

**LECTURE NO. NINE**

## **THE GENETIC CODE**

### *Gene expression*

Genetic information is encoded in the base sequence of DNA molecules as a series of genes. Gene expression is the term used to describe how cells decode the information to synthesize proteins required for cellular function. The expression of a gene involves the synthesis of a complementary RNA molecule whose sequence specifies the amino acid sequence of a protein. The DNA sequence of the gene is colinear with the amino acid sequence of the polypeptide.

### **Genetic code:**

Amino acids are encoded by 64 base triplets called codons which encode the 20 amino acids. Most amino acids have more than one codon. This is known as the degeneracy of the genetic code and it helps to minimize the effect of mutations. Codons that specify the same amino acid are called synonyms and differ at their third base, known as the 'wobble' position. AUG is the initiation codon and encodes methionine. There are three stop codons: UAG, UGA and UAA.

### **Reading frames:**

Three possible sets of codons can be read from any sequence depending on which base is chosen as the start of a codon. Each set of codons is known as a reading frame. The initiation codon determines the reading frame of the protein coding sequence. Other reading frames tend to contain stop codons and are not used for protein synthesis. An open reading frame is a sequence of codons bounded by start and stop codons.

University of the code:

The genetic code applies universally with all organisms using the same codons for each amino acid. However, some exceptions to the standard codon usage occur in mitochondrial genomes and in some unicellular organisms.

### Gene expression:

The information required by an organism to reproduce itself is carried by its DNA, encoded in the base sequence and organized as a series of genes. Gene expression is the term used to describe the process by which cells decode and make use of this information to synthesize the proteins that are responsible for cellular function. During gene expression, information is copied from DNA to RNA by the synthesis of an RNA molecule whose base sequence is complementary to that of the DNA template. The RNA then directs the synthesis of a protein whose amino acid sequence is specified by the base sequence of the RNA. For every gene the DNA sequence is colinear with the amino acid sequence of the polypeptide it encodes such that the 5'→3' base sequence of the coding strand specifies the amino acid sequence of the encoded polypeptide from the amino to the carboxy terminus.

### Genetic code:

The genetic code describes how base sequences are converted into amino acid sequences during protein synthesis. The DNA sequence of a gene is divided into a series of units of three bases. Each set of three bases is called a codon and specifies a particular amino acid. The four bases in DNA and RNA can combine as a total of  $4^3 = 64$  codons which specify the 20 amino acids found in proteins (Table 1). Because the number of codons is greater, all of the amino acids, with the exceptions of methionine and tryptophan, are encoded by more than one codon. This feature is referred to as the degeneracy or the redundancy of the genetic code. Codons which specify the same amino acid are called synonyms and tend to be similar. For example, ACU, ACC, ACA and ACG all specify the amino acid threonine. Variations between synonyms tend to occur at the third position of the codon, which is known as the wobble position. The degeneracy of the genetic code minimizes the effects of mutations so that alterations to the base sequence are teas

likely to change the amino acid encoded and possible deleterious effects on protein function are avoided. Of the 64 possible codons, 61 encode amino acids. The remaining three, **UAG**, **UGA**, and **UAA**, do not encode amino acids but instead act as signals for protein synthesis to stop and as such are known as termination codons or stop codons. The codon for methionine, **AUG**, is the signal for protein synthesis to start and is known as the initiation codon. Thus all **polypeptides** start with methionine although mis is sometimes removed later.

|                         |                | Second position |                 |                 |                |   |                         |
|-------------------------|----------------|-----------------|-----------------|-----------------|----------------|---|-------------------------|
|                         |                | U               | C               | A               | G              |   |                         |
| First position (5'-end) | U              | UUU <i>phe</i>  | UCU             | UAU <i>tyr</i>  | UGU <i>cys</i> | U | Third position (3'-end) |
|                         | UUC            | UCC             | UAC             | UGC             | C              |   |                         |
|                         | UUA <i>leu</i> | UCA <i>ser</i>  | UAA <i>Stop</i> | UGA <i>Stop</i> | A              |   |                         |
|                         | UUG            | UCG             | UAG <i>Stop</i> | UGG <i>trp</i>  | G              |   |                         |
| C                       | CUU            | CCU             | CAU <i>his</i>  | CGU             | U              |   |                         |
| CUC                     | CCC <i>pro</i> | CAC             | CGC             | C               |                |   |                         |
| CUA <i>leu</i>          | CCA            | CAA <i>gln</i>  | CGA             | A               |                |   |                         |
| CUG                     | CCG            | CAG             | CGG             | G               |                |   |                         |
| A                       | AUU            | ACU             | AAU <i>asn</i>  | AGU <i>ser</i>  | U              |   |                         |
| AUC <i>ile</i>          | ACC <i>thr</i> | AAC             | AGC             | C               |                |   |                         |
| AUA                     | ACA            | AAA <i>lys</i>  | AGA             | A               |                |   |                         |
| AUG <i>met</i>          | ACG            | AAG             | AGG             | G               |                |   |                         |
| G                       | GUU            | GCU             | GAU <i>asp</i>  | GGU             | U              |   |                         |
| GUC                     | GCC <i>ala</i> | GAC             | GGC             | C               |                |   |                         |
| GUA <i>val</i>          | GCA            | GAA <i>glu</i>  | GGA             | A               |                |   |                         |
| GUG                     | GCG            | GAG             | GGG             | G               |                |   |                         |

