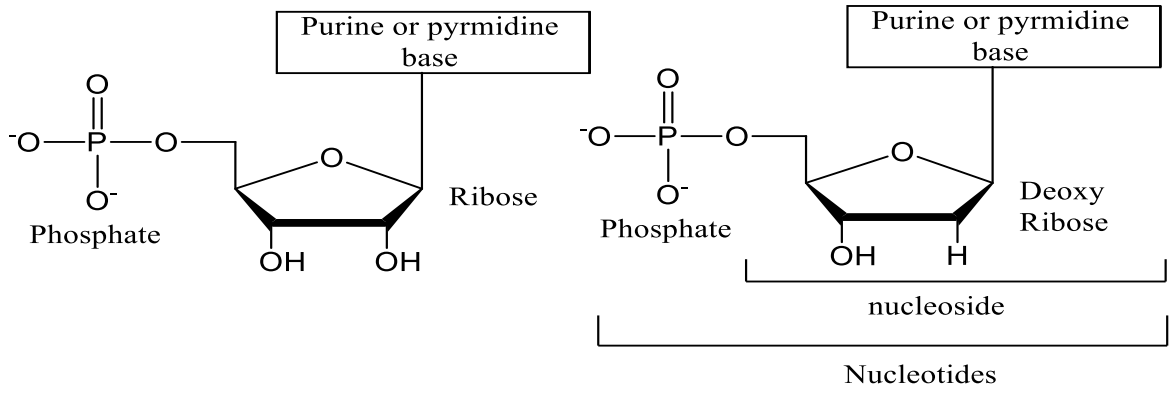
Nucleic Acid

Nucleic acids are large biomolecules that play essential roles in all cells and viruses. A major function of nucleic acids involves the storage and expression of genomic information. Deoxyribonucleic acid, or DNA, encodes the information cells need to make proteins. A related type of nucleic acid, called ribonucleic acid (RNA), comes in different molecular forms that play multiple cellular roles, including protein synthesis.

Nucleic acids are made of nitrogen-containing bases, phosphate groups, and sugar molecules. Each type of nucleic acid has a distinctive structure and plays a different role in our cells.

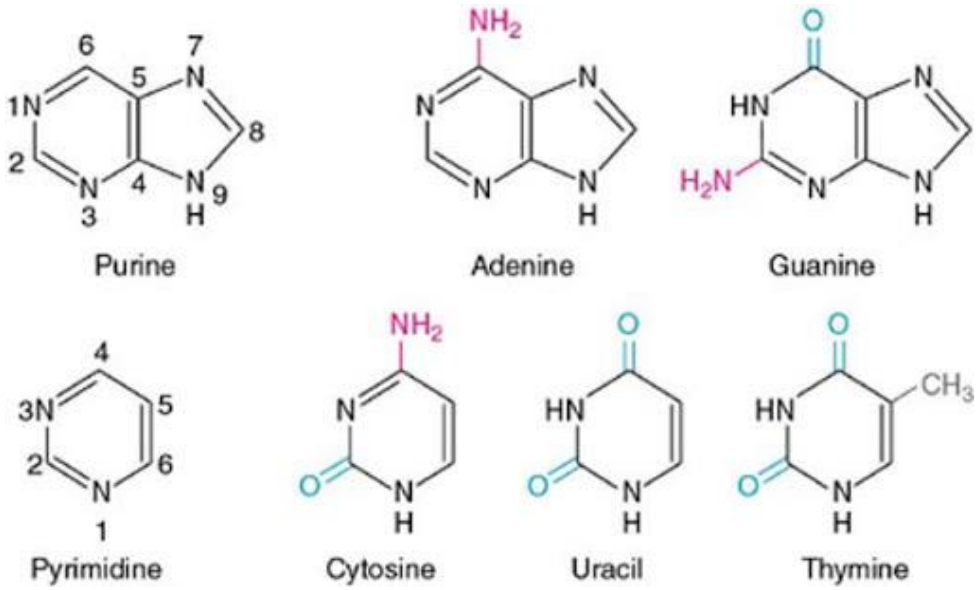
Nucleotides and Nucleoside

Nucleotides have three characteristic components: (1) a nitrogenous (nitrogen-containing) base, (2) a pentose, and (3) a phosphate (Figure 1). The molecule without the phosphate group is called a **nucleoside**. The nitrogenous bases are derivatives of two parent compounds, pyrimidine and purine. The bases and pentoses of the common nucleotides are heterocyclic compounds. The carbon and nitrogen atoms in the parent structures are conventionally numbered to facilitate the naming and identification of the many derivative compounds (see Fig. 1).



