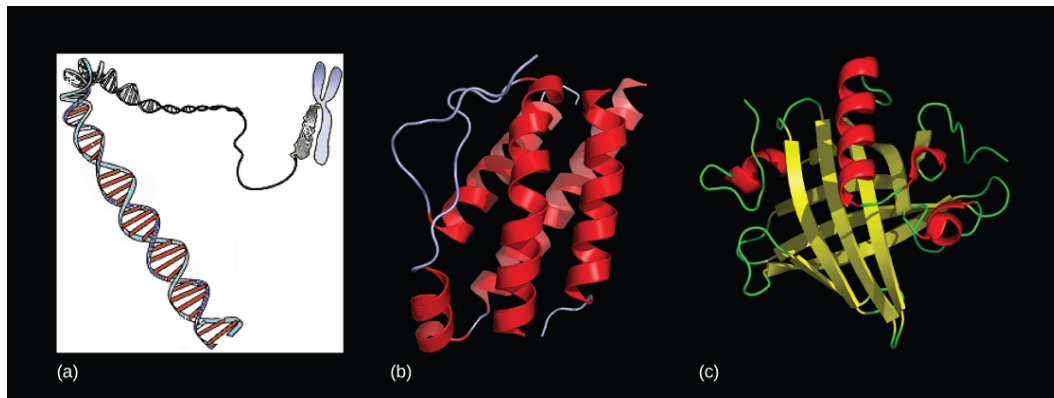# 15 | GENES AND PROTEINS

Figure 15.1 Genes, which are carried on (a) chromosomes, are linearly organized instructions for making the RNA and protein molecules that are necessary for all of the processes of life. The (b) interleukin-2 protein and (c) alpha-2u-globulin protein are just two examples of the array of different molecular structures that are encoded by genes. (credit "chromosome: National Human Genome Research Institute; credit "interleukin-2": Ramin Herati/Created from PDB 1M47 and rendered with Pymol; credit "alpha-2u-globulin": Darren Logan/rendered with AISMIG)

Chapter Outline

15.1: The Genetic Code

15.2: Prokaryotic Transcription

15.3: Eukaryotic Transcription

15.4: RNA Processing in Eukaryotes

15.5: Ribosomes and Protein Synthesis

Introduction

Since the rediscovery of Mendel's work in 1900, the definition of the gene has progressed from an abstract unit of heredity to a tangible molecular entity capable of replication, expression, and mutation (**Figure 15.1**). Genes are composed of DNA and are linearly arranged on chromosomes. Genes specify the sequences of amino acids, which are the building blocks of proteins. In turn, proteins are responsible for orchestrating nearly every function of the cell. Both genes and the proteins they encode are absolutely essential to life as we know it.

15.1 | The Genetic Code

By the end of this section, you will be able to do the following:

- Explain the "central dogma" of DNA-protein synthesis
- Describe the genetic code and how the nucleotide sequence prescribes the amino acid and the protein sequence

The cellular process of transcription generates messenger RNA (mRNA), a mobile molecular copy of one or more genes with an alphabet of A, C, G, and uracil (U). Translation of the mRNA template on ribosomes converts nucleotide-based genetic information into a protein product. That is the central dogma of DNA-protein synthesis.

