7.1 From Gene to Protein

As you learned in Section 6.1, Gregor Mendel experimented with garden peas and laid the foundation for the science of genetics. His experimental results led to his hypothesis that certain factors are responsible for the patterns of inheritance that he observed in his pea plants. He proposed that the height of the pea plant, the colour of the flower, and the shape and colour of the seed are determined by these factors, which are passed from parent to offspring. These factors are now known as genes. They direct the production of proteins that determine the physical characteristics of organisms. They also direct the production of other essential proteins, such as antibodies, enzymes, and hormones. Proteins drive cellular processes (such as metabolism), determine physical characteristics, and manifest genetic disorders by their absence or by their presence in an altered form.

One Gene–One Polypeptide Hypothesis

Two key pieces of research, involving defects in metabolism, led scientists to the discovery of how genes encode for proteins (or, in other words, specify the amino acid sequence). The first piece of research began in 1896 with Archibald Garrod, an English physician. He studied alkaptonuria, a human disease that is detected easily because a patient's urine turns black in air. Garrod studied families of these patients and concluded that alkaptonuria is an inherited trait. He found that people with alkaptonuria excrete a particular chemical in their urine. Garrod concluded that people without the disease are able to break down the chemical, whereas people with alkaptonuria cannot. By 1908, Garrod had concluded that the disease was an "inborn error of metabolism." Later research has shown that alkaptonuria is caused by a mutation of a gene that normally codes for an enzyme that breaks down the amino acid tyrosine. The altered gene codes for a defective enzyme that cannot fully break down tyrosine. When tyrosine is not broken down fully, a chemical known as alkapton accumulates. Alkapton turns black in air. Garrod's work was the first to show a specific relationship between genes and metabolism.

The second piece of research that led scientists to the gene–protein connection was conducted by George Beadle and Edward Tatum in the 1940s. Working with the orange bread mould *Neurospora crassa*, they collected data that showed a direct relationship between genes and enzymes. Normal *Neurospora* grows readily on a minimal medium (MM): a medium that contains several salts, sucrose, and a vitamin, but none of the other, more complex chemicals required by cells. The researchers reasoned that the mould uses the simple chemicals in the medium to synthesize all of the more complex molecules it needs for growth and reproduction. Beadle and Tatum exposed spores of *Neurospora* to X-rays. X-rays are mutagenic and cause random mutations in genes. They observed that some of the X-rayed spores did not grow unless additional nutrients, such as amino acids and vitamins, were added to the MM. Because of the mutation(s) caused by the radiation, the spores were unable to synthesize all of the complex molecules they needed to grow. Only when additional nutrients were added did the spores mature into mould.

Beadle and Tatum hypothesized that each mutated strain had a defect in a gene that coded for one of the enzymes needed to synthesize a particular nutrient that was not in the MM. The original strain could make the nutrient for itself from raw materials in the MM, but the mutated strain could not. It could grow only if the researchers added the nutrient to the MM. For example, the mutated strain that required the addition of the amino acid arginine had a defect in a gene that coded for an enzyme required in the synthesis of arginine. The assembly of arginine from raw materials is a multistep process, with different enzymes responsible for each step. Therefore, different "*arg*" mutants might differ in the particular enzyme that is defective and therefore in which step of the synthesis of arginine is blocked.

Beadle and Tatum deduced that the biosynthesis of arginine required a number of steps, and each step was controlled by a gene that coded for the enzyme for this step (**Figure 1**). Beadle and Tatum had shown the direct relationship between genes and enzymes, which they put forward as the **one gene–one enzyme hypothesis**. Their work was a keystone in the development of molecular biology. As a result of their work, they were awarded a Nobel Prize in 1958. Later, scientists recognized that the one gene–one enzyme hypothesis needed to be expanded to include other assemblies of amino acids. Many proteins are not enzymes, and many proteins consist of more than one subunit. Since this subunit is called a polypeptide, Beadle and Tatum's hypothesis was restated as the **one gene–one polypeptide hypothesis**.