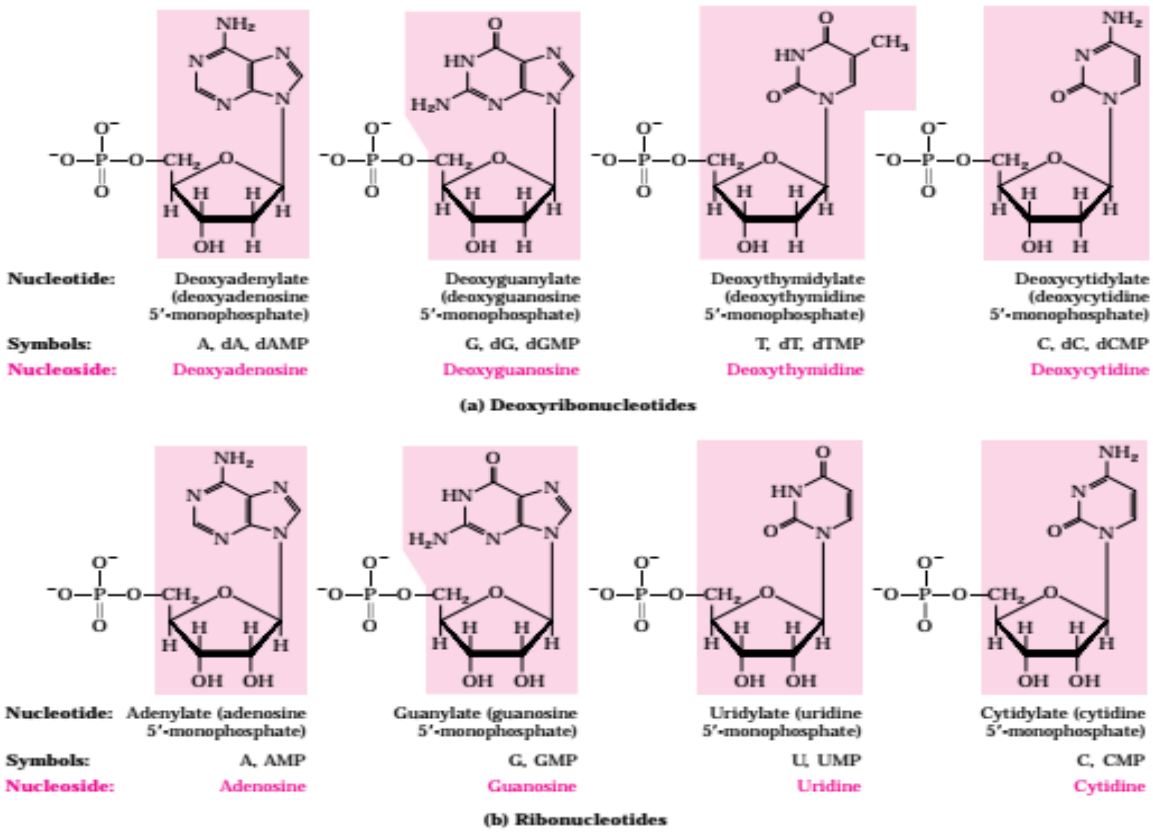
(Figure 1)

Both DNA and RNA contain two major purine bases, adenine (A) and guanine (G), and two major pyrimidines. In both DNA and RNA one of the pyrimidines is cytosine (C), but the second major pyrimidine is not the same in both: it is thymine (T) in DNA and uracil (U) in RNA. Only rarely does thymine occur in RNA or uracil in DNA. The structures of the five major bases are shown in (Figure 2)



(Figure 2)

The base of a nucleotide is joined covalently (at N-1 of pyrimidines and N-9 of purines) in an N-β-glycosyl bond to the 1 carbon of the pentose, and the phosphate is esterified to the 5 carbon. The N-β-glycosyl bond is formed by removal of the elements of water (a hydroxyl group from the pentose and hydrogen from the base), as in O-glycosidic bond formation (see Fig. 3).



