20.19: Information Storage

How can DNA and RNA molecules act as blueprints for the manufacture of proteins? The exact details were unraveled in the early 1960s mainly by Marshall Nirenberg (born 1927) at the National Institutes of Health and H. G. Khorana (born 1922) at the University of Wisconsin, work which earned them the Nobel prize in 1968. They showed that each amino acid in a protein is determined by a specific **codon** of three nitrogenous bases in the DNA or RNA chain. The details of this **genetic code** are given in the table below. As an example of how this code works, let us take the section of RNA shown in Figure 3 on Nucleic Acid Structure. This has the sequence UCAUGG. This is part of the instructions for building a polypeptide chain containing the amino acid serine (UCA) followed by the amino acid tryptophan (UGG).

Table 20.19.1The Genetic Code for RNA

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

(First Base in Codon shown as the leftmost column label)

> **📌 Note**
>
> (*a*) A termination codon is indicated by TERM. (*b*) AUG, the codon for methionine is also the initiation codon. All protein synthesis begins at this codon, though this initial methionine is often removed during post-transcriptional processing.

Since each codon corresponds to three places in the nucleic acid chain and since there are four kinds of nitrogenous bases to fill each place, there are a total of $4^3 = 64$ different possible codons. Since there are only 20 amino acids, the genetic code is *degenerate*—several different codons correspond to the same amino acid. This degeneracy acts as a safeguard against errors in reading the code. Thus UCU, UCC, UCA, and UCG all correspond to serine. If a mistake is made in reading the third base in this triplet, no harm is done since serine is still produced. On the molecular level **transfer RNAs (tRNA)**, the molecules reading the codons and providing the correct amino acid, can pair with multiple codons. This only occurs in terms of the third base in the codon. For instance, G pairs with C, but is also capable of pairing with U. Some tRNAs even employ a fifth nitrogenous base, inosinate(I) which is capable of pairing with A, U or C. This use of multiple pairing with the third codon by tRNA is called the *wobble hypothesis*, and was first proposed by Francis Crick. Notice that while a tRNA can pair with multiple codon in the wobble hypothesis, it can only pair with codons for the same amino acid, and each codon is still specific to only one amino acid.[1]

There are three additional features of the genetic code. First, AUG, the codon for Methionine also serves as an **initiation codon**, and, with help from other signals, is where protein systhesis begins. A second feature is that reading RNA for protein synthesis goes from the 5' carbon end of the nucleic acid to the 3' carbon end. A final important feature of the genetic code is the existence of three **termination codons**. These correspond to an instruction for ending a polypeptide chain. How these features work is best illustrated by an example.

> **✔ Example 20.19.1: RNA**
>
> Decode the RNA fragment
>
> 5' A C C U U A U G A C G C C U G U C C A U U A A C G A U 3'
>
> **Solution**

First, we must decide which direction to read the RNA code. Synthesis goes from the 5' end to the 3' end, so this segment is read left to right. Had it been displayed 3' to 5', we would have needed to read it from right to left.

Second, we need to look for an initiation codon, AUG. This codon appears starting at the sixth letter in. Thus, we can divide the sequence up like this, with the start codon bold:

AC|CUU|**AUG**|ACG|CCU|GUC|CAU|UAA|CGA|U

Third, let us see if there is a stop codon in this sequence. Sure enough, the fifth codon after the start codon, UAA is a stop codon. Thus, the entire sequence to be translated, in bold:

AC|CUU|**AUG|ACG|CCU|GUC|CAU|UAA**|CGA|U

which translates to the amino acid sequence:

Met-Thr-Pro-Val-His-STOP

Notice in the example, that if we had not started with the initiation codon, an entirely different protein would have been formed. Look at what would have happened if we had simply started at the beginning of the sequence:

ACC|UUA|UGA|CGC|CUG|UCC|AUU|AAC|GAU|

a stop codon appears in a new place, and the translated protein is:

Thr-Leu-STOP

This highlights the importance of the **reading frame**, the place where codons start being read. Notice that, since codons are 3 bases long, any sequence has three different reading frames. Without the initiation codon, there would be no way to identify the correct reading frame. In addition to the AUG initiation codon, other element regulate initiation. In bacteria, a sequence of bases before the initiation codon, called the **Shine-Dalgarno sequence** precedes the AUG codon, specifying where to begin translation. A different set up occurs in eukaryotes. An initiation complex forms, but instead of having a specific sequence connected to the initiation codon, the complex slides along the mRNA strand, until it finds the AUG initiation codon.[2]

1. ↑ Nelson, D.L., Cox, M.M. *Lehninger Principles of Biochemistry(5thed).* New York: W.H. Freeman and Company, 2008. pp. 1070-1072.
2. ↑ Nelson, D.L., Cox, M.M. *Lehninger Principles of Biochemistry(5thed).* New York: W.H. Freeman and Company, 2008. pp. 1088-1090.