Protein sequences consist of 20 commonly occurring amino acids; therefore, it can be said that the protein alphabet consists of 20 "letters" (**Figure 15.2**). Different amino acids have different chemistries (such as acidic versus basic, or polar and nonpolar) and different structural constraints. Variation in amino acid sequence is responsible for the enormous variation in protein structure and function.
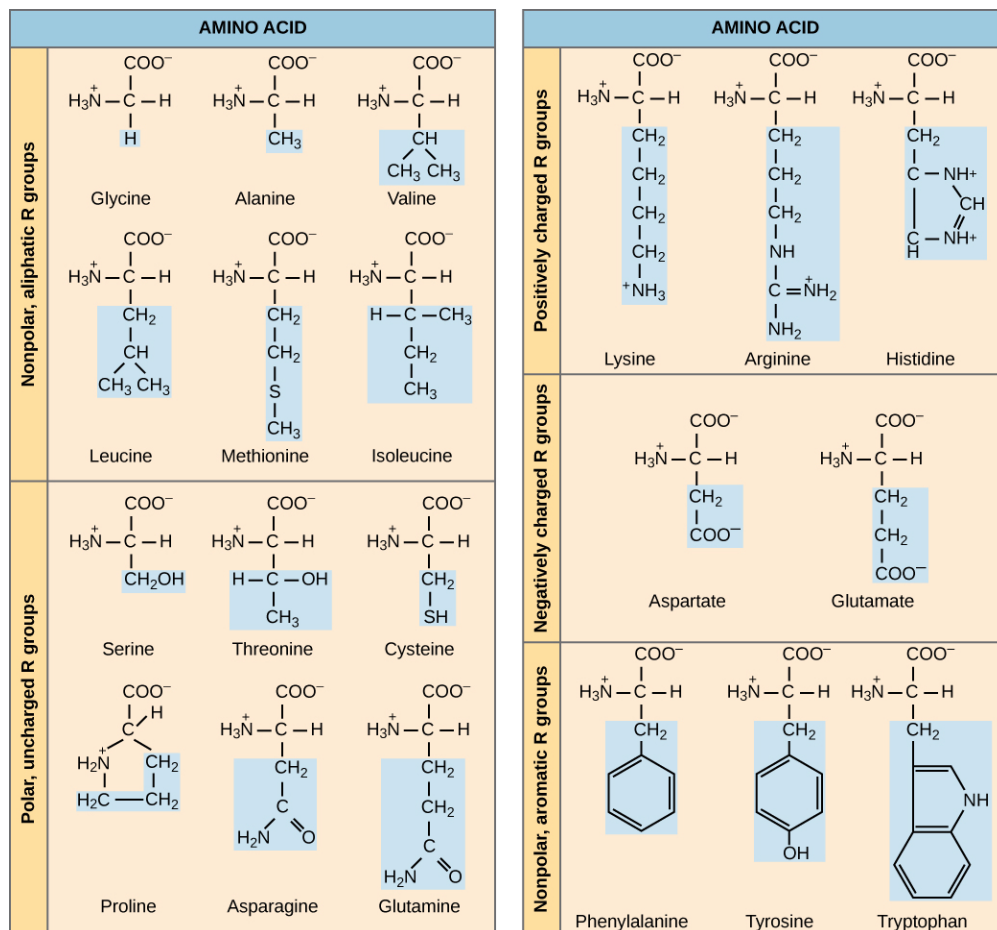
Figure 15.2 Structures of the 20 amino acids found in proteins are shown. Each amino acid is composed of an amino group (NH_3^+), a carboxyl group (COO⁻), and a side chain (blue). The side chain may be nonpolar, polar, or charged, as well as large or small. It is the variety of amino acid side chains that gives rise to the incredible variation of protein structure and function.

