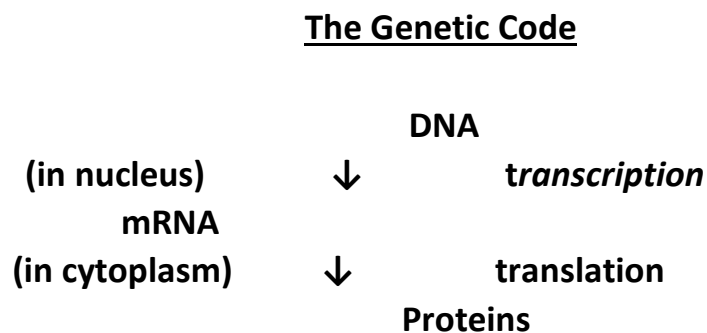## Genetic code

One of the first questions about the genetic code to be addressed was: How many nucleotides are necessary to specify a single amino acid? This basic unit of the genetic code—the set of bases that encode a single amino acid—is a codon

The **genetic code** is the set of rules by which information encoded in genetic material (DNA or mRNA sequences) is translated into proteins (amino acid sequences) by living cells. The code defines a mapping between tri-nucleotide sequences, called **codons, and amino acids**

The Genetic Code

DNA
(in nucleus)   ↓   t*ranscription*
mRNA
(in cytoplasm)   ↓   **translation**
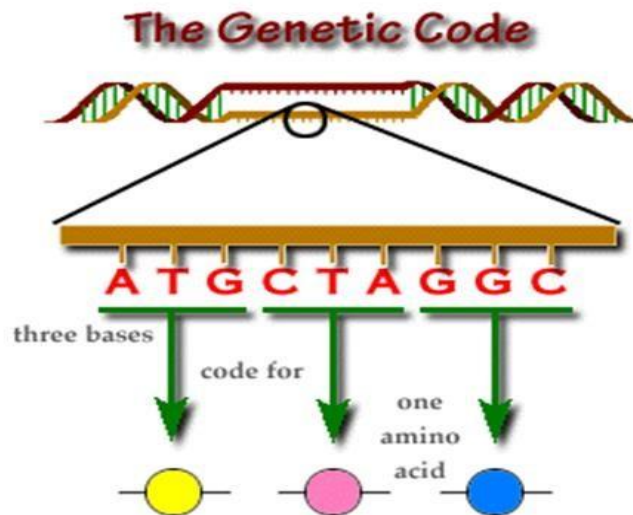**Proteins**

DNA provides the information that ultimately codes for a specific protein to be produced.

DNA must first be "transcribed" into mRNA (messenger RNA).

Information for the genetic code is read as a series of three consecutive bases or **codons.**

Each **codon** ultimately corresponds to a specific amino acid that will be added to a growing polypeptide chain

Cracking the Code

Crick reasoned that:
The 4 different nucleotides in DNA (G, C, A, T) must code for 20 amino acids.
So if:
1 nucleotide coded for 1 amino acid
- we would only code for 4 amino acids
2 nucleotides coded for 1 amino acid
- we would get $4^2 = 16$ amino acids
3 nucleotides coded for 1 amino acid
- we would get $4^3 = 64$ combinations
- this would include codes for stops and starts
- more than one three letter code can be used for each amino
      acid (code is redundant)

There would be 4 4 16 possible codons—still not enough to

code for all 20 amino acids. With three nucleotides per codon, there are 4 4 4 64 possible

codons, which is more than enough to specify 20 different amino acids. Therefore, a triplet code

requiring three nucleotides per codon is the most efficient way to encode all 20 amino acids.

Some of the most important properties of genetic codes are as follows:

The genetic code consists of a sequence of nucleotides in DNA or RNA. There are four letters in the code, corresponding to the four bases—A, G, C, and U (T in DNA).
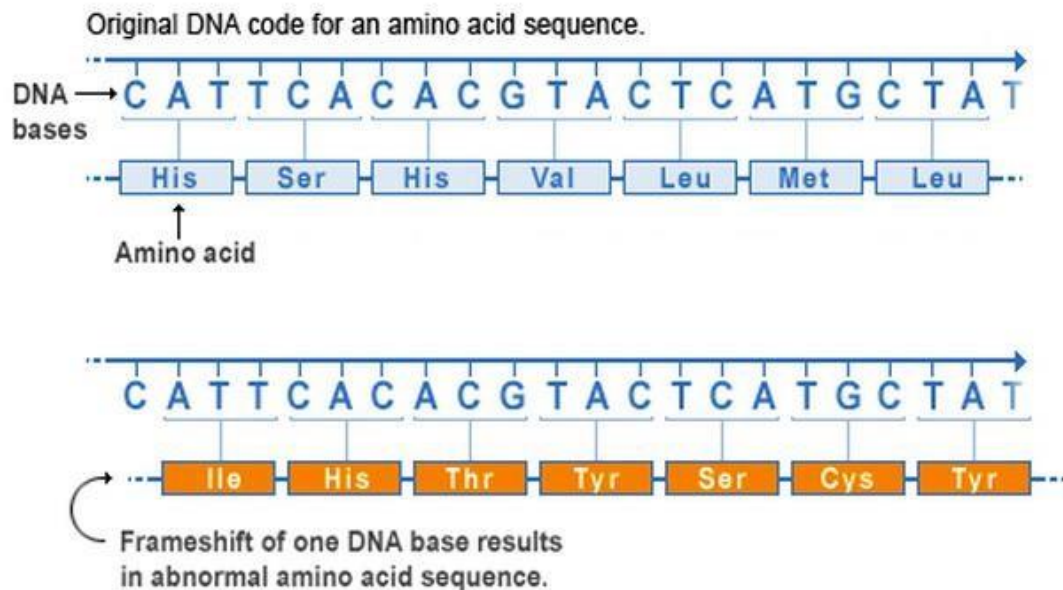
1. The code is a triplet codon:
The nucleotides of mRNA are arranged as a linear sequence of codons, each codon consisting of three successive nitrogenous bases, i.e., the code is a triplet codon. . Each amino acid is encoded by a sequence of three consecutive nucleotides, called a codon.

