13.2 INFORMATION FLOW FROM DNA TO PROTEIN: AN OVERVIEW

LEARNING OBJECTIVES

3 Outline the flow of genetic information in cells, from DNA to RNA to polypeptide.

4 Compare the structures of DNA and RNA.

5 Explain why the genetic code is said to be redundant and virtually universal, and discuss how these features may reflect its evolutionary history.

Although the sequence of bases in DNA determines the sequence of amino acids in polypeptides, cells do not use the information in DNA directly. Instead, a related nucleic acid, **ribonucleic acid (RNA),** links DNA and protein. When a gene that codes for a protein is expressed, first an RNA copy is made of the information in the DNA. It is this RNA copy that provides the information that directs polypeptide synthesis.

Like DNA, RNA is a polymer of nucleotides, but it has some important differences (**FIG. 13-3**). RNA is usually single-stranded, although internal regions of some RNA molecules may have complementary sequences that allow the strand to fold back and pair to form short, double-stranded segments. As shown in Figure 13-3, the sugar in RNA is **ribose,** which is similar to deoxyribose of DNA but has a hydroxyl group at the 2′ position. (Compare ribose with the deoxyribose of DNA, shown in Figure 12-4, which has a hydrogen at the 2′ position.) The base **uracil** substitutes for thymine and, like thymine, is a pyrimidine that can form two hydrogen bonds with adenine. Hence, uracil and adenine are a complementary pair.

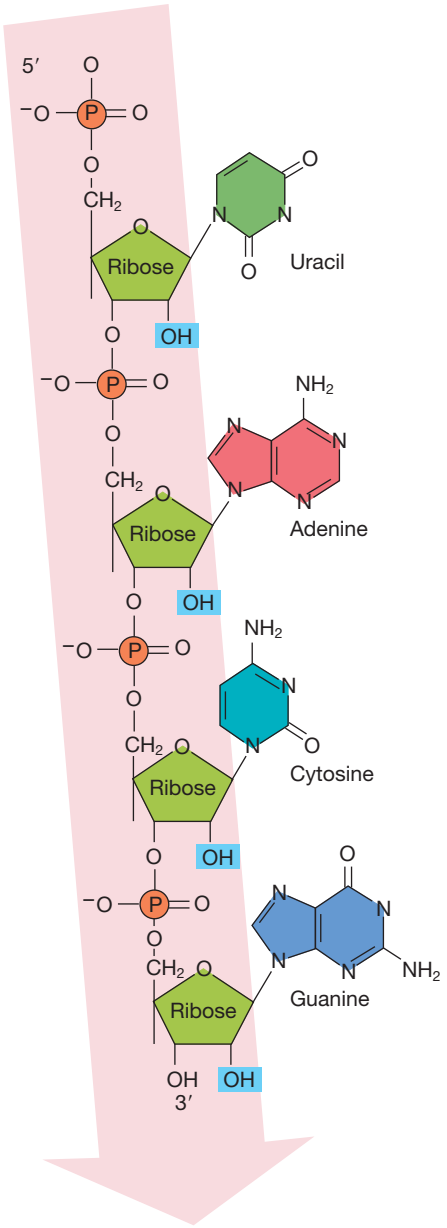


FIGURE 13-3 The nucleotide structure of RNA

The nucleotide subunits of RNA are joined by 5′ ⟶ 3′ phosphodiester linkages, like those found in DNA. Adenine, guanine, and cytosine are present, as in DNA, but the base uracil replaces thymine. All four nucleotide types contain the five-carbon sugar ribose, which has a hydroxyl group (*blue*) on its 2′ carbon atom.

DNA is transcribed to form RNA

The process by which RNA is synthesized resembles DNA replication in that the sequence of bases in the RNA strand is determined by complementary base pairing with one of the DNA strands, the **template strand** (**FIG. 13-4**). Because RNA synthesis takes the information in one kind of nucleic acid (DNA) and copies it as another nucleic acid (RNA), this process is called **transcription** ("copying").

Three main kinds of RNA molecules are transcribed: messenger RNA, transfer RNA, and ribosomal RNA. **Messenger RNA (mRNA)** is a single strand of RNA that carries the information for making a protein. Each of the 45 or so kinds of **transfer RNAs (tRNAs)** is a single strand of RNA that folds back on itself to form a specific shape. Each kind of tRNA bonds with only one kind of amino acid and carries it to the *ribosome.* (Because there are more kinds of tRNA molecules than there are amino acids, many amino acids are carried by two or more kinds of tRNA molecules.) **Ribosomal RNA (rRNA),** which is in a globular form, is an important part of the structure of ribosomes and has catalytic functions needed during protein synthesis.