(Figure 3)

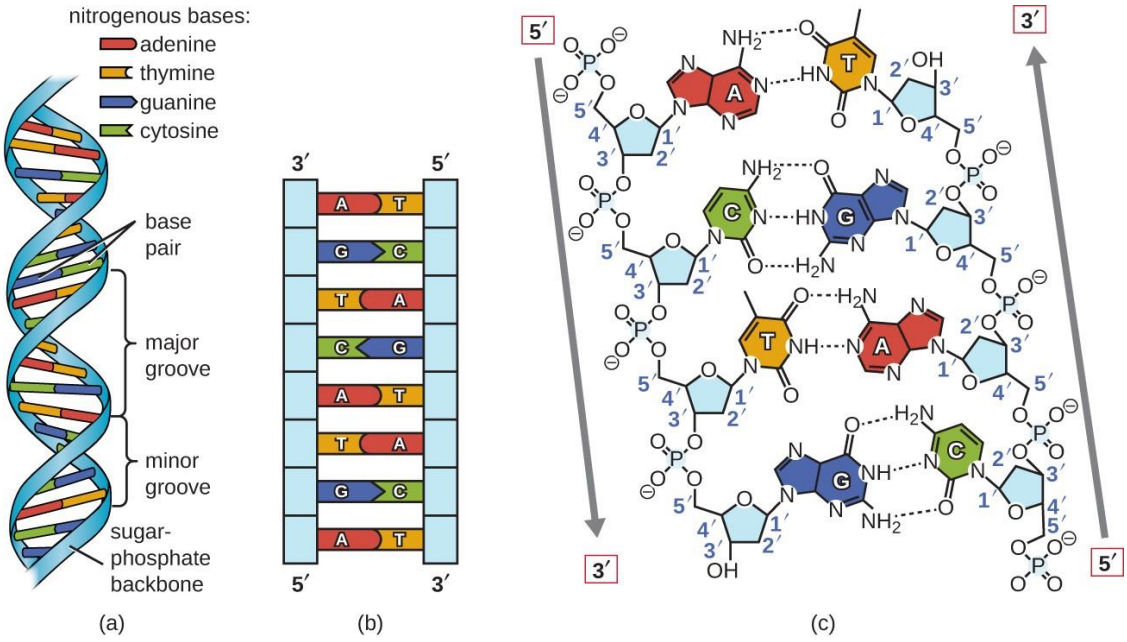
Deoxyribonucleic acid (DNA)

In 1953 James D. Watson and Francis H.C. Crick proposed a three-dimensional structure for DNA based on low-resolution X-ray crystallographic data and on Erwin Chargaff's observation that, in naturally occurring DNA, the amount of T equals the amount of A and the amount of G equals the amount of C. Watson and Crick, who shared a Nobel Prize in 1962 for their efforts, postulated that two strands of polynucleotides coil around each other, forming a double helix. The two strands, though identical, run in opposite directions as determined by the orientation of the 5′ to 3′ phosphodiester bond. The sugar-phosphate chains run along the outside of the helix, and the bases lie on the inside, where they are linked to complementary bases on the other strand through hydrogen bonds.

DNA is a molecule composed of two polynucleotide chains that coil around each other to form a double helix carrying genetic instructions for the development, functioning, growth and reproduction of all known organisms and many viruses.

The two DNA strands are known as polynucleotides as they are composed of simpler monomeric units called nucleotides.

Each nucleotide is composed of one of four nitrogen-containing nucleobases (cytosine [C], guanine [G], adenine [A] or thymine [T]), a sugar called deoxyribose, and a phosphate group. The nucleotides are joined to one another in a chain by covalent bonds (known as the phospho-diester linkage) between the sugar of one nucleotide and the phosphate of the next, resulting in an alternating sugar-phosphate backbone. The nitrogenous bases of the two separate polynucleotide strands are bound together, according to base pairing rules (A with T and C with G), with hydrogen bonds to make double-stranded DNA (Figure 4).



(Figure 4)

Both strands of double-stranded DNA store the same biological information. This information is replicated as and when the two strands separate. A large part of DNA (more than 98% for humans) is non-coding, meaning that these sections do not serve as patterns for protein sequences. The two strands of DNA run in opposite directions to each other and are thus antiparallel. Attached to each sugar is one of four types of nucleobases (informally, bases). It is the sequence of these four nucleobases along the backbone that encodes genetic information.

