

# Replication.. Transcription.. Translation

## THE DNA

DNA acts as the genetic material in most of the organisms. RNA though it also acts as a genetic material in some viruses, mostly functions as a messenger. RNA has additional roles as well. It functions as adapter, structural, and in some cases as a catalytic molecule. we are going to discuss the structure of DNA, its replication, the process of making RNA from DNA (transcription), the genetic code that determines the sequences of amino acids in proteins, the process of protein synthesis (translation) and elementary basis of their regulation. The determination of complete nucleotide sequence of human genome during last decade has set in a new era of genomics. In the last section, the essentials of human genome sequencing and its consequences will also be discussed.

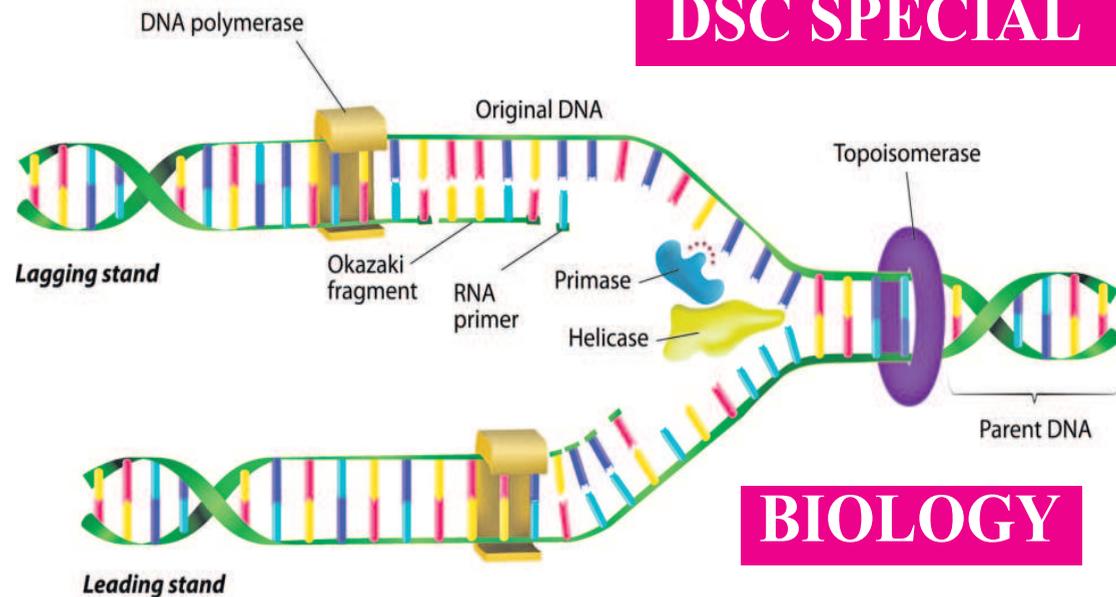
- Let us begin our discussion by first understanding the structure of the most interesting molecule in the living system, that is, the DNA. In subsequent sections, we will understand that why it is the most abundant genetic material, and what its relationship is with RNA.

## THE DNA

- DNA is a long polymer of deoxyribonucleotides. The length of DNA is usually defined as number of nucleotides (or a pair of nucleotide referred to as base pairs) present in it. This also is

### The salient features of the Double-helix structure of DNA

- It is made of two polynucleotide chains, where the backbone is constituted by sugar-phosphate, and the bases project inside.
- The two chains have anti-parallel polarity. It means, if one chain has the polarity 5'→3', the other has 3'→5'.
- The bases in two strands are paired through hydrogen bond (H-bonds) forming base pairs (bp). Adenine forms two hydrogen bonds with Thymine from opposite strand and vice-versa.
- Similarly, Guanine is bonded with Cytosine with three H-bonds.
- As a result, always a purine comes opposite to a pyrimidine. This generates approximately



## DSC SPECIAL

## BIOLOGY

the characteristic of an organism.

- For example, a bacteriophage known as  $\phi \times 174$  has 5386 nucleotides, Bacteriophage lambda has 48502 base pairs (bp), Escherichia coli has  $4.6 \times 10^6$  bp, and haploid content of human DNA is  $3.3 \times 10^9$  bp. Let us discuss the structure of such a long polymer.