one gene–one enzyme hypothesis the hypothesis, proposed by Beadle and Tatum, that each gene is unique and codes for the synthesis of a single enzyme

one gene–one polypeptide hypothesis the hypothesis that each gene is unique and codes for the synthesis of a single polypeptide; the restated version of the one gene–one enzyme hypothesis

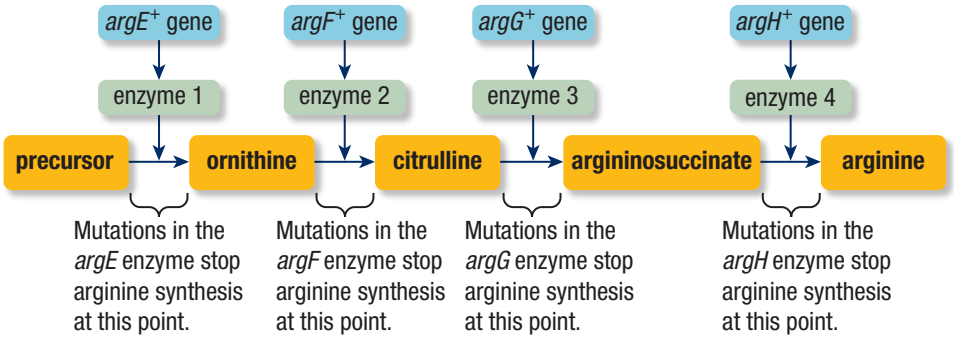

Figure 1 Arginine is synthesized in a biochemical pathway. Each step of the pathway is catalyzed by an enzyme, and each enzyme is coded by a different gene.

Connection between DNA, RNA, and Protein

In 1956, Francis Crick gave the name **central dogma** to the flow of information from DNA to RNA to protein (**Figure 2**). The process has two major steps: transcription and translation.

central dogma the fundamental principle of molecular genetics, which states that genetic information flows from DNA to RNA to proteins

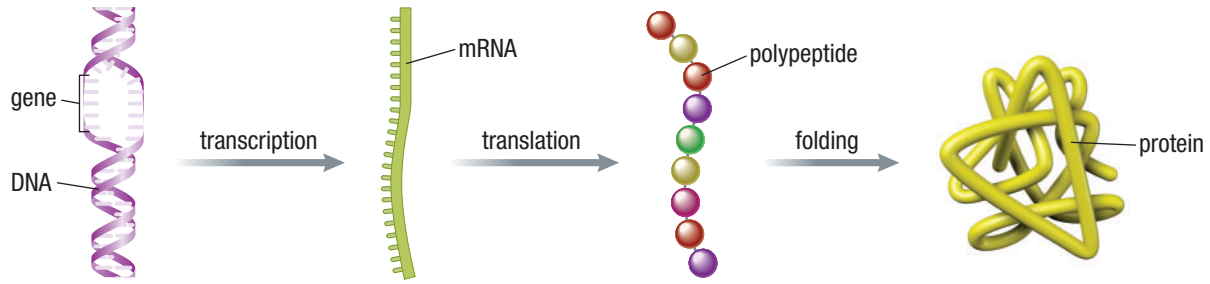

Figure 2 The central dogma of molecular genetics states that genetic information flows from DNA to RNA to protein.

Transcription is the mechanism by which the information encoded in DNA is transcribed into a complementary RNA copy. In other words, the information in one type of nucleic acid, DNA, is copied onto another type of nucleic acid, RNA. Transcription occurs in the nucleus of a eukaryotic cell. Unlike DNA, RNA is able to exit the nucleus and enter the cytosol.

transcription mechanism by which the information coded in nucleic acids of DNA is copied into the nucleic acids of RNA; something rewritten in the same language

Translation is the assembly of amino acids into a polypeptide using the information encoded in the RNA. It takes place on the ribosomes in the cytosol. The term "translation" (changing from one language to another) helps us visualize the flow of genetic information. RNA contains the information for a polypeptide in the language of bases, but this information must be translated into the language of amino acids. Transcription and translation will be discussed in more detail in Sections 7.2 and 7.3, respectively.

translation mechanism by which the information coded in the nucleic acids of RNA is copied into the amino acids of proteins