Ribonucleic acid (RNA)

Ribonucleic acid (RNA) is a molecule that is present in the majority of living organisms and viruses. It is made up of nucleotides, which are ribose sugars attached to nitrogenous bases and phosphate groups. The nitrogenous bases include adenine, guanine, uracil, and cytosine. RNA mostly exists in the single-stranded form, but there are special RNA viruses that are double-stranded. The RNA molecule can have a variety of lengths and structures. An RNA virus uses RNA instead of DNA as its genetic material and can cause many human diseases. Transcription is the process of RNA formation from DNA, and translation is the process of protein synthesis from RNA. The means of RNA synthesis and the way that it functions differs between eukaryotes and prokaryotes. Specific RNA molecules also regulate gene expression and have the potential to serve as therapeutic agents in human diseases.

Three main types of RNA are involved in protein synthesis. They are messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA).

mRNA

mRNA is transcribed from DNA and contains the genetic blueprint to make proteins. Prokaryotic mRNA does not need to be processed and can proceed to synthesize proteins immediately. In eukaryotes, a freshly transcribed RNA transcript is considered a pre-mRNA and needs to undergo maturation to

form mRNA. A pre-mRNA contains non-coding and coding regions known as introns and exons, respectively. During pre-mRNA processing, the introns are spliced, and the exons are joined together to produce matured RNA.

tRNA

tRNAs are RNA molecules that translate mRNA into proteins. The primary function of a tRNA is to carry amino acids on its 3' acceptor site to a ribosome complex with the help of aminoacyl-tRNA synthetase. Aminoacyl-tRNA synthetases are enzymes that load the appropriate amino acid onto a free tRNA to synthesize proteins. The type of amino acid on a tRNA is dependent on the mRNA codon, which is a sequence of three nucleotides that codes for an amino acid. The anticodon arm of the tRNA is the site of the anticodon, which is complementary to an mRNA codon and dictates which amino acid to carry.

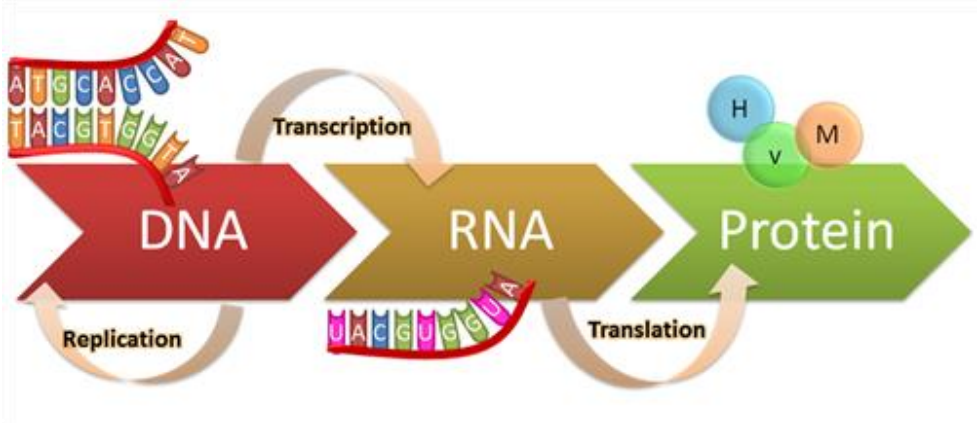
rRNA

rRNA forms ribosomes, which are essential in protein synthesis. A ribosome contains a large and small ribosomal subunit. In prokaryotes, a small 30S and large 50S ribosomal subunit make up a 70S ribosome. In eukaryotes, the 40S and 60S subunit form an 80S ribosome. The ribosomes contain an exit (E), peptidyl (P), and acceptor (A) site to bind aminoacyl-tRNAs and link amino acids together to create polypeptides.

Protein synthesis

It consists of two processes **transcription** and **translation**. In eukaryotic cells, transcription takes place in the nucleus. During transcription, DNA is used as a template to make a molecule of messenger RNA (mRNA). The molecule of mRNA then leaves the nucleus and goes to a ribosome in the cytoplasm, where translation occurs. During translation, the genetic code in mRNA is read and used to make a polypeptide. These two processes are summed up by the central dogma of molecular biology: DNA → RNA → Protein.

RNA strands are created using DNA strands as a template in a process called transcription, where DNA bases are exchanged for their corresponding bases except in the case of thymine (T), for which RNA substitutes uracil (U). Under the genetic code, these RNA strands specify the sequence of amino acids within proteins in a process called translation.