2- Frameshift read:

Evidently, the genetic message once initiated at a fixed point is read in a definite frame in a series of three letter words. The framework would be disturbed as soon as there is a deletion or addition of one or more bases.

When such frame shift mutations were intercrossed, then in certain combinations they produce wild type normal gene. It was concluded that one of them was deletion and the other an addition, so that the disturbed order of the frame due to mutation will be restored by the other

Original DNA code for an amino acid sequence.

DNA bases → C A T T C A C A C G T A C T C A T G C T A T

His — Ser — His — Val — Leu — Met — Leu

↑ Amino acid

C A T T C A C A C G T A C T C A T G C T A T

Ile — His — Thr — Tyr — Ser — Cys — Tyr

Frameshift of one DNA base results in abnormal amino acid sequence.

U.S. National Library of Medicine

2. The code is non-overlapping:

In translating mRNA molecules the codons do not overlap but are "read" sequentially (Fig. 38.27). Thus, a non-overlapping code means that a base in a mRNA is not used for different codons, it has been shown that an overlapping code can mean coding for four amino acids from six bases.

However, in actual practice six bases code for not more than two amino acids. For example, in case of an overlapping code, a single change (of substitution type) in the base sequence will be reflected in substitutions of more than one amino acid in corresponding protein. Many examples have accumulated since 1956 in which a single base substitution results into a single amino acid change in insulin, tryptophan synthelase, , haemoglobin, et

How is the reading frame read?
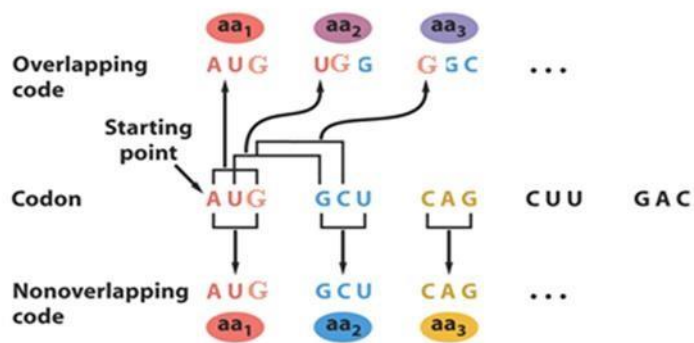
c.

4. The code is non-ambiguous:

Non-ambiguous code means that a particular codon will always code for the same amino acid. In case of ambiguous code, the same codon could have different meanings or in other words, the same codon could code two or more than two different amino acids. Generally, as a rule, the same codon shall never code for two different amino acids.

However, there are some reported exceptions to this rule: the codons AUG and GUG both may code for methionine as initiating or starting codon, although GUG is meant for valine. Likewise, GGA codon codes for two amino acids glycine and glutamic acid.

5. The code has polarity:

The code is always read in a fixed direction, i.e., in the 5′→3′ direction. In other words, the codon has a polarity. It is apparent that if the code is read in opposite directions, it would specify two different proteins, since the codon would have reversed base sequence:

Codon :	UUG	AUC	GUC	UCG	CCA	ACA	AGG
Polypeptide:→Leu		Ile	Val	Ser	Prof	Thr	Arg
	Val	Leu	Leu	Ala	Thr	Thr	Gly←

6. The code is degenerate:

More than one codon may specify the same amino acid; this is called degeneracy of the code. For example, except for tryptophan and methionine, which have a single codon each, all other 18 amino acids have more than one codon. Thus, nine amino acids, namely phenylalanine, tyrosine, histidine, glutamine, asparagine, lysine, aspartic acid, glutamic acid and cysteine, have two codons each. Isoleucine has three codons. Five amino acids, namely valine, proline, threonine, alanine and glycine, have four codons each. Three amino acids, namely leucine, arginine and serine, have six codons each .