RNA is translated to form a polypeptide

Following transcription, the transcribed information in the mRNA is used to specify the amino acid sequence of a polypeptide (see Fig. 13-4). This process is called **translation** because it involves conversion of the "nucleic acid language" in the mRNA molecule into the "amino acid language" of protein.

In translation of the genetic instructions to form a polypeptide, a sequence of three consecutive bases in mRNA, called a **codon,** specifies one amino acid. For example, one codon that corresponds to the amino acid phenylalanine is 5′—UUC—3′. Because each codon consists of three nucleotides, the code is described as a **triplet code.** The assignments of codons for amino acids and for start and stop signals are collectively named the **genetic code** (**FIG. 13-5**).

KEY POINT Protein synthesis requires two major steps: DNA $\xrightarrow{\text{transcription}}$ RNA $\xrightarrow{\text{translation}}$ protein

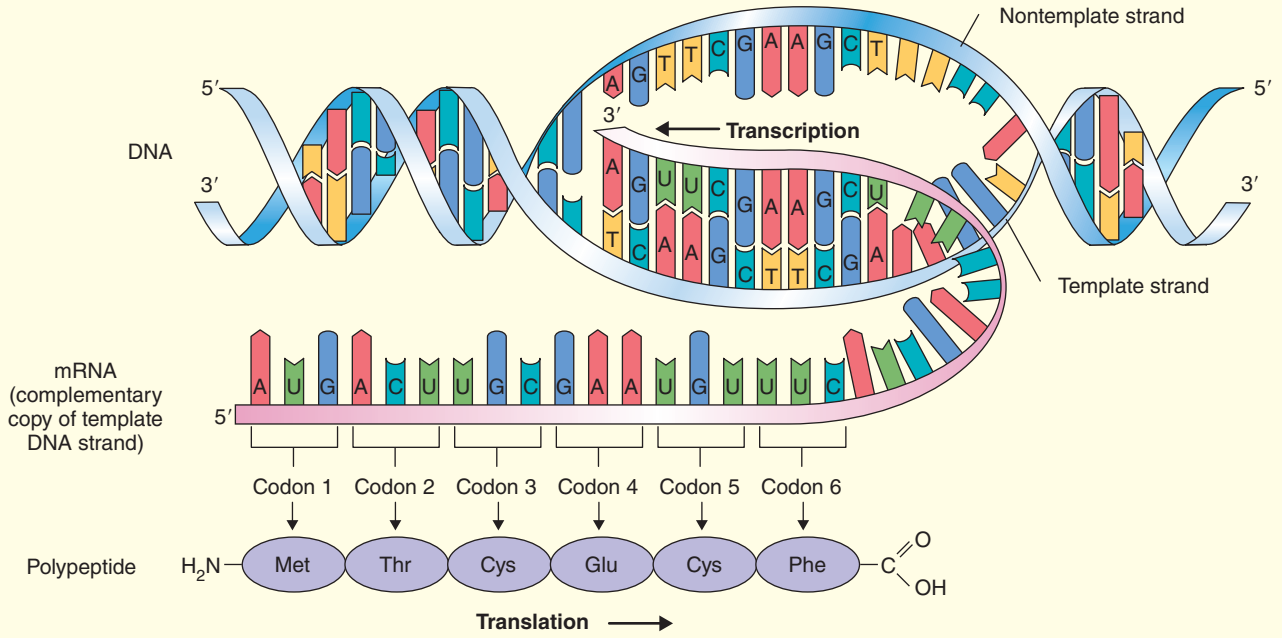


FIGURE 13-4 An overview of transcription and translation

In transcription, messenger RNA is synthesized as a complementary copy of one of the DNA strands, the template strand. Messenger RNA carries genetic information in the form of sets of three bases, or codons, each of which specifies one amino acid. Codons are translated consecutively, thus specifying the linear sequence of amino acids in the polypeptide chain. Translation requires tRNA and ribosomes (*not shown*). The figure depicts transcription and translation in bacteria. In eukaryotes, transcription takes place in the nucleus and translation occurs in the cytosol.