RNA: Ribonucleic Acid

Like DNA, ribonucleic acid (RNA) is a carrier of genetic information. However, RNA differs from DNA in many ways (**Table 1** and **Figure 3**). First, ribonucleic acid contains a ribose sugar rather than a deoxyribose sugar. A ribose sugar has a hydroxyl group on its 2′ carbon. Second, instead of thymine, RNA contains the base uracil. Uracil is similar in structure to thymine, except thymine has a methyl group on its 1′ carbon. Uracil in the RNA pairs with adenine in the DNA strand. Third, DNA is double stranded, whereas RNA is single stranded. When a gene is transcribed into RNA, only a single-stranded complementary copy is made. In the complementary copy, uracil is substituted for thymine.

Table 1 Comparison of DNA and RNA

Deoxyribonucleic acid	Ribonucleic acid
• double stranded	• single stranded
• adenine pairs with thymine	• adenine pairs with uracil
• guanine pairs with cytosine	• guanine pairs with cytosine
• deoxyribose sugar	• ribose sugar

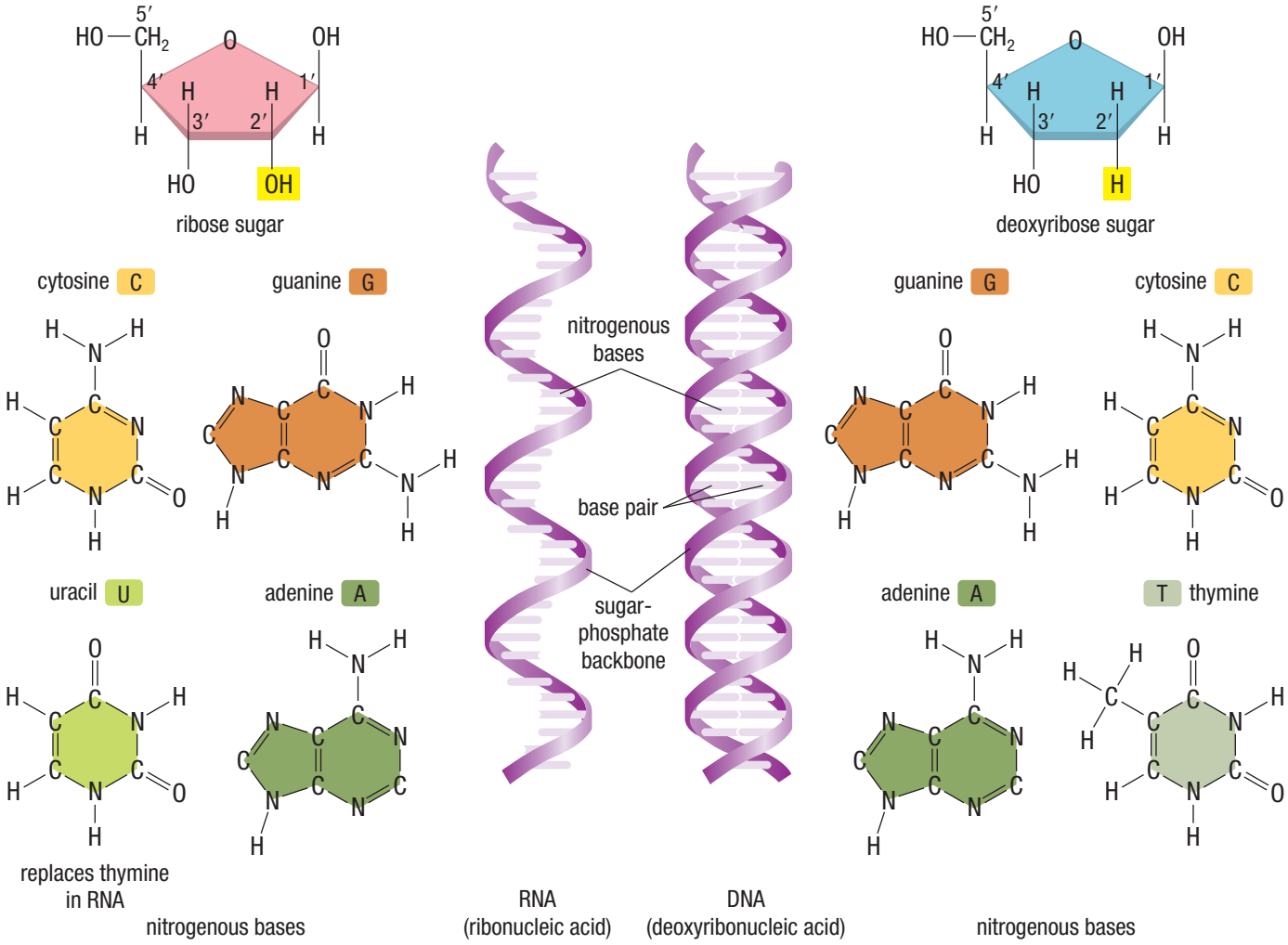

Figure 3 DNA and RNA are similar, with a few important differences: RNA is single stranded, not double stranded; RNA has uracil in place of thymine in DNA; and RNA contains a hydroxyl group (OH) at the 2′ position of the sugar, whereas DNA has a hydrogen (H) at that position (thus "deoxy" to denote the absence of oxygen).

Three major types of RNA molecules are involved in protein synthesis: messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA) (**Table 2**). **Messenger RNA (mRNA)** acts as the intermediary between DNA and the ribosomes. mRNA is the RNA version of the gene encoded by DNA. It varies in length, depending on the gene that has been transcribed; the longer the gene, the longer the mRNA is.

