*.
- Rough estimates have been made of the rate of mutation during DNA replication for both *Escherichia coli* and eukaryotes, and the numbers are similar: About one nucleotide in every 10^{10} is altered and passed on to the next generation of cells.
- **Mutagens** are chemical or physical agents that interact with DNA to cause mutations.
- Physical agents include high-energy radiation like X-rays and ultraviolet light.
  - Ultraviolet light can cause disruptive thymine dimers in DNA.
- Chemical mutagens cause mutations in different ways.
  - Some chemicals are nucleotide analogs that may be substituted into DNA, but pair incorrectly during DNA replication.
  - Other mutagens interfere with DNA replication by inserting into DNA and distorting the double helix.
  - Still others cause chemical changes in bases that change their pairing properties.
- Researchers have developed various methods to test the mutagenic activity of different chemicals.

- These tests are often used as a preliminary screen of chemicals to identify those that may cause cancer.
- This makes sense because most carcinogens are mutagenic and most mutagens are carcinogenic.

Concept 17.6 While gene expression differs among the domains of life, the concept of a gene is universal

- Although bacteria and eukaryotes carry out transcription and translation in similar ways, they differ in cellular machinery and in the details of the processes.
- Archaea share many aspects of the mechanisms of gene expression with eukaryotes, as well as a few with bacteria.
- Bacterial and eukaryotic RNA polymerases differ significantly from each other, while the single archaeal RNA polymerase resembles the three eukaryotic ones.
- Archaea and eukaryotes use a complex set of transcription factors, unlike the smaller set of accessory proteins in bacteria.
- Transcription is terminated differently in bacteria and eukaryotes.
  - Little is known about this process in archaea, although it appears to be more similar to the eukaryotic process.
- Bacterial and eukaryotic ribosomes differ slightly.
  - Archaeal ribosomes are the same size as bacterial ribosomes, but their sensitivity to chemical inhibitors is similar to that of eukaryotic ribosomes.
- Initiation of translation is slightly different in bacteria and eukaryotes.
  - The archaeal process is similar to the bacterial process.
- Gene expression in eukaryotes differs from that of bacteria because of the greater compartmental organization of the eukaryotic cell.
  - In the absence of a nucleus, a bacterial cell can simultaneously transcribe and translate the same gene and the new protein quickly diffuses to its operating site.
- Since archaea lack a nuclear envelope, transcription and translation are likely coupled.
- In eukaryotes, the nuclear envelope segregates transcription from translation and provides a compartment for extensive RNA processing between these processes.
  - This provides additional steps whose regulation helps coordinate the elaborate activities of a eukaryotic cell.

What is a gene? We revisit the question.

- In spite of the differences in gene expression between living things, the gene itself is a unifying concept among all forms of life.
- The Mendelian concept of a gene views it as a discrete unit of inheritance that affects phenotype.
- Morgan and his colleagues assigned genes to specific loci on chromosomes.
- We can also view a gene as a specific nucleotide sequence along a region of a DNA molecule.

- We can define a gene functionally as a DNA sequence that codes for a specific polypeptide chain.
- All these definitions are useful in certain contexts.
- Most eukaryotic genes contain large introns that have no corresponding segments in polypeptides.
    - Promoters and other regulatory regions of DNA are not transcribed, but they must be present for transcription to occur.
    - Our molecular definition must also include the various types of RNA that are not translated into polypeptides, such as rRNA, tRNA, and other RNAs.
- This is our definition of a gene: *A gene is a region of DNA that can be expressed to produce a final functional product that is either a polypeptide or an RNA molecule*.
- A typical gene is expressed by transcription into RNA and then translation into a polypeptide that forms a protein of specific structure and function.
    - Proteins, in turn, bring about an organism's observable phenotype.
- A given type of cell expresses only a subset of its genes, and gene expression is precisely regulated.