**Course: Advanced Bioinformatics**

**Module title: DNA and parts**

**Module no. : 7**

It is considered that life on this planet i.e., Earth has evolved from a common root for all organisms. The ancestor is supposed to be a singleton (single) cell organism or unicellular organism. Later on, a diversity of organisms has been produced due to the distinctions in the genetic characteristics of organisms. Now a day, more than 100 million living species exist in this world (Earth). Cell is a basic unit or tenet of all living organisms either these contains micro-molecules or macromolecules among them. Cell shares and carries all distinctiveness of life, which are necessary for a living organism. Some of the organisms consist of a single cell called prokaryotes i.e., bacteria, archaea, and protists. Most of the organisms are composed of multi-cells, these are known as eukaryotes i.e., fungi, human beings and animals. Cells play an important part in the lives of all these living organisms. Our bodies are composed of more than 10¹³ cells and these cells vary from each other in their shapes, functions, and sizes. The components of a cell are also known as intra-cellular components or micro-molecules. Some of the known processes carry out cell division, mitosis/meiosis in all kinds of living organisms. These all kinds of living cells, work as a computer taps, which carry the genetic information among their materials, in form of chromosomes and genes. Mainly, there are two types of cells, known as eukaryotic cells and prokaryotic cells. Eukaryotic cells consist of nucleus, which store deoxyribonucleic acid (DNA) among them. Nuclear membrane is used to separate out the nucleus from other antibodies of eukaryotic cell. Therefore, nucleus is called the core element of eukaryotic cell. Prokaryotic cell does not contain any kind of nucleus inside it and it stores deoxyribonucleic acid in its cytoplasm. The nucleus, its parts and all components of a cell are self-possessed of de-oxyribonucleic acid (DNA), ribonucleic acid (RNA) and an assortment of protein structures.

De-oxyribonucleic Acid (DNA)

Deoxyribonucleic acid (DNA) acts as a computer tape or store house that can keep the information for a longer time. It (DNA) monitors, controls and supervises, the complete cell activities within the bodies of molecular organisms. When we study its structure, single DNA molecule looks like a double stranded helical structure in shape. These helical strands are considered to be moving in the opposite direction to each other; so they form an anti-parallel shape. Each single strand of DNA molecule is composed of large and long series of polynucleotide. The nucleotides held closer to each other are anti-parallel and form a long chain of DNA molecule. If one strand of nucleotides represents 5' to 3' in direction, then the other strand will be 3' to 5' in direction on the same strand. These strands coil around each other forming a helical structure which looks like that it has been derived from any of the single strand of DNA molecule. Each nucleotide strand is said to be an organic compound which is further composed of three basic groups named deoxyribose (sugar), phosphate and nitrogen bases. Sugar and phosphate groups are common among all of these nucleotides having different nitrogen bases. Sugar is just like a pentose-ring structure that has been made up of various atoms. In a single molecule, sugar joins oxygen at 1I position and carbon atoms at 2 to 5 positions in a clockwise direction. A diversity of different molecules of phosphate group is attached to the carbon atoms at different location. The nitrogen bases are commonly known as simple bases. Four different types of nitrogen bases are available in every nucleotide. These bases include adenine (A), cytosine (C), guanine (G), and thymine (T) and they are represented by a single letter i.e., A, C, G, and T. The nucleotide strands are linked to eachother with hydrogen bonding among them. Adenine (A) and Thymine (T) bases compliment eachother forming a double covalent bond which is two-faced. Cytosine (C) and Guanine (G) form a triple covalent bond on the opposite strands. The A+T base pairing is called purine and C+G base pairing called purimidines. Due to this triple hydrogen bonding, purimidines is more stable than purines.

The DNA molecule is composed of coding regions and non-coding regions/spacer molecules. The coding regions make 3% to 5% part of the total DNA. These regions contain a large number of genes and templates, which produce protein structures as an output and play an important part during different biological process, i.e., replication, transcription and translation processes in wet labs. The non-coding region constitutes 95% to 97% of DNA molecule. These non-coding regions do not play any active part during the replication, transcription, and translation process. However, these regions provide different kinds of information, which act as punctuations for replication, transcription and translation processes. Coding regions are composed of chromaton protein and chromosomes. There are 23 (twenty-three) pairs of chromosomes. Among them, 22 pairs are known as autosomes. The 23rd pair named sex chromosome consists of X chromosome and Y chromosome, in male. In female, both are X chromosomes. A single chromosome contains hundreds to thousands of genes.