messenger RNA (mRNA) the end product of the transcription of a gene; mRNA is translated by ribosomes into a protein

Table 2 Different Types of RNA

Types of RNA	Characteristics and key functions
messenger RNA (mRNA)	• varies in length, depending on the gene that has been copied • acts as the intermediary between DNA and the ribosomes • is translated into protein by ribosomes • is the RNA version of the gene encoded by DNA
transfer RNA (tRNA)	• functions as the delivery system of amino acids to ribosomes as they synthesize proteins • is very short, only 70 to 90 base pairs long
ribosomal RNA (rRNA)	• binds with proteins to form the ribosomes • varies in length

The role of **transfer RNA (tRNA)** is to transfer the appropriate amino acid to the ribosome to build a protein, as dictated by the mRNA template. tRNA is comparatively short in length, averaging 70 to 90 ribonucleotides (**Figure 4**). The single-stranded RNA molecule loops in on itself, forming antiparallel double strands, which are complementary to each other. **Ribosomal RNA (rRNA)** is a structural component of a ribosome and varies in length. Along with proteins, it forms the ribosome, which is the construction site for the assembly of polypeptides.

transfer RNA (tRNA) a carrier molecule that binds to a specific amino acid and adds the amino acid to the growing polypeptide chain

ribosomal RNA (rRNA) an RNA molecule within the ribosome that bonds the correct amino acid to the polypeptide chain

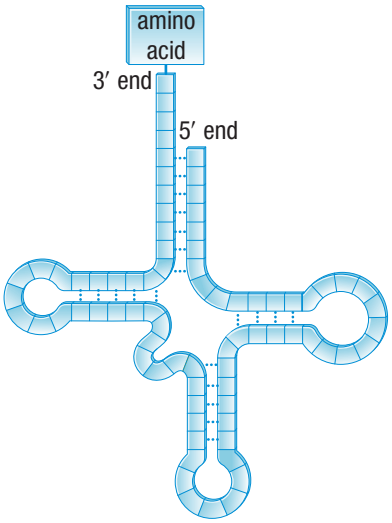

Figure 4 tRNA is composed of a single strand of RNA that loops in on itself to form antiparallel double-stranded areas.

