Cell Biology: Cellular Organization, Division and Processes

Prof. Shikha Laloraya
Department of Biochemistry
Indian Institute of Science – Bangalore

Lecture - 03
DNA: The genetic material, Part - 1

Hello everyone. In today's lecture, we will discuss DNA, deoxyribonucleic acid that is the genetic material, which is present within the cells, and also how it is organized.

(refer time: 00:35)

Before DNA was discovered, it was observed that organisms possess certain traits that appeared to be inherited by the offspring according to certain quantifiable and predictable set of rules. There were two main paradigms in biology at that time and they were proposed by the great scientists Gregor Mendel and Charles Darwin, prior to the discovery of DNA. Based on breeding experiments done with pea plants and also other systems, Mendel proposed principles of heredity that are known as Mendel's laws. In 1865, 66, just to give you an idea about the timelines, and these findings were later promoted by a later scientist Bateson, whose book cover is shown here. So, Mendel's laws were basically, dominance, that is in offspring having determinants of two varieties of a trait, only one, that is a dominant trait is expressed; now we know there are exceptions also; Segregation: organisms had two determinants of a trait now known as alleles and they segregated from each other during gametogenesis. And independent assortment, that is the determinants of unrelated traits were inherited independently from each other. That is they were independently assorted into gametes during meiosis.

On the other hand, Darwin had proposed the theory of evolution by natural selection. According to this, variations were there in populations but only the fittest survive, ultimately resulting in evolution. So, it was in this backdrop that scientists wondered about the nature of the substance or the molecule that might be responsible for determining heritable traits.

(refer time: 02:45)

Based on the paradigms of inheritance and evolution, understood at that time, it seemed reasonable that the genetic material should provide a rationale for the faithful inheritance of traits that were observed, and also it should provide a rationale for the differences in traits or what is the cause of variations, that now we know arise from mutations. So as far as evolution is concerned, people wondered, how do these selectable variations arise? And there should be a possible mechanism for this.

At that time, by light microscopy, chromosomes were observed to appear in dividing cells as thread like structures and then these structures, they appear to be segregated equally during cell division between the daughter cells. So, these were very exciting observations and people wondered

whether, you know, they may have something to do with genetics, although it was not clear at that time. But chromosomes are also analyzed chemically and it was found that chromosomes had DNA, deoxyribonucleic acid, as well as proteins, as their components. So scientists were wondering, what is the genetic material? And of course, they knew that chromosomes had DNA and proteins. So, these were two of the chief contenders. And people knew a lot about proteins that they are made up of many building blocks, 20 amino acids, and that there are many different types of proteins. So, it could have seemed at that time that you know, that organisms have so many different qualities and traits and could there be some connection between proteins and the determination of those traits. And DNA, on the other hand, was also an important constituent of chromosomes. And it was known at the chemical level it was analyzed and it was known that DNA is composed of a sugar, phosphates and four bases.

(refer time: 05:05)

So, while people were wondering about this, in the 1940s there were some interesting observations related to a pathogenic bacterium, *Streptococcus pneumoniae*. It was observed that there were two strains of this bacterium, smooth, which was pathogenic, and rough, which was not. And these strains also of course, as the name suggests, differed from each other based on their appearance, smooth versus rough. And when they were injected into mice in case of the smooth strain the mouse died, whereas, in case of the rough strain the mouse survived. And interestingly, it was observed also that when this R strain which is non-pathogenic is cultivated or mixed with either the inactivated heat killed S strain or cell free extracts from this pathogenic strain, then, it was observed that when injected into the mice, now, the rough strain could actually kill the mouse. And therefore, it became pathogenic simply by being cultivated in the presence of the inactivated S strain or derivative cell extracts from it. So, these are very interesting observations, and this phenomenon was referred to as transformation.

Later on, scientists Oswald Avery and colleagues further tested the constituents of this smooth strain for the transforming potential. So, they took different macromolecules: RNA, protein, DNA etc. and then they incubated them with the R strain as per this experiment already described, and they found that only DNA from the S cells could transform the R cells to pathogenic form. So, it was concluded that the transforming principle is DNA and not these other macromolecules.

(refer time: 07:25)

So, what is DNA made up off? Basically, it consists of sugar, bases and a phosphate. So, the sugar present in DNA is 2 prime deoxyribose, you can see the C2 prime (C2') it has, it does not have a hydroxyl group, whereas a hydroxyl group at this position is present in RNA or ribonucleic acid that we will be discussing soon. So, this sugar deoxyribose, it has got 5 carbon atoms as you can see and it is a cyclized form of a pentose.

There are 4 carbons in the ring, C 1 to 4 prime, and along with an oxygen atom which is termed as O 4 prime and the fifth carbon is not a member of the ring, but it is attached to the C 4 prime. So, a phosphate group is also part of DNA and this is attached to the oxygen attached to the C 5 prime carbon atom in a nucleotide and in C 3 prime, in a dinucleotide or a poly nucleotide and that bond is referred to as the phosphodiester bond.