■ Initiation ■ Termination

### Reading frames:

In addition to identifying the start of protein synthesis, the initiation codon determines the reading frame of the RNA sequence. Depending on which base is chosen as the start of a codon, three possible sets of codons may be read from any base sequence. In practice, during protein synthesis, normally only one reading frame contains useful information; the other two reading frames usually contain several stop codons which prevent them from being used to

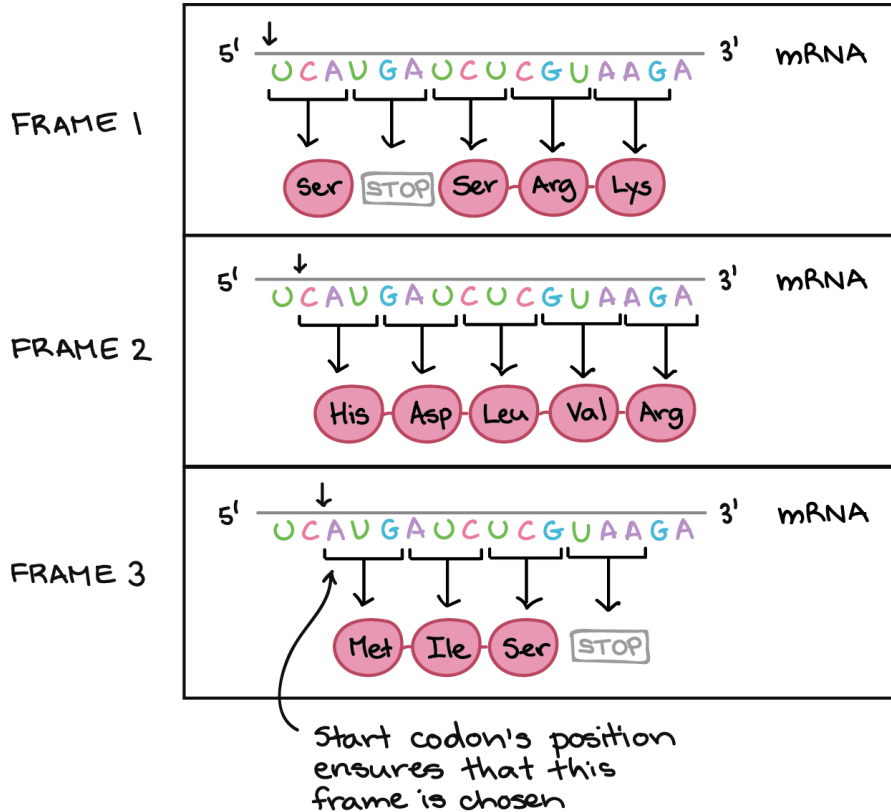


Fig. 1. Every DNA sequence can be read as three separate reading frames depending on which base is chosen as the start of the codon.

direct **protein** synthesis (Fig. 1). A set of **codons** that runs continuously and is **bounded** at the start by an initiation **codon** and at the end by a termination **codon** is known as an **open reading frame (ORF)**. This characteristic is used to **identify** protein coding **DNA** sequences in **genome** sequencing projects.

### Universality of the code:

Initially the genetic code was believed to apply universally, that is all organisms would recognize individual codons as the same **amino** acids. However, it has now been shown that some variation in the code exists, although this is rare. For example, **mitochondria** have a small DNA **genome** containing about 20 genes in which deviations from the genetic code occur. Changes are mostly associated with start and stop codons. For example, **UGA**, which is normally a termination codon, codes for **tryptophan** whereas **AGA** and **AGG** which normally encode **arginine** are

termination codons, and **AUA**, normally **isoleucine**, specifies **methionine**. It is thought that these changes tend to be viable because the **mitochondrion** is a closed system. A few examples of **nonstandard** codon usage have now been found outside **mitochondrial genomes** in unicellular organisms. For example **UAA** and **UAG** which are normally stop codons, encode **glutamic acid** in some protozoa.

.