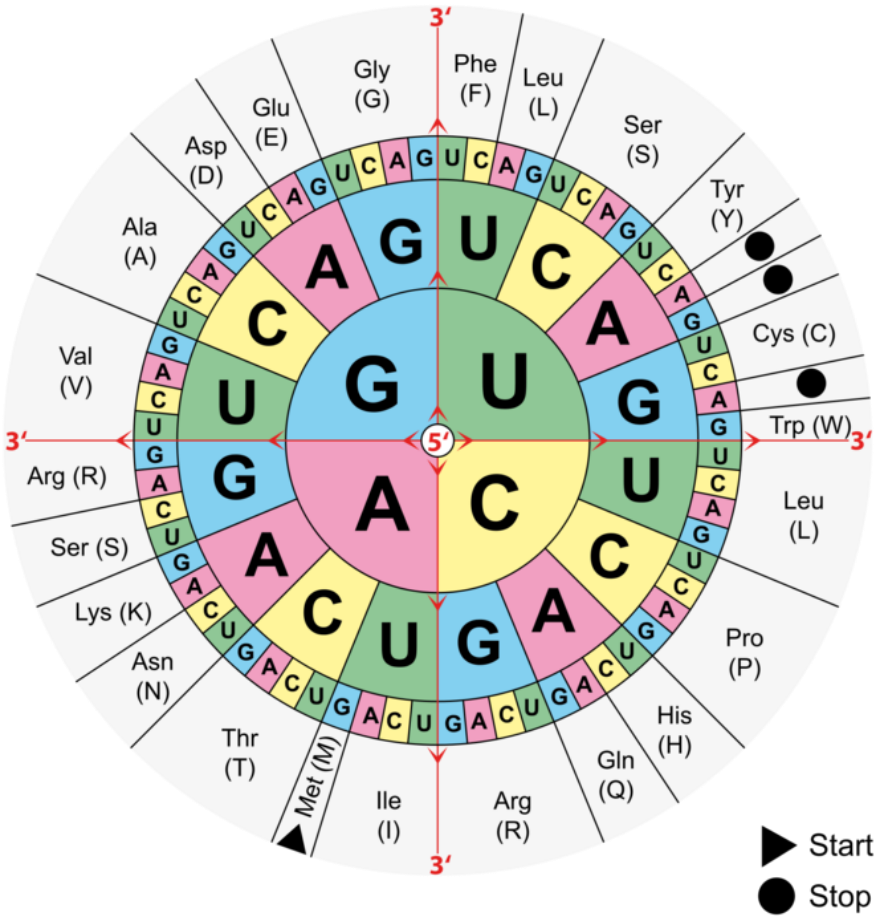


RNAs have a broader range of functions, and several classes are found in cells. Ribosomal RNAs (rRNAs) are components of ribosomes, the complexes that carry out the synthesis of proteins. Messenger RNAs (mRNAs) are intermediaries, carrying genetic information from one or a few genes to a ribosome, where the corresponding proteins can be synthesized. Transfer RNAs (tRNAs) are adapter molecules that faithfully translate the information in mRNA into a specific sequence of amino acids.

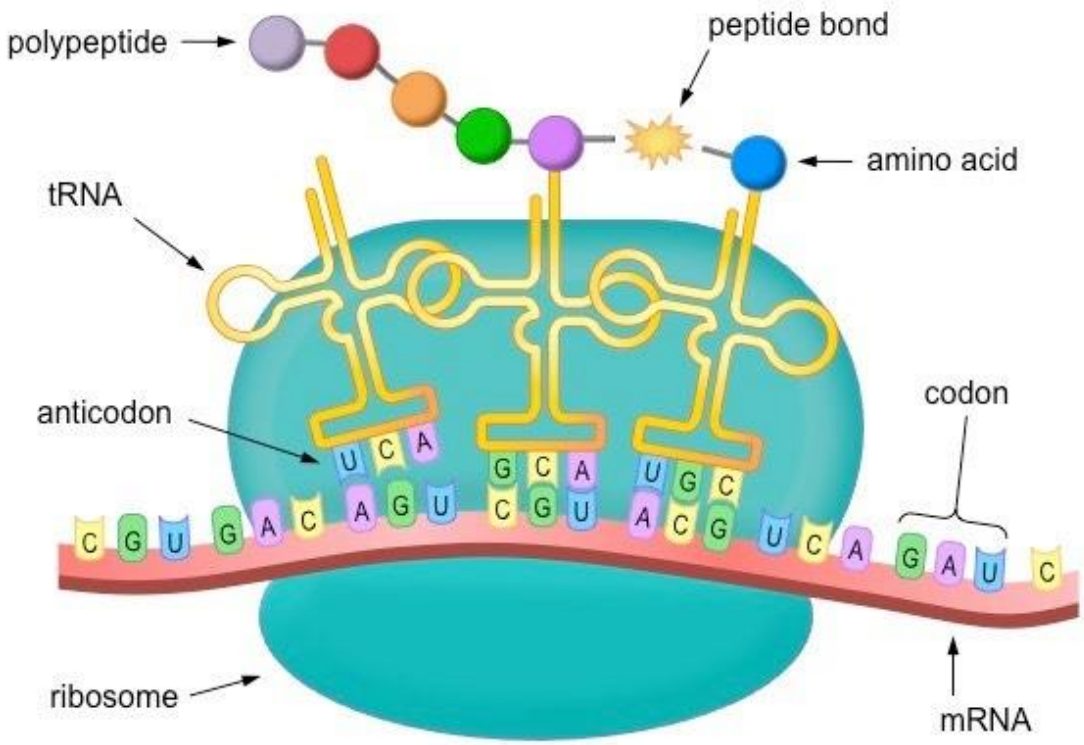
The genetic code is the set of rules used by living cells to translate information encoded within genetic material (DNA or mRNA sequences of nucleotide triplets, or codons) into proteins. Translation is accomplished by the ribosome, which links proteinogenic amino acids in an order specified by messenger RNA (mRNA), using transfer RNA (tRNA) molecules to carry amino acids and to read the mRNA three nucleotides at a time. The genetic code is highly similar among all organisms and can be expressed in a simple table with 64 entries.