First letter (5′ end)	Second letter: U	Second letter: C	Second letter: A	Second letter: G	Third letter (3′ end)
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met or start	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G

= Stop codon

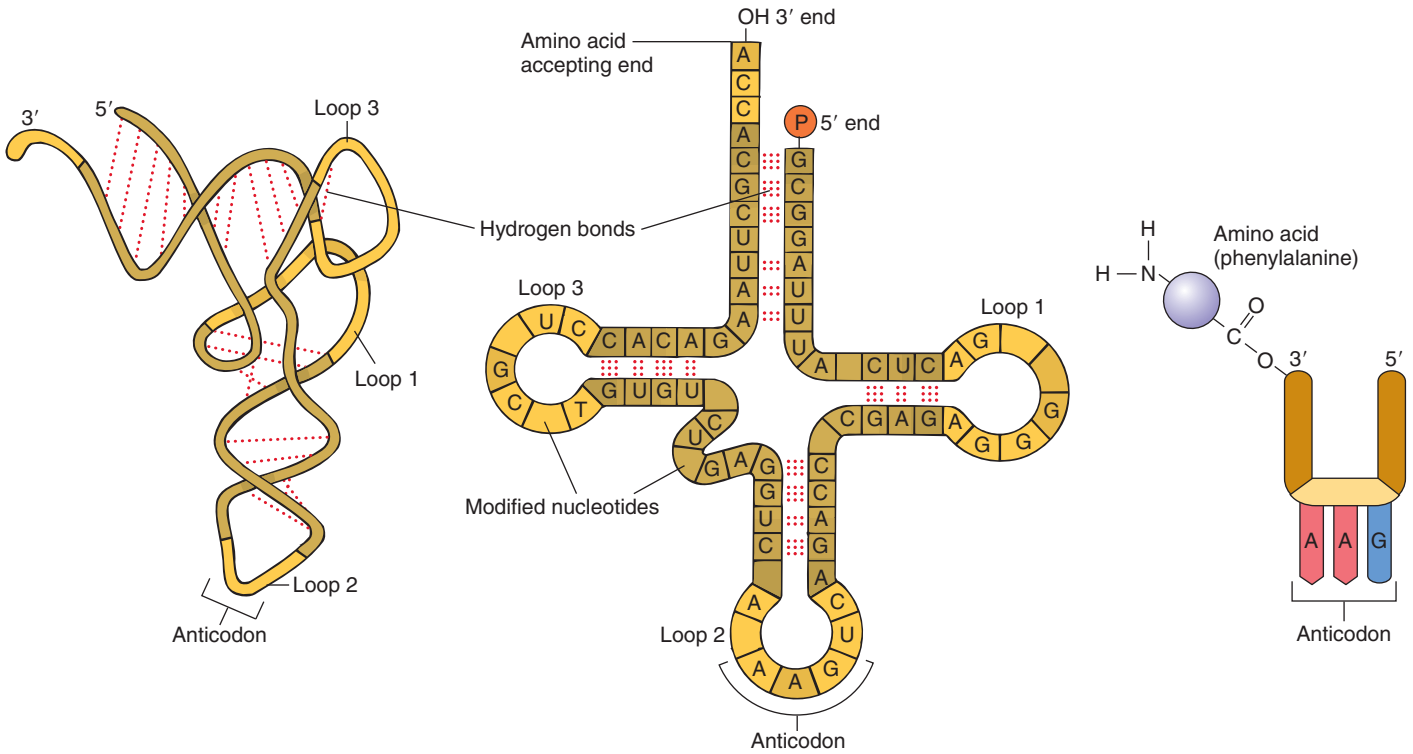
= Start codon

Transfer RNAs are crucial parts of the decoding machinery because they act as "adapters" that connect amino acids and nucleic acids. This mechanism is possible because each tRNA can (1) link with a specific amino acid and (2) recognize the appropriate mRNA codon for that particular amino acid (**FIG. 13-6**). A particular tRNA can recognize a particular codon because it has a sequence of three bases, called the **anticodon,** that hydrogen-bonds with the mRNA codon by complementary base pairing. The exact anticodon that is complementary to the codon for phenylalanine in our example is 3′—AAG—5′.

Translation requires that each tRNA anticodon be hydrogen-bonded to the complementary mRNA codon and that the amino acids carried by the tRNAs be linked in the order specified by the sequence of codons in the mRNA. **Ribosomes,** the site of translation, are organelles composed of two different subunits, each containing protein and rRNA. (The structure and function of ribosomes were introduced in Chapter 4.) Ribosomes attach to the 5′

FIGURE 13-5 The genetic code

The genetic code specifies all possible combinations of the three bases that compose codons in mRNA. Of the 64 possible codons, 61 specify amino acids (see Fig. 3-17 for an explanation of abbreviations). The codon AUG specifies the amino acid methionine and also signals the ribosome to initiate translation ("start"). Three codons—UAA, UGA, and UAG—do not specify amino acids; they terminate polypeptide synthesis ("stop").



