

# The genetic code

Deciphering **the genetic code** dates back to 1961, when in vitro translation experiments revealed that a **triplet of nucleotides** always codes for a certain amino acid.

- these triplets are called **codons**
- The 4 bases of DNA and RNA can combine as  $4^3=64$  **codons**, that specify the 20 **proteinogenic amino acids** for protein formation
- **the genetic code is degenerate**
  - **the number of codons** is greater than the number of **amino acids** – all amk except tryptophan and methionine are encoded by **multiple triplets**
  - the genetic code is thus characterized by **redundancy** – we call it degenerate
  - **codons** that specify the same amino acid are called **synonymous**
  - **variations** between synonymous codons mainly concern the **3rd position** in the triplet, the decisive role is usually played by the first two bases of the codon
  - the degeneration of the genetic code minimizes **the effect of mutations**
  - due to degeneration, it is also impossible to determine the structure of the mRNA that encoded the protein, but on the contrary, we can determine the **structure of the proteins** quite precisely
- out of **664 codons** , **61** codons encode amino acids – the remaining 3 (UAG, UGA and UAA) are so-called **termination codons or stop codons and terminate** proteosynthesis (the tRNA having a complementary anticodon usually carries water instead of amk)
- **on the other hand, proteosynthesis** is initiated at the site of the **initiation codon AUG**, which codes for **methionine**
- it is located in one of the first exons and determines the **reading frame** of the RNA sequence
- each **RNA sequence** can be read by three sets of codons, depending on which base is chosen as **the start codon**
- a set of codons that is limited by **an initiation and a termination codon** is called an open reading frame - **ORF, open reading frame**
- the genetic code is considered **universal**
  - it is used by all organisms **in the same way**
  - however, there are exceptions - notably **mitochondria** and some unicellular organisms
  - ne.g. in mitochondria, UGA (usually terminating) codes for **tryptophan** and conversely AGA and AGG (usually arginine) are **terminating**
- not all synonymous codons are used in different species with the same frequency - some more, some less - the so-called **codon dialect**

---

Proteosynthesis - the biosynthesis of proteins - represents a process in which the genetic information stored and transmitted in DNA speech is expressed, i.e. realized in the form of proteins. Proteosynthesis is also called translation, because it represents the biochemical transfer of information from the speech of nucleic acids to the speech of proteins.

**Genetic information** is linear, it is stored in the form of a polynucleotide strand (usually) DNA, which is a polymer of nucleotides. Recall that it uses four letters - **nucleotides A, G, C and T** (adenine, guanine, cytosine, thymine). In contrast, proteins are composed both biochemically and stereochemically of completely different letters called amino acids. The number of these amino acids today reaches the number 22. During the formation of proteins, they also combine, polymerize into a linear fiber, and it can be said that there is colinearity (concurrency) between the order of nucleotides in DNA and the sequence of amino acids in the protein encoded by this DNA. Practically speaking, translation involves the creation of a protein fiber according to a polynucleotide fiber. By convention, polynucleotide sequences are written *in the 5'-3' direction and protein amino acid sequences from the amino N-terminus to the carboxyl C-terminus, corresponding to their reading directions.*

As already deduced by **F. Jacob** and **J. Monod**, the "conversion" of DNA into protein does not take place directly, but through an intermediary, which is mRNA. Each mRNA is a copy of a certain section of one of the DNA strands roughly the length of one or more genes. Like DNA, it therefore carries the same specific message for the production of one or more proteins. However, the nucleotide T is replaced in the mRNA by another pyrimidine nucleotide U and the sugar component here is ribose and not deoxyribose as in DNA. Additionally RNA is naturally single-stranded [1]. However, all this makes RNA much more labile than DNA. It can be easily and quickly broken down by the cell's natural mechanisms when, for example, the required level of the given protein has already been reached and its excess would be of no use to the cell or could even be harmful (e.g. an excess mRNA encoding proto-oncogenes). It is reported that in an *E. coli* cell, the average half-life of mRNA is about 2-3 minutes, in contrast, the mRNA encoding  $\beta$ -globin in a eukaryotic cell has a half-life of more than 10 hours.