Codon is a sequence of three nucleotides on an mRNA strand that encodes a specific amino acid.

tRNA is a type of RNA that is folded into a specific three-dimensional structure. It carries and transfers an amino acid to the polypeptide chain that the ribosome is assembling. One end of the tRNA contains an **anticodon**, a sequence of three nucleotides that is complementary to the three nucleotides in the corresponding codon on the mRNA. Each anticodon is specific to one and only one codon.



The process of protein synthesis explained in figure 5.



(Figure 5)

Types of mutations

There are many different ways that DNA can be changed, resulting in different types of mutation. Here is a quick summary of a few of these:

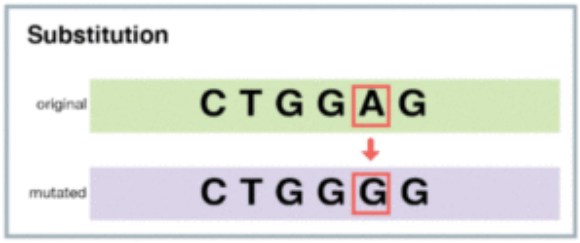
1-Substitution

A substitution is a mutation that exchanges one base for another (i.e., a change in a single "chemical letter" such as switching an A to a G). Such a substitution could:

1. Change a codon to one that encodes a different amino acid and cause a small change in the protein produced. For example, **sickle cell anemia** is caused

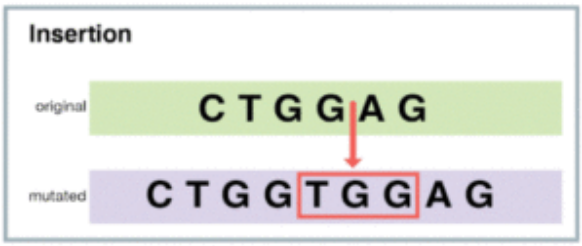
by a substitution in the beta-hemoglobin gene, which alters a single amino acid in the protein produced.

2. Change a codon to one that encodes the same amino acid and causes no change in the protein produced. These are called **Silent mutations**.
3. Change an amino-acid-coding codon to a single "stop" codon and cause an incomplete protein. This can have serious effects since the incomplete protein probably won't function.



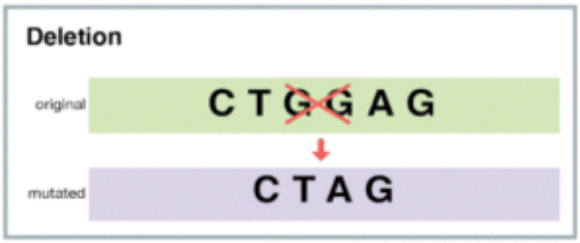
2- Insertion

Insertions are mutations in which extra base pairs are inserted into a new place in the DNA.



3- Deletion

Deletions are mutations in which a section of DNA is lost, or deleted.



4-Frameshift

Since protein-coding DNA is divided into codons three bases long, insertions and deletions can alter a gene so that its message is no longer correctly parsed. These changes are called frame shifts.