The Genetic Code*

OpenStax

This work is produced by OpenStax-CNX and licensed under the Creative Commons Attribution License 4.0

Abstract

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

• Explain the “central dogma” of 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 converts nucleotide-based genetic information into a protein product. Protein sequences consist of 20 commonly occurring amino acids; therefore, it can be said that the protein alphabet consists of 20 letters (Figure 1). Each amino acid is defined by a three-nucleotide sequence called the triplet codon. Different amino acids have different chemistries (such as acidic versus basic, or polar and nonpolar) and different structural constraints. Variation in amino acid sequence gives rise to enormous variation in protein structure and function.

*Version 1.9: Feb 25, 2016 9:40 pm +0000
†http://creativecommons.org/licenses/by/4.0/
Figure 1: Structures of the 20 amino acids found in proteins are shown. Each amino acid is composed of an amino group ($\text{NH}_2^+$), 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.

1 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 2), which states that genes specify the sequence of mRNAs, which in turn specify the sequence of 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 2: 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.
1.1 The Genetic Code Is Degenerate and Universal

Given the different numbers of “letters” in the mRNA and protein “alphabets,” scientists theorized that combinations of nucleotides corresponded to single amino acids. 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, protein synthesis was completely abolished. When three nucleotides were inserted, the protein was synthesized and functional. This demonstrated that three nucleotides specify each amino acid. 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 4). 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 3).
In addition to instructing 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.

The genetic code is universal. With a few 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.

http://cnx.org/content/m44522/1.9/
Transcribe a gene and translate it to protein using complementary pairing and the genetic code at this site\(^1\).

\(^1\text{http://openstaxcollege.org/l/create_protein}\)
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. This nuance of the genetic code ensures that a single-nucleotide substitution mutation might either specify the same amino acid but have no effect or specify a similar amino acid, preventing the protein from being rendered completely nonfunctional.

**Question:** Which Has More DNA: A Kiwi or a Strawberry?

**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. 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 kiwifruit 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 it more quantitative?

### 2 Section Summary

The genetic code refers to the DNA alphabet (A, T, C, G), the RNA alphabet (A, U, C, G), and the polypeptide alphabet (20 amino acids). The Central Dogma describes the flow of genetic information in the cell from genes to mRNA to proteins. Genes are used to make mRNA by the process of transcription; mRNA is used to synthesize proteins by the process of translation. The genetic code is degenerate because 64 triplet codons in mRNA specify only 20 amino acids and three nonsense codons. Almost every species on the planet uses the same genetic code.

### 3 Review Questions

**Exercise 1**

The AUC and AUA codons in mRNA both specify isoleucine. What feature of the genetic code explains this?

- a. complementarity
- b. nonsense codons
- c. universality
- d. degeneracy

*(Solution on p. 11.)*
Exercise 2

How many nucleotides are in 12 mRNA codons?

a. 12
b. 24
c. 36
d. 48

(Solution on p. 11.)

4 Free Response

Exercise 3

Imagine if there were 200 commonly occurring amino acids instead of 20. Given what you know about the genetic code, what would be the shortest possible codon length? Explain.

Exercise 4

Discuss how degeneracy of the genetic code makes cells more robust to mutations.

(Solution on p. 11.)
Solutions to Exercises in this Module

to Exercise (p. 9)
D

to Exercise (p. 10)
C

to Exercise (p. 10)
For 200 commonly occurring amino acids, codons consisting of four types of nucleotides would have to be at least four nucleotides long, because $4^4 = 256$. There would be much less degeneracy in this case.

to Exercise (p. 10)
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. This nuance of the genetic code ensures that a single-nucleotide substitution mutation might either specify the same amino acid and have no effect, or may specify a similar amino acid, preventing the protein from being rendered completely nonfunctional.

Glossary

**Definition 5: Central Dogma**
states that genes specify the sequence of mRNAs, which in turn specify the sequence of proteins

**Definition 5: codon**
three consecutive nucleotides in mRNA that specify the insertion of an amino acid or the release of a polypeptide chain during translation

**Definition 5: colinear**
in terms of RNA and protein, three “units” of RNA (nucleotides) specify one “unit” of protein (amino acid) in a consecutive fashion

**Definition 5: degeneracy**
(of the genetic code) describes that a given amino acid can be encoded by more than one nucleotide triplet; the code is degenerate, but not ambiguous

**Definition 5: nonsense codon**
one of the three mRNA codons that specifies termination of translation

**Definition 5: reading frame**
sequence of triplet codons in mRNA that specify a particular protein; a ribosome shift of one or two nucleotides in either direction completely abolishes synthesis of that protein