Bases present in DNA or heterocyclic nitrogen containing molecules. There are two types: Pyrimidines and Purines. So pyrimidine has got a 6 member aromatic ring with 2 nitrogen atoms at positions 1 and 3. And it is one which gets attached to the sugar. And it has got 4 carbon atoms and various side chains or groups. The purines have got two fused 5 and 6 atom rings and the positions 1, 3, 7 and 9 are nitrogens, and the rest are carbons. And it is the nine position that gets attached to the sugar. The main bases present in DNA are: cytosine and thymine are the pyrimidines, and guanine and adenine are the purines.

So, the chemical constituents of DNA were known, and it had been found that the ratio of the amounts of A to T and G to C are close to unity in DNA by Chargaff in 1952.

(refer time: 10:09)

Several groups of scientists were interested in solving the structure of DNA, particularly after the experiments of Avery demonstrating that it was likely to be the genetic material. These included Wilkins et al, Franklin and Gosling, who were trying to solve its structure by X-ray diffraction, and Watson and Crick, who were trying to deduce the structure based on the information that was available to them.

Three papers were published in the same issue of Nature in 1953, reporting their conclusions regarding the structure of DNA. Watson and Crick hypothesized, based on published experimental data, including (Chargaff's), and stereo chemical arguments, which are based on model building that they were engaged in, that the DNA consists of two helical chains which are coiled around the same axis, these are right-handed helices but they run in opposite directions, that is they are anti parallel.

They proposed that the chain consists of phosphodiester groups joining the beta D de deoxyribofuranose residues with 3 prime, 5 prime linkages. The bases were on the inside and the phosphates were on the outside, according to this model, which was different from an earlier model proposed by Linus Pauling, and also Watson and Crick had come up with that model, which was not published by them though.

So, they also proposed that in this current or new model proposed by them, the residues recur every 3.4 angstroms in the z direction and there are 10 residues for one full turn after 34 angstroms. And the distance of the phosphate from the axis is 10 angstroms thus implying that it should have a diameter of 20 angstroms. And that the planes of the bases are perpendicular to the axis. Purines paired with pyrimidines of the opposite strand, and the base pairing was rather specific, that is A would only pair with T, and G with C.

Wilkins et al had an X-ray photograph of DNA and they interpreted it and they also proposed that the helix would have a pitch of 34 angstroms, a diameter of 20 angstroms with 10 nucleotides in 1 turn and also that the nitrogenous bases would be in the central region of the helix. Franklin and Gosling had obtained a very high-quality X-ray diffraction image, famously known as Photo 51 nowadays, which was the label that they had given to that photograph of DNA from calf thymus.

And this was actually a critical evidence in determining the structure of DNA; it is the structure of the more hydrated B-form of DNA. And this pattern which they observed in the X ray diffraction image which you can also find by looking up this paper as well as this other article, it was consistent with a helical conformation. And they also proposed that there are 10 residues per turn. And the helical

structure had a diameter of 20 angstroms. The phosphates were on the outside, sugar and bases turn inward. And there were 2 coaxial molecules with a cylindrical repeat unit of 34 angstroms. So very similar conclusions by the three groups, and this gave rise to what we now understand as the double helix, Watson and Crick model of DNA structure, of B-DNA.

(refer time: 14:14)

So according to the model, DNA is a double helix, a beautiful helix shown here- this is an artist's depiction, but it aims to be accurate in terms of dimensions. And this helix as you can see, it resembles a spiral ladder with the bases that would be the rungs of the ladder in this case. So, DNA is a double helix with the sugar, phosphate backbone on the outside and the bases on the inside, and the two strands, they run in opposite directions. That is, they are anti parallel, and also they are complementary with a specific base pairing which happens via the hydrogen bonds.

(refer time: 15:04)

So let us try to understand the structure of the double helix in more detail starting with the components. So, one of the main components is sugar: 2 prime deoxyribose. I already mentioned that in case of RNA at this position, there is a hydroxyl group; the phosphate is attached to the C 5 prime of one sugar and the base is attached to this oxygen attached to this C 1 prime of the ring and it forms an N-glycosidic bond.

The N 1 of a pyrimidine and N 9 of the purine are connected to the sugar at the C 1 prime position. There is also a phosphate group, I already mentioned, which is attached to the C 5 prime. These are the structures of the bases present in the DNA. And what you can observe is that they have certain differences, they have for example, cytosine has got an amino group, so has guanine; and thymine and adenine-they have a distinct groups attached to them (e.g. methyl group in case of T and amino group in case of A), which accounts for the differences among them.

(refer time: 16:39)

The basic building block or repeating unit of DNA is a nucleotide. It consists of a base, a sugar, and a phosphate. The nucleotide AMP, adenosine monophosphate, is shown here with one phosphate and deoxyribose sugar and the base adenine; the phosphate is attached at the C 5 prime position whereas the base is attached via the N-glycosidic bond at C 1 prime of the sugar.