The Central Dogma: DNA Encodes RNA; RNA Encodes Protein

The flow of genetic information in cells from DNA to mRNA to protein is described by the **central dogma** (**Figure 15.3**), which states that genes specify the sequence of mRNAs, which in turn specify the sequence of amino acids making up all proteins. The decoding of one molecule to another is performed by specific proteins and RNAs. Because the information stored in DNA is so central to cellular function, it makes intuitive sense that the cell would make mRNA copies of this information for protein synthesis, while keeping the DNA itself intact and protected. The copying of DNA to RNA is relatively straightforward, with one nucleotide being added to the mRNA strand for every nucleotide read in the DNA strand. The translation to protein is a bit more complex because three mRNA nucleotides correspond to one amino acid in the polypeptide sequence. However, the translation to protein is still systematic and **colinear**, such that nucleotides 1 to 3 correspond to amino acid 1, nucleotides 4 to 6 correspond to amino acid 2, and so on.
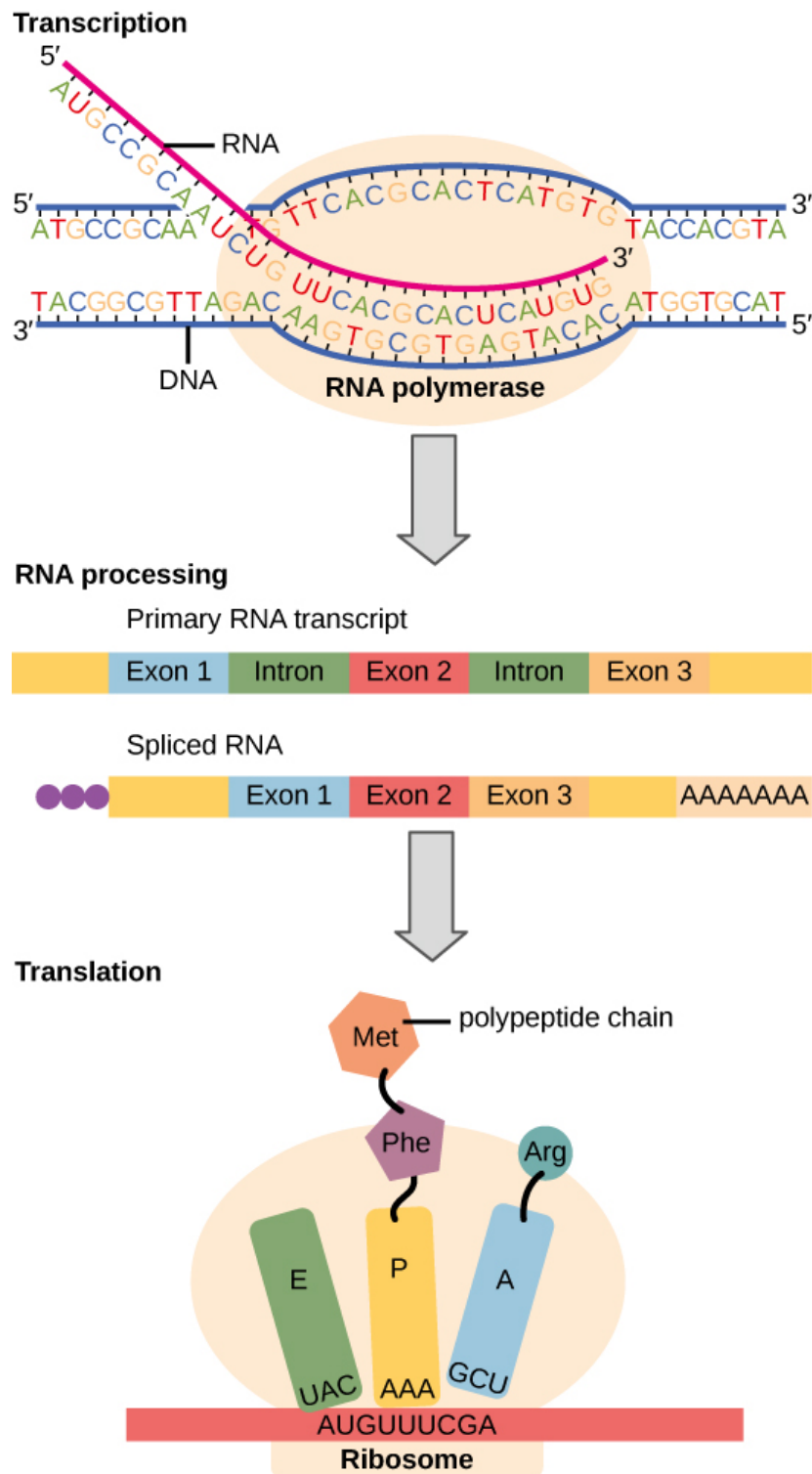
Figure 15.3 Instructions on DNA are transcribed onto messenger RNA. Ribosomes are able to read the genetic information inscribed on a strand of messenger RNA and use this information to string amino acids together into a protein.