(a) The 3-D shape of a tRNA molecule is determined by hydrogen bonds formed between complementary bases.

(b) One loop contains the anticodon; these unpaired bases pair with a complementary mRNA codon. The amino acid attaches to the terminal nucleotide at the hydroxyl (OH) 3′ end.

(c) This stylized diagram of an aminoacyl-tRNA shows that the amino acid attaches to tRNA by its carboxyl group, leaving its amino group exposed for peptide bond formation.

FIGURE 13-6 Three representations of a tRNA molecule

end of the mRNA and travel along it, allowing the tRNAs to attach sequentially to the codons of mRNA. In this way the amino acids carried by the tRNAs take up the proper position to be joined by *peptide bonds* in the correct sequence to form a polypeptide.

Biologists cracked the genetic code in the 1960s

Before the genetic code was deciphered, scientists had become interested in how a genetic code might work. The Watson and Crick model of DNA showed it to be a linear sequence of four different nucleotides. If each nucleotide coded for a single amino acid, the genetic code could specify only 4 amino acids, not the 20 found in the vast variety of proteins in the cell.

Scientists saw that the DNA bases could serve as a four-letter "alphabet" and hypothesized that three-letter combinations of the four bases would make it possible to form a total of 64 "words," more than enough to specify all the naturally occurring amino acids. In 1961, Crick and British scientist Sydney Brenner concluded from experimental evidence that the code used nonoverlapping triplets of bases. They predicted that the code is read, one triplet at a time, from a fixed starting point that establishes the **reading frame** for the genetic message.

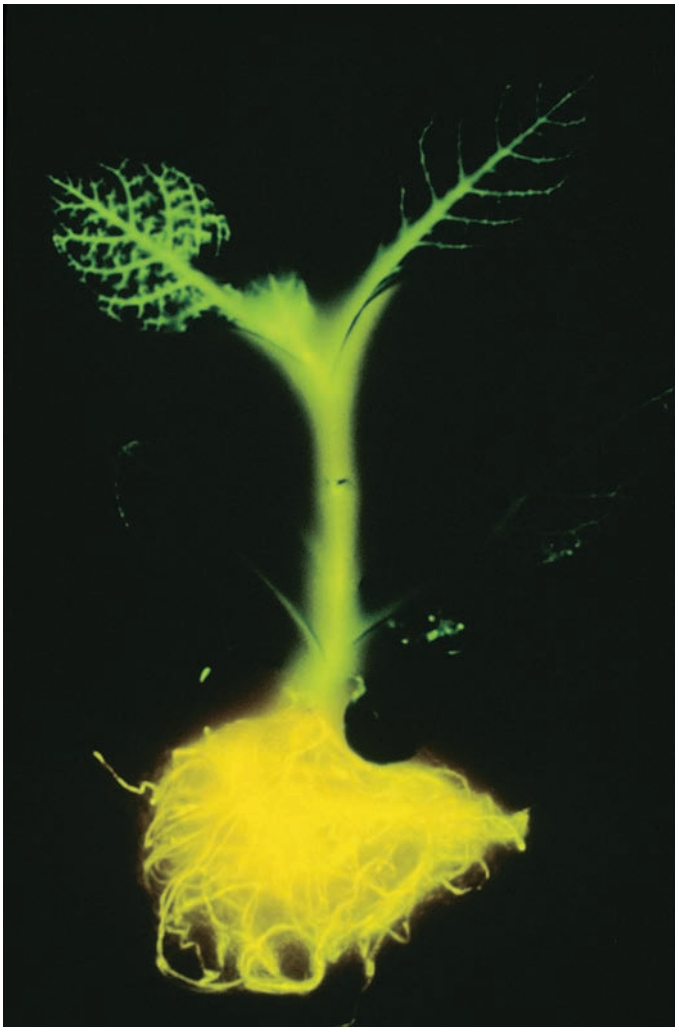
Marshall Nirenberg, a U.S. biochemist, and his postdoctoral researcher, Heinrich Matthaei, studied protein synthesis outside living cells in purified *cell-free systems* derived from the bacterium *Escherichia coli*. Nirenberg and Matthaei obtained the first experimental evidence indicating the assignment of triplets to specific amino acids. By constructing artificial mRNA molecules with known base sequences, they determined which amino acids would be incorporated into protein. For example, when they added the synthetic mRNA polyuridylic acid (UUUUUUUUU . . .) to a mixture of purified ribosomes, aminoacyl-tRNAs (amino acids linked to their respective tRNAs), and essential cofactors needed to synthesize polypeptide, only phenylalanine was incorporated into the resulting polypeptide chain. The inference was inescapable that the UUU triplet codes for phenylalanine. Similar experiments showed that polyadenylic acid (AAAAAAAAA . . .) codes for a polypeptide of lysine and that polycytidylic acid (CCCCCCCCC . . .) codes for a polypeptide of proline.