Genetic code – coding rules for the 22 genetically encoded amino acids. The international one-letter abbreviations of amino acids and for the three-foot triplets their historical names Amber, Opal and Ochre are also given. The synthesis of the vast majority of proteins begins with methionine, which is coded by the codon AUG (universal start signal), exceptionally by valine (codon GUG) or another amino acid. The table clearly shows the degeneration of the genetic code expressed mainly by variability in the third codon position

<b>Ala</b>	A	GCU, GCC, GCA, GCG	<b>Phe</b>	F	UUU, UUC
<b>Arg</b>	R	CGU, CGC, CGA, CGG, AGA, AGG	<b>Pro</b>	P	CCU, CCC, CCA, CCG
<b>Asn</b>	N	AAU, AAC	<b>Ser</b>	S	UCU, UCC, UCA, UCG, AGU, AGC
<b>Asp</b>	D	GAU, GAC	<b>Thr</b>	T	ACU, ACC, ACA, ACG
<b>Cys</b>	C	UGU, UGC	<b>Trp</b>	W	UGG
<b>Gln</b>	Q	CAA, CAG	<b>Tyr</b>	Y	UAU, UAC
<b>Glu</b>	E	GAA, GAG	<b>Val</b>	V	GUU, GUC, GUA, GUG
<b>Gly</b>	G	GGU, GGC, GGA, GGG	<b>Start</b>		AUG, (GUG)
<b>His</b>	H	CAU, CAC	<b>Stop</b>		UAG ( <i>Amber</i> )
<b>Ile</b>	I	AUU, AUC, AUA	<b>Stop</b>		UGA ( <i>Opal</i> )
<b>Leu</b>	L	CUU, CUC, CUA, CUG, UUA, UUG	<b>Stop</b>		UAA ( <i>Ochre</i> )
<b>Lys</b>	K	AAA, AAG	<b>Pyl</b>	O	UAG
<b>Met</b>	M	AUG	<b>Sec</b>	U	UGA (+ <i>SECIS</i> )

What is the *encryption key*, or *code* between the four nucleotides in mRNA and the 22 amino acids in protein? It turned out that this so-called **genetic code** is three-letter, i.e. the order of three nucleotides – a **triplet** always codes for a single amino acid. From four different nucleotides in mRNA,  $4^3 = 64$  different triplets or code words, **codons** can thus be formed (see table). This is absolutely enough to specify each of the 22 amino acids, and also to create the signals that determine where in the mRNA translation should start (usually from the AUG codon) and where it should end (three stop signals– UAA, UAG and UGA). They represent a kind of period after the sentence, i.e. in our case after the synthesized protein.

The assignment of individual amino acids to individual codons, i.e. *deciphering the genetic code*, was achieved experimentally around the mid-1960s. As can be seen from the table, only two amino acids, namely methionine and tryptophan, are encoded by only one codon (AUG or UGG, respectively). For the rest of the amino acids, they are encoded by two, three, four or even six codons. Codons for the same amino acid are referred to as *synonymous* and usually differ in the third nucleotide. The existence of multiple codons for a single amino acid means that the **code is degenerate**. On the other hand, however, it is **unambiguous**, because each codon always specifies only one single amino acid. The genetic code is also almost (see below) **universal**, as it is used for translation in all organisms. This fact is one of the most convincing evidences for the common ancestor of all living organisms.

**Codons** follow one another in mRNA, without gaps and without overlapping in the so-called **reading frame**. This is set by the first – initiation – codon, which is most often AUG coding for methionine. The vast majority of newly synthesized protein chains therefore begin with this amino acid. However, methionine is very often – either before the completion of translation, or after it, already from the finished protein – cleaved. Thus, the information in the mRNA is read and translated continuously in sections three nucleotides long, and their sequence is set by the first initiation codon [2].

Not all code words for one amino acid are used in the organism with the same frequency. On the contrary, some of the synonyms are hardly used in genetic information (or only for special proteins), in contrast to the so-called common codons. At the same time, the frequency of use of individual codons is already specific for each single-celled organism, a different typical selection applies to vertebrates and another again, for example, to higher plants. We are talking about the so-called **codon dialect**.

## Links

## Related Articles

## References

1. Živa 2006, 5: 198
2. Živa 2006, 5: 198