Transcription and Translation: An Overview

The first step of protein synthesis is transcription. In transcription, the enzyme **RNA polymerase** creates an RNA molecule with a base sequence that is complementary to one strand of the DNA sequence of a given gene (**Figure 5**, next page). Transcription follows the same basic rules of complementary base pairing and nucleic acid chemistry, which you first encountered in DNA replication. For example, because the DNA template strand is read in the 3′ to 5′ direction, the mRNA will be formed in the complementary 5′ to 3′ direction. For each of the several thousand genes that can be expressed in a given cell, one DNA strand is the **template strand** and

RNA polymerase an enzyme that reads a DNA strand and creates a complementary strand of RNA

template strand the DNA strand that is copied into an mRNA molecule during gene transcription

precursor mRNA (pre-mRNA) the initial RNA transcription product

is read by the RNA polymerase. The template strand is a DNA strand that is transcribed into a precursor mRNA molecule. This **precursor mRNA** (or **pre-mRNA**), which cannot be used to produce a protein, is then modified to become an mRNA strand (to be discussed in Section 7.2). The mRNA can now exit the nucleus and enter the cytosol, where ribosomes are found.

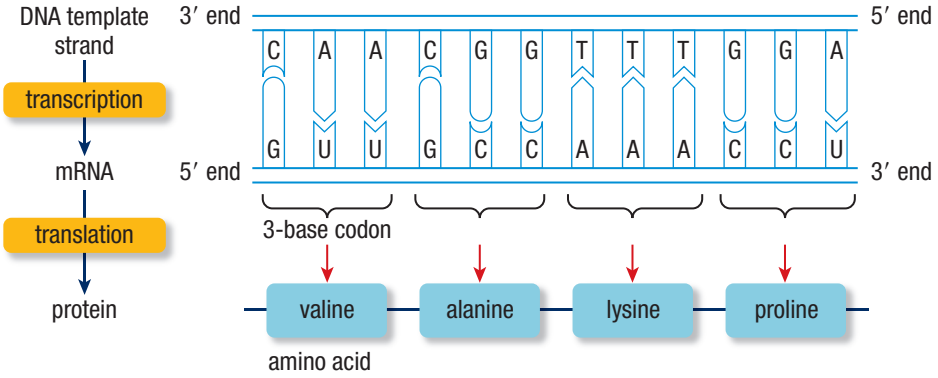

Figure 5 The relationship between a gene, the codons in an mRNA, and the amino acid sequence of a polypeptide

Translation is the second step of protein synthesis. In translation, the mRNA molecule associates with a ribosome (**Figure 6**). As the ribosome moves along the mRNA, the amino acids coded for by the mRNA are delivered by tRNA to the ribosome. The amino acids are joined together, one by one, to form the polypeptide encoded by the gene.

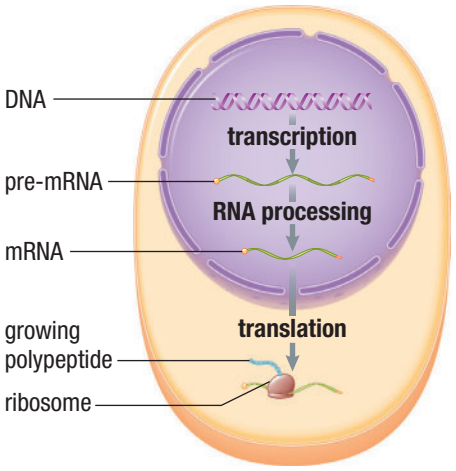

Figure 6 During transcription, RNA polymerase synthesizes a precursor-mRNA molecule (pre-mRNA, discussed in Section 7.2), which contains extra segments that are later removed. The final mRNA exits the nucleus through a nuclear pore and is translated by ribosomes in the cytosol, resulting in the synthesis of a polypeptide.

The Genetic Code: Three-Letter Words with a Four-Letter Alphabet

genetic code the specific coding relationship between bases and the amino acids they specify; the genetic code can be expressed in terms of either DNA or RNA bases

The specific amino acid coded for by particular DNA (or complementary RNA) bases is determined by the **genetic code**. The DNA "alphabet" consists of the four letters A, T, G, and C, and the RNA "alphabet" consists of the four letters A, U, G, and C. However, while there are only four RNA bases, there are 20 amino acids. How is nucleotide information in an mRNA molecule translated into the amino acid sequence of a polypeptide? Scientists realized that the four bases in an mRNA must be used in combinations of at least three to provide the capacity to code for 20 amino acids. If the code used only single letters (bases), only 4 different amino acids could be specified (A, T, C, and G, or 4^1). If the code used 2-letter combinations, only 16 different amino acids could be specified (AA, AT, AC, AG, TA, TT, . . . , or 4^2).