By using mixed nucleotide polymers (such as a random polymer of A and C) as artificial messengers, researchers such as H. Gobind Khorana, then at the University of Wisconsin, assigned the other codons to specific amino acids. However, three of the codons—UAA, UGA, and UAG—did not specify any amino acids. These codons are now known to be the signals that terminate the coding sequence for a polypeptide chain. By 1967, the genetic code was completely "cracked," and scientists had identified the coding assignments of all 64 possible codons shown in Figure 13-5. In 1968, Nirenberg and Khorana received the Nobel Prize in Physiology or Medicine for their work in deciphering the genetic code.

Remember that the genetic code is an mRNA code. The tRNA anticodon sequences, as well as the DNA sequence from which the message is transcribed, are complementary to the sequences in Figure 13-5. For example, the mRNA codon for the amino acid methionine is 5′—AUG—3′. It is transcribed from the DNA base sequence 3′—TAC—5′, and the corresponding tRNA anticodon is 3′—UAC—5′.

The genetic code is virtually universal

Perhaps the single most remarkable feature of the code is that it is essentially universal. Over the years, biologists have examined the genetic code in a diverse array of species and found it the same in organisms as different as bacteria, redwood trees, jellyfish, and humans (**FIG. 13-7**). These findings strongly suggest that the code is an ancient legacy that evolved early in the history of life (discussed in Chapter 21).



Keith V. Wood/Visuals Unlimited

FIGURE 13-7 Genetically engineered tobacco plant
This plant glows as it expresses the luciferase gene from a firefly. Luciferase is an enzyme that catalyzes a reaction that produces a flash of light. This classic experiment indicated that animal genes can be expressed in plants.

Scientists have discovered some minor exceptions to the universality of the genetic code. In several unicellular protozoa, UAA and UGA code for the amino acid glutamine instead of functioning as stop signals. Other exceptions are found in mitochondria, which contain their own DNA and protein synthesis machinery for some genes. These slight coding differences vary with the organism, but keep in mind that in each case, all the other coding assignments are identical to those of the standard genetic code.

The genetic code is redundant

Given 64 possible codons and only 20 common amino acids, it is not surprising that more than one codon specifies certain amino acids. This redundancy in the genetic code has certain characteristic patterns. The codons CCU, CCC, CCA, and CCG are "synonymous" in that they all code for the amino acid proline. The only difference among the 4 codons involves the nucleotide at the 3′ end of the triplet. Although the code may be read three nucleotides at a time, only the first two nucleotides seem to contain specific information for proline. A similar pattern can be seen for several other amino acids. Only methionine and tryptophan are specified by single codons. Each of the other amino acids is specified by 2 to 6 different codons.

Crick first proposed this apparent breach of the base-pairing rules as the **wobble hypothesis.** He reasoned that the third nucleotide of a tRNA anticodon (which is the 5′ base of the anticodon sequence) may sometimes form hydrogen bonds with more than one kind of third nucleotide (the 3′ base) of a codon. Investigators later established this experimentally by determining the anticodon sequences of tRNA molecules and testing their specificities in artificial systems. Some tRNAs bond exclusively to one codon, but other tRNA molecules pair with as many as three separate codons that differ in their third nucleotide but that specify the same amino acid. Thus, the wobble hypothesis accounts for the possible variation in base pairing between the third base of a codon and the corresponding base in its anticodon. "Wobble" results in several acceptable forms of base pairing between mRNA and tRNA.

Review

- Sketch a simple flow diagram that shows the relationships among the following: RNA, translation, DNA, transcription, and polypeptide.
- How are the structures of DNA and RNA similar? How are they different?

13.3 TRANSCRIPTION

LEARNING OBJECTIVES

6 Compare the processes of transcription and DNA replication, identifying both similarities and differences.
7 Compare bacterial and eukaryotic mRNAs, and explain the functional significance of their structural differences.

Now that we have presented an overview of information flow from DNA to RNA to polypeptide, let us examine the entire pro-