### Structure of Polynucleotide Chain

- Let us recapitulate the chemical structure of a polynucleotide chain (DNA or RNA). A nucleotide has three components – a nitrogenous base, a pentose sugar (ribose in case of RNA, and deoxyribose for DNA), and a phosphate group. There are two types of nitrogenous bases – Purines (Adenine and Guanine), and Pyrimidines (Cytosine,

- Uracil and Thymine). Cytosine is common for both DNA and RNA and Thymine is present in DNA. Uracil is present in RNA at the place of Thymine. A nitrogenous base is linked to the pentose sugar through a N-glycosidic linkage to form a nucleoside, such as adenosine or deoxyadenosine, guanosine or deoxyguanosine, cytidine or deoxycytidine and uridine or deoxythymidine.
- When a phosphate group is linked to 5'-OH of a nucleoside through phosphoester linkage, a corresponding nucleotide (or deoxynucleotide depending upon the type of sugar present) is formed. Two nucleotides are linked through 3'-5' phosphodiester linkage to form a dinucleotide.
- More nucleotides can be joined in such a manner to form a

- polynucleotide chain. A polymer thus formed has at one end a free phosphate moiety at 5'-end of ribose sugar, which is referred to as 5'-end of polynucleotide chain. Similarly, at the other end of the polymer the ribose has a free 3'-OH group which is referred to as 3'-end of the polynucleotide chain.
- The backbone in a polynucleotide chain is formed due to sugar and phosphates. The nitrogenous bases linked to sugar moiety project from the backbone. In RNA, every nucleotide residue has an additional -OH group present at 2'-position in the ribose. Also, in RNA the uracil is found at the place of thymine (5-methyl uracil, another chemical name for thymine).
- DNA as an acidic substance

present in nucleus was first identified by Friedrich Meischer in 1869. He named it as 'Nuclein'. However, due to technical limitation in isolating such a long polymer intact, the elucidation of structure of DNA remained elusive for a very long period of time. It was only in 1953 that James Watson and Francis Crick, based on the X-ray diffraction data produced by Maurice Wilkins and Rosalind Franklin, proposed a very simple but famous Double Helix model for the structure of DNA. One of the hallmarks of their proposition was base pairing between the two strands of polynucleotide chains. However, this proposition was also based on the observation of Erwin Chargaff that for a double stranded DNA, the ratios between Adenine and Thymine and Guanine and Cytosine are constant and equals one.

- The base pairing confers a very unique property to the polynucleotide chains. They are said to be complementary to each other, and therefore if the sequence of bases in one strand is known then the sequence in other strand can be predicted. Also, if each strand from a DNA (let us call it as a parental DNA) acts as a template for synthesis of a new strand, the two double stranded DNA (let us call them as daughter DNA) thus, produced would be identical to the parental DNA molecule. Because of this, the genetic implications of the structure of DNA became very clear.

uniform distance between the two strands of the helix.

- The two chains are coiled in a right-handed fashion. The pitch of the helix is 3.4 nm (a nanometre is one billionth of a metre, that is  $10^{-9}$  m) and there are roughly 10 bp in each turn. Consequently, the distance between a bp in a helix is approximately equal to 0.34 nm.
- The plane of one base pair stacks over the other in double helix. This, in addition to H-bonds, confers stability of the helical structure. The proposition of a double helix structure for DNA and its simplicity in explaining the genetic implication became revolutionary. Very soon, Francis Crick proposed the Central dogma in molecular

biology, which states that the genetic information flows from DNA→RNA→Protein.

### Packaging of DNA Helix

- Taken the distance between two consecutive base pairs as 0.34 nm ( $0.34 \times 10^{-9}$  m), if the length of DNA double helix in a typical mammalian cell is calculated (simply by multiplying the total number of bp with distance between two consecutive bp, that is,  $6.6 \times 10^9$  bp  $\times$   $0.34 \times 10^{-9}$  m/bp), it comes out to be approximately 2.2 metres. A length that is far greater than the dimension of a typical nucleus (approximately 10–6 m).
- If the length of E. coli DNA is 1.36 mm, can you calculate the number of base pairs in E.coli?

- In prokaryotes, such as, E. coli, though they do not have a defined nucleus, the DNA is not scattered throughout the cell. DNA (being negatively charged) is held with some proteins (that have positive charges) in a region termed as 'nucleoid'. The DNA in nucleoid is organised in large loops held by proteins.
- In eukaryotes, this organisation is much more complex. There is a set of positively charged, basic proteins called histones. A protein acquires charge depending upon the abundance of amino acid residues with charged side chains. Histones are rich in the basic amino acid residues lysines and arginines. Both the amino acid residues carry positive charges in their

side chains. Histones are organised to form a unit of eight molecules called as histone octamer. The negatively charged DNA is wrapped around the positively charged histone octamer to form a structure called nucleosome. A typical nucleosome contains 200 bp of DNA helix. Nucleosomes constitute the repeating unit of a structure in nucleus called chromatin, thread-like stained (coloured) bodies seen in nucleus. The nucleosomes in chromatin are seen as 'beads-on-string' structure when viewed under electron microscope (EM). Theoretically, how many such beads (nucleosomes) do you imagine are present in a mammalian cell?