The code degeneracy is basically of two types: partial and complete. Partial degeneracy occurs when first two nucleotides are identical but the third (i.e., 3′ base) nucleotide of the degenerate codons differs, e.g., CUU and CUC code for leucine, Complete degeneracy occurs when any of the four bases can take third position and still code for the same amino acid (e.g., UCU, UCC, UCA and UCG code for serine).

7. Some codes act as start codons:

In most organisms, AUG codon is the start or initiation codon, i.e., the polypeptide chain starts either with methionine (eukaryotes) or N- formylmethionine (prokaryotes). Methionyl or N-formylmethionyl-tRNA specifically binds to the initiation site of mRNA containing the AUG initiation codon. In rare cases, GUG also serves as the initiation codon, e.g., bacterial protein synthesis. Normally, GUG codes for valine, but when normal AUG codon is lost by deletion, only then GUG is used as initiation codon.

8. Some codes act as stop codons:

Three codons UAG, UAA and UGA are the chain stop or termination codons. They do not code

for any of the amino acids. These codons are not read by any tRNA molecules (via their

anticodons), but are read by some specific proteins, called release factors (e.g., RF-1, RF-2, RF-

3 in prokaryotes and RF in eukaryotes). These codons are also called nonsense codons, since

they do not specify any amino acid.

Genetic Code: Where to begin and stop?

- Synthesis of all polypeptide chains in prokaryotic and eukaryotic
cells begins with the amino acid **methionine**.
   -In most mRNAs, the **start (_initiator_) codon** specifying this
   amino-terminal methionine is AUG.
   -Exceptions: GUG in prokaryotes and CUG in eukaryotes
- The three codons UAA, UGA, and UAG do not specify amino acids
but constitute **stop (_termination_) codons** that mark the carboxyl
terminus of polypeptide chains.
- The sequence of codons that runs from a specific start codon to a stop
codon is called a **_reading frame_**.
   -This precise linear array of ribonucleotides in groups of three specifies
   the precise linear sequence of amino acids in a polypeptide chain.
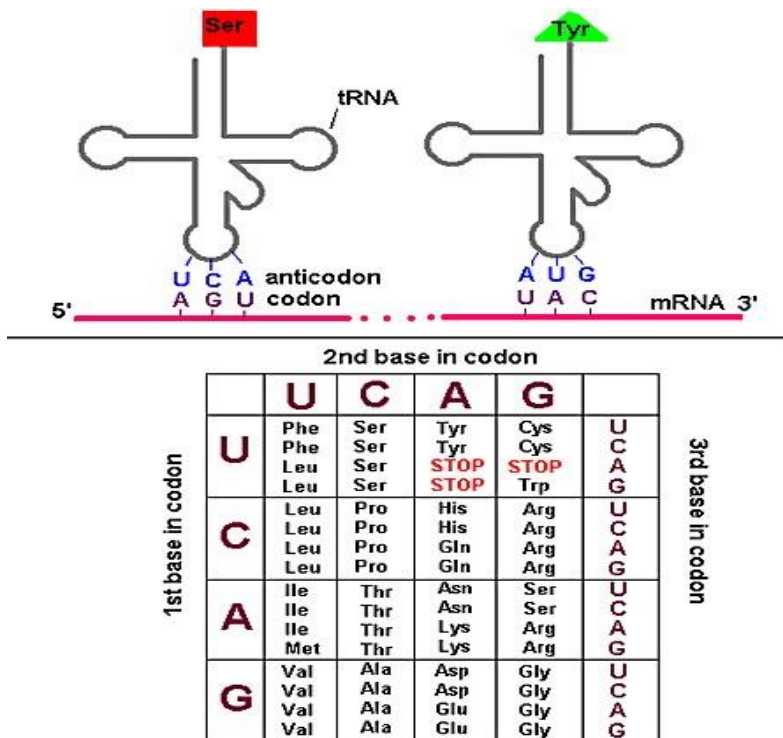
Reading frame
UCAUGGGCACCGGAGUGAAACGCUGAGGGUG
↓
Functional Protein
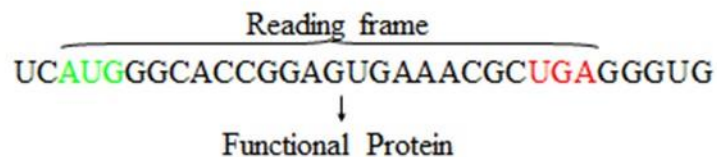
9. The code is universal:

Same genetic code is found valid for all organisms ranging from bacteria to man.

Such universality of the code was found that E. coli (Bacterium), Xenopus laevis

(Amphibian) and guinea pig (mammal) amino acyl-tRNA use almost the same code

10. **Isoaccepting** tRNAs are tRNAs with different anticodons that accept the same amino acid; wobble allows the anticodon on one type of tRNA to pair with   more than one type of codon on mRNA.

The Genetic Code