The Genetic Code Is Degenerate and Universal

Each amino acid is defined by a three-nucleotide sequence called the triplet codon. Given the different numbers of "letters" in the mRNA and protein "alphabets," scientists theorized that single amino acids must be represented by combinations of nucleotides. Nucleotide doublets would not be sufficient to specify every amino acid because there are only 16 possible two-nucleotide combinations (4^2). In contrast, there are 64 possible nucleotide triplets

(4^3), which is far more than the number of amino acids. Scientists theorized that amino acids were encoded by nucleotide triplets and that the genetic code was **"degenerate."** In other words, a given amino acid could be encoded by more than one nucleotide triplet. This was later confirmed experimentally: Francis Crick and Sydney Brenner used the chemical mutagen proflavin to insert one, two, or three nucleotides into the gene of a virus. When one or two nucleotides were inserted, the normal proteins were not produced. When three nucleotides were inserted, the protein was synthesized and functional. This demonstrated that the amino acids must be specified by groups of three nucleotides. These nucleotide triplets are called **codons**. The insertion of one or two nucleotides completely changed the triplet **reading frame**, thereby altering the message for every subsequent amino acid (**Figure 15.5**). Though insertion of three nucleotides caused an extra amino acid to be inserted during translation, the integrity of the rest of the protein was maintained.

Scientists painstakingly solved the genetic code by translating synthetic mRNAs in vitro and sequencing the proteins they specified (**Figure 15.4**).

Second letter

		U	C	A	G	
First letter	**U**	UUU ⎫ Phe UUC ⎭ UUA ⎫ Leu UUG ⎭	UCU ⎫ UCC ⎪ Ser UCA ⎪ UCG ⎭	UAU ⎫ Tyr UAC ⎭ **UAA Stop** **UAG Stop**	UGU ⎫ Cys UGC ⎭ **UGA Stop** UGG Trp	U C A G
	C	CUU ⎫ CUC ⎪ Leu CUA ⎪ CUG ⎭	CCU ⎫ CCC ⎪ Pro CCA ⎪ CCG ⎭	CAU ⎫ His CAC ⎭ CAA ⎫ Gln CAG ⎭	CGU ⎫ CGC ⎪ Arg CGA ⎪ CGG ⎭	U C A G
	A	AUU ⎫ AUC ⎪ Ile AUA ⎪ **AUG Met**	ACU ⎫ ACC ⎪ Thr ACA ⎪ ACG ⎭	AAU ⎫ Asn AAC ⎭ AAA ⎫ Lys AAG ⎭	AGU ⎫ Ser AGC ⎭ AGA ⎫ Arg AGG ⎭	U C A G
	G	GUU ⎫ GUC ⎪ Val GUA ⎪ GUG ⎭	GCU ⎫ GCC ⎪ Ala GCA ⎪ GCG ⎭	GAU ⎫ Asp GAC ⎭ GAA ⎫ Glu GAG ⎭	GGU ⎫ GGC ⎪ Gly GGA ⎪ GGG ⎭	U C A G

Figure 15.4 This figure shows the genetic code for translating each nucleotide triplet in mRNA into an amino acid or a termination signal in a protein. (credit: modification of work by NIH)

In addition to codons that instruct the addition of a specific amino acid to a polypeptide chain, three of the 64 codons terminate protein synthesis and release the polypeptide from the translation machinery. These triplets are called **nonsense codons**, or *stop codons*. Another codon, AUG, also has a special function. In addition to specifying the amino acid methionine, it also serves as the start codon to initiate translation. The **reading frame** for translation is set by the AUG start codon near the 5' end of the mRNA. Following the start codon, the mRNA is read in groups of three until a stop codon is encountered.

The arrangement of the coding table reveals the structure of the code. There are sixteen "blocks" of codons, each specified by the first and second nucleotides of the codons within the block, e.g., the "AC*" block that corresponds to the amino acid threonine (Thr). Some blocks are divided into a pyrimidine half, in which the codon ends with U or C, and a purine half, in which the codon ends with A or G. Some amino acids get a whole block of four codons, like alanine (Ala), threonine (Thr) and proline (Pro). Some get the pyrimidine half of their block, like histidine (His) and asparagine (Asn). Others get the purine half of their block, like glutamate (Glu) and lysine (Lys). Note that some amino acids get a block and a half-block for a total of six codons.