De-oxyribonucleic acid (DNA) performs two types of processes, i.e., replication and transcription. In replication process, the parental DNA divides itself into two double helical daughter strands through meiosis/mitosis process, keeping same information in daughter molecules. Replication is cyclic and dynamic process, in which every daughter double strand molecule, is further divided into two daughter double strand molecules. During replication of DNA, only the sequences of nitrogenous bases get copied. During transcription process, DNA is converted into RNA polymerase or messenger ribonucleic acid (mRNA), ribosomal ribonucleic acid (rRNA) and transfer ribonucleic acid (tRNA). RNA polymerase itself cannot initiate the transcription process, however, the transcription factor in the promoter region of gene directs to RNA polymerase to start this process. In translation, each mRNA molecule is converted into protein structure. Protein structures are coded through linear sequence of amino acids. The combination of three nucleotides represents the sequence of amino acids. The triplet of nucleotides is called codon

Ribonucleic Acid (RNA)

Ribonucleic acid (RNA) is another organic molecule, which is similar to DNA molecule but different in its internal details like structure and shape to some extent. RNA is a polymer of ribo-nucleotides; its molecule has a single strand, which folds itself at the back forming a double helical structure. Most of the RNAs (tRNA, rRNA, mRNA) are found in cytoplasm, whereas few mRNAs exist in the nucleus. The rRNA has been used for protein synthesis. RNA consists of a long chain of ribo nucleotides and each ribo nucleotide is composed of ribose sugar, phosphate, and nitrogen bases similar to DNA. The phosphate molecule in RNA is similar to DNA phosphate molecule. Ribose molecule of RNA is different from DNA’s molecule, which contains deoxyribose. Ribose contains one more oxygen atom at 2' position in RNA. RNA molecule also consists of four different types of nitrogen bases like DNA molecule. The four types of bases called, adenine (A), cytosine (C), guanine (G) and uracil (U). In RNA, thymine (T) replaces a nitrogen base uracil (U). There also exists a base pairing between complement bases in RNA molecule. Double covalent bond between A and U, and triple covalent is formed between C and G. RNA molecule is unstable as compared to DNA molecule and it plays an intermediate role for the conversion of genetic information from DNA molecule to protein structures. Therefore, it is highly important to understand its internal details for transformation.

Majorly, three different types of RNA’s exist, which are known as, mRNA, tRNA, and rRNA. They perform different functions during transcription process in bodies of organisms. Transcription process is enveloped (from DNA to RNA) by three small processes, called initiation, elongation, and termination. The mRNA acts as message transporter from nucleus to the sites, where actually protein is being synthesize. It consists of single strand, which is complement to single strand of DNA molecule. Three consecutively joint mRNA make one codon, which further define the arranged sequences of amino acids. The tRNA carries amino acids for ribosome. It also binds one amino acid to one mRNA molecule. It forms the anti-codon by joining three amino acids in a specific sequence. The rRNA is the acting component of RNA that remains functional during the production of protein structures.

Protein Structures

Protein comes from the Greek word protoas, which means primary importance. Protein structures are highly weighing organic compounds which are made up by extensive linking of amino acids. Proteins play an important role in our lives. Proteins perform different kinds of tasks in human bodies. Some proteins act as an enzyme that acts as a catalyst and triggers biochemical functions in our bodies. Some proteins are hard which make tendons, muscles, nails, bones and fibrous that makes connecting tissues in our bodies. In short, digestion of food in small intestine, carrying oxygen into the blood, making epithelial cell of our skin etc are only possible due to proteins. There are four types of protein structures which have been classified on the basis of poly peptide chains. These include primary, secondary, tertiary and quaternary structures. These protein structures have different shapes due the different arrangements of sequence of amino acids. Some of them make are linear, coil, helix, parallel sheet, folds, loops, and helices.

Protein structures made up of sequence of amino acids. These amino acids have been coded from genes. The basic function of Protein structure is synthesis, which is due to the formation of a peptide bond between two amino acids. This reaction is done repeatedly and as a result forming a long chain of polypeptides. There are commonly twenty (20) different types of amino acids in a protein structure and different combinations are possible in a codon. Each amino acid is coded by one to six codons. When a ribosome reaches one of the three condones for which there is no matching tRNA, the ribosome fall off and synthesized protein is released.