But if the code used 3-letter combinations, 64 different amino acids could be specified (AAA, AAT, AAC, . . . , or 4^3)—more than enough to code for 20 amino acids.

Scientists now know that the genetic code is indeed a 3-letter code. Each 3-letter combination is called a **codon**. The codons are in the 5′ to 3′ order in the mRNA. Figure 5 (page 316) illustrates the relationship between a gene, codons on an mRNA, and the amino acid sequence of a polypeptide. Of the 64 codons, 61 specify amino acids (**Figure 7** and **Table 3**). These are known as "sense codons." For example, one of these codons, AUG, specifies the amino acid methionine. It is usually the first codon translated in any mRNA in both prokaryotes and eukaryotes. Therefore, AUG is called a **start codon**, or **initiator codon**.

codon a group of three base pairs that code for an individual amino acid

start codon (initiator codon) the codon that signals the start of a polypeptide chain and initiates translation

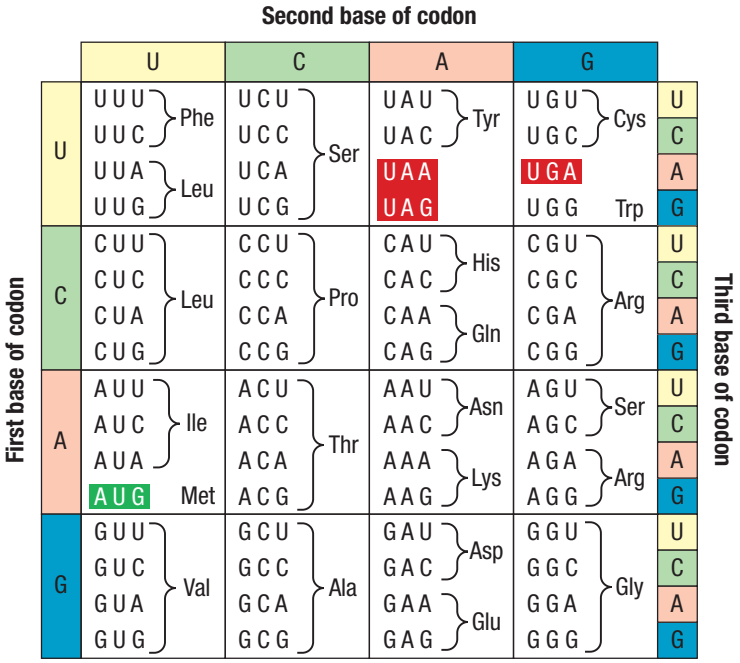

Figure 7 The genetic code, written in the form in which the codons appear in mRNA: The AUG initiator codon, which codes for methionine, is shown in green; the three terminator codons are shown in red. The triplet sequences are in the 5′ to 3′ order.

Table 3 Amino Acids and Their Abbreviations

Amino acid	Three-letter abbreviation	Amino acid	Three-letter abbreviation
alanine	Ala	leucine	Leu
arginine	Arg	lysine	Lys
asparagine	Asn	methionine	Met
aspartic acid	Asp	phenylalanine	Phe
cysteine	Cys	proline	Pro
glutamic acid	Glu	serine	Ser
glutamine	Gln	threonine	Thr
glycine	Gly	tryptophan	Trp
histidine	His	tyrosine	Tyr
isoleucine	Ile	valine	Val

The three codons that do not specify amino acids—UAA, UAG, and UGA—are called **stop codons** (also called "nonsense codons" or "termination codons"). They act as "periods," indicating the end of a polypeptide-encoding sentence. When a ribosome reaches one of the stop codons, polypeptide synthesis stops and the newly synthesized polypeptide chain is released from the ribosome.