The specification of a single amino acid by multiple similar codons is called "degeneracy." Degeneracy is believed to be a cellular mechanism to reduce the negative impact of random mutations. Codons that specify the same amino acid typically only differ by one nucleotide. In addition, amino acids with chemically similar side chains are encoded by similar codons. For example, aspartate (Asp) and glutamate (Glu), which occupy the GA* block, are both negatively charged. This nuance of the genetic code ensures that a single-nucleotide substitution mutation might specify the same amino acid but have no effect or specify a similar amino acid, preventing the protein from being rendered completely nonfunctional.

The genetic code is nearly universal. With a few minor exceptions, virtually all species use the same genetic code for protein synthesis. Conservation of codons means that a purified mRNA encoding the globin protein in horses could be transferred to a tulip cell, and the tulip would synthesize horse globin. That there is only one genetic code is powerful evidence that all of life on Earth shares a common origin, especially considering that there are about 10^{84} possible combinations of 20 amino acids and 64 triplet codons.

LINK TO LEARNING

Transcribe a gene and translate it to protein using complementary pairing and the genetic code at this **site (http://openstaxcollege.org/l/create_protein)** .
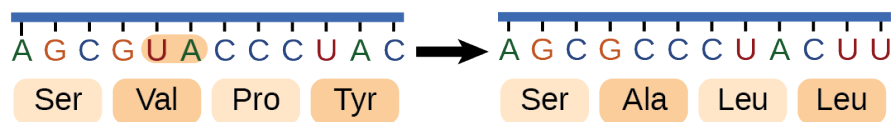

Figure 15.5 The deletion of two nucleotides shifts the reading frame of an mRNA and changes the entire protein message, creating a nonfunctional protein or terminating protein synthesis altogether.

scientific method CONNECTION

Which Has More DNA: A Kiwi or a Strawberry?

Figure 15.6 Do you think that a kiwi or a strawberry has more DNA per fruit? (credit "kiwi": "Kelbv"/Flickr; credit: "strawberry": Alisdair McDiarmid)

Question: Would a kiwi and strawberry that are approximately the same size (**Figure 15.6**) also have approximately the same amount of DNA?

Background: Genes are carried on chromosomes and are made of DNA. All mammals are diploid, meaning they have two copies of each chromosome. However, not all plants are diploid. The common strawberry is octoploid ($8n$) and the cultivated kiwi is hexaploid ($6n$). Research the total number of chromosomes in the cells of each of these fruits and think about how this might correspond to the amount of DNA in these fruits' cell nuclei. What other factors might contribute to the total amount of DNA in a single fruit? Read about the technique of DNA isolation to understand how each step in the isolation protocol helps liberate and precipitate DNA.

Hypothesis: Hypothesize whether you would be able to detect a difference in DNA quantity from similarly sized strawberries and kiwis. Which fruit do you think would yield more DNA?

Test your hypothesis: Isolate the DNA from a strawberry and a kiwi that are similarly sized. Perform the experiment in at least triplicate for each fruit

1. Prepare a bottle of DNA extraction buffer from 900 mL water, 50 mL dish detergent, and two teaspoons of table salt. Mix by inversion (cap it and turn it upside down a few times).

2. Grind a strawberry and a kiwi by hand in a plastic bag, or using a mortar and pestle, or with a metal bowl and the end of a blunt instrument. Grind for at least two minutes per fruit.

3. Add 10 mL of the DNA extraction buffer to each fruit, and mix well for at least one minute.

4. Remove cellular debris by filtering each fruit mixture through cheesecloth or porous cloth and into a funnel placed in a test tube or an appropriate container.

5. Pour ice-cold ethanol or isopropanol (rubbing alcohol) into the test tube. You should observe white, precipitated DNA.

6. Gather the DNA from each fruit by winding it around separate glass rods.

Record your observations: Because you are not quantitatively measuring DNA volume, you can record for each trial whether the two fruits produced the same or different amounts of DNA as observed by eye. If one or the other fruit produced noticeably more DNA, record this as well. Determine whether your observations are consistent with several pieces of each fruit.

Analyze your data: Did you notice an obvious difference in the amount of DNA produced by each fruit? Were your results reproducible?

Draw a conclusion: Given what you know about the number of chromosomes in each fruit, can you conclude that chromosome number necessarily correlates to DNA amount? Can you identify any drawbacks to this procedure? If you had access to a laboratory, how could you standardize your comparison and make