stop codon a codon that signals the end of a polypeptide chain and causes the ribosome to terminate translation

Only two amino acids, methionine and tryptophan, are specified by a single codon. All the rest are represented by at least two, and some by as many as six, codons. In other words, there are many synonyms in the genetic code. For example, UGU and UGC both specify cysteine, whereas CCU, CCC, CCA, and CCG all specify proline. This feature is known as redundancy and is called the wobble hypothesis. The presence of this redundancy allows the third base in a codon to change (wobble), while still allowing the codon to code for the same amino acid. Notice how both cysteine codons follow the pattern UG_ and all proline codons follow the pattern CC_. The wobble hypothesis is discussed in more detail in Section 7.3.

The genetic code is universal. With a few exceptions, the same codons specify the same amino acids in all living organisms, and also in all viruses. The universality of the genetic code indicates that it was established, in its present form, very early in the evolution of life and has remained virtually unchanged through billions of years of evolutionary history. Minor exceptions to the universality of the genetic code have been found in a few organisms, including yeast, some protozoans, and a prokaryote, and in the DNA of mitochondria and chloroplasts.

7.1 Review

Summary

- Beadle and Tatum's experiments exposing bread mould to X-ray radiation resulted in their one gene–one enzyme hypothesis, later expanded to the one gene–one polypeptide hypothesis.
- A gene is a DNA sequence of bases that specifies the exact sequence of assembly of amino acids, by a ribosome, into a unique polypeptide.
- The central dogma outlines the flow of information from DNA to mRNA to protein.
- DNA differs from RNA. RNA is a single strand, it contains a ribose sugar instead of a deoxyribose sugar, and it contains uracil instead of thymine.
- Three types of RNA are involved in protein synthesis: messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA).
- In transcription, the information encoded in DNA is passed to a complementary mRNA molecule. The code is rewritten using the same nucleotides that are found in DNA, except for uracil, which replaces the thymine in DNA.
- In translation, the sequence of nucleotides in an mRNA molecule specifies an amino acid sequence in a polypeptide. A ribosome uses the mRNA sequence to assemble a polypeptide, with the help of tRNA.
- The genetic information that specifies a single amino acid is a sequence of three bases, called a codon. A single codon (AUG) signals the start of translation, and three codons (UAA, UAG, and UGA) signal the termination.

Questions

1. Describe Beadle and Tatum's experiment. How did their results support their one gene–one enzyme hypothesis? K/U
2. Beadle and Tatum observed that different mutant strains of *Neurospora* required the addition of different nutrients to the minimal medium (MM) in order to grow. Explain this observation. K/U
3. Using a Venn diagram, compare the structure of DNA and RNA. K/U C
4. What are the three major classes of RNA, and what is the function of each? K/U
5. Compare and contrast transcription and translation in terms of their purpose and location. K/U
6. The sequence of a fragment of one strand of DNA is AATTGCATATACGGGAAATACGACCGG. Transcribe this sequence into mRNA. K/U T/I
7. A genetic code that is based on triplet codons can generate 64 unique combinations. How many combinations of codons could be generated using a four-base codon? How many could be generated using a five-base codon? T/I
8. The following mRNA strand is being used to assemble a polypeptide strand by a ribosome:
   5′-AUGCUUGCUCAUCGGGGUUUUAA-3′
   (a) Write out the amino acids that will be assembled, in their correct order.
   (b) Provide an alternative mRNA sequence with four or more changes that would translate to the same amino acid sequence. T/I
9. Write out all possible RNA base codons that could code for the following amino acid sequences. T/I
   (a) Arg-Trp
   (b) Leu-Pro
   (c) Met-Phe-Trp
10. Differentiate between a stop codon and a start codon. K/U
11. What is the wobble hypothesis? K/U
12. Suppose that you are studying a protein with the primary sequence Met-Leu-His-Asn-Ala. Write out one double-stranded DNA sequence for this peptide. (Yes, there are MANY possible answers!) Include the start codon and a stop codon. Indicate the 3′ and 5′ ends of the DNA. Which is the template strand? T/I