A DNA polymerase can use nucleotide triphosphates as precursors and can add a second nucleotide via a phosphodiester bond at the C 3 prime of the first connected to the C 5 prime of the next nucleotide and thus a poly nucleotide would have a free 5 prime phosphate and a free 3 prime hydroxyl end. For the growth of the chain, the hydroxyl group at the 3 prime positions on the terminal nucleotide attacks the linkage between the alpha and the beta phosphate groups of the incoming NTP. One nucleotide gets added and a pyrophosphate ion is released, resulting in the formation of the trinucleotide as shown in this example.

(refer time: 18:07)

This slide shows base pairing found within DNA. So each base has got unique locations of hydrogen bond acceptors and donors based on these different groups, which are attached to them. For

example, in case of adenine this amino group is there at C 6 position and it can serve as a donor for 2 hydrogen bonds. And also in this case, the lone pairs of electrons are the nitrogens 1, 3 and 7 impart a partial negative charge so they can act as hydrogen bond acceptors.

And likewise, guanine has got this CO group and an amino group at 2 and this oxygen here is a hydrogen bond acceptor. The amino would be a hydrogen bond donor. So, the base pairing is specific, it is based on the potential of these groups to form the hydrogen bonds and their orientations,

(refer time: 19:24)

relative to each other. So the base pairing that occurs in DNA is shown here. You can see that cytosine and guanine they are bonded by 3 hydrogen bonds in this configuration and adenine and thymine are bonded by two hydrogen bonds.

(refer time: 19:45)

This diagram shows a specific base pairing scheme which is present in the double helix. So, in this diagram, you can note that the two strands they are anti parallel, okay, 5 prime to 3 Prime, and 5 prime to 3 prime in this direction. So, this scheme, it suggests a mode for replication, or copying of the genetic information. The bases as you know, they are complementary. So, it was famously stated in the paper by Watson and Crick that "it has not escaped our notice that the specific pairing that we have postulated immediately suggests a possible copying mechanism for the genetic material". When you have one strand and you know the sequence of it, the other strand as per rule has to be complementary. So, this was thought to be very exciting by everyone in the field.

(Video Starts: 20:45)

This animation shows the three dimensional structure of DNA. Note the phosphate groups, these large yellow spheres on the outside and the bases containing these navy blue atoms on the inside; also observe the major groove which is broader and the minor groove, which is narrower.

(Video Ends: 21:07)

(refer time: 21:08)

So, to summarize what I have said so far, DNA is a polymer consisting of 4 nucleotides, bases linked to phosphorylated sugars, and the nitrogenous bases- they are purines and pyrimidines. Purines, adenine and guanine, the pyrimidines, cytosine and thymine. There is a 5-carbon sugar, which is part of the DNA molecule building block, it is 2 prime deoxyribose. And there is a phosphate group attached to it to the C 3 prime and 5 prime of successive sugars forming this phosphodiester linkage. Again, it is a double helix-it is got 2 antiparallel strands which are twisted around each other, purines pair with pyrimidines of the other strand via hydrogen bonds, A with T via 2 hydrogen bonds and G with C via 3 hydrogen bonds. And in the B form of DNA the rise per base pair is 3.4 angstroms. There is one full turn every 34 angstroms and there are 10 bases per turn. And the diameter of the helix is 2 nanometers.

(refer time: 22:35)

So here is a space filling model of 1.5 turns of a DNA double helix. And you can see that each turn, that is from here to here, is made up of 10 nucleotide pairs. And the center to center distance between adjacent pairs is 3.4 angstroms. The coiling of the two strands around each other, it creates two grooves, okay, the wider one is referred to as the major groove and the narrower one is referred to as the minor groove. So, the bases in the major groove, they are more accessible and probably can be more readily recognized by DNA binding proteins.

(refer time: 23:25)

This slide illustrates the semi-conservative replication of DNA. So, the two parental strands, they would separate and each would serve as a template for the synthesis of the new daughter strands shown here in gray by complementary base pairing. And this model, in fact, we know this to be true and this was also confirmed experimentally later on, by Meselson and Stahl, using DNA labelled with the heavy isotope of nitrogen, N^{15} and coupled with density gradient centrifugation using cesium chloride density gradients, it was observed that intermediate density DNA could be observed after shifting from the heavy to the light isotope of nitrogen in one round of DNA replication (more details of this experiment will be discussed in a later lecture in this course). And this was consistent with semi-conservative mode of replication. So, there was a very nice boost to this model.

(refer time: 24:34)

In this animation, made and provided by Drew Berry and Etsuko Uno, we can see how the DNA is organized within the cell.

(Video Starts: 24:45)

In the cell, the DNA whose backbone is negatively charged, associates with basic histone proteins that form an octamer around which the DNA is wrapped. The histone octamer has got 2 each of 4 histones, H2A, H2B, H3 and H4, and there is also a linker that is associated with histone H1; the histones are shown here in blue; this unit of chromatin organization is referred to as a nucleosome. It forms a 11 nanometer diameter fiber, which is further folded to form a 30 nanometer diameter fiber. Note that the histone tails are exposed to the outside; these can be modified. This thread is further folded to form what we are familiar with as a folded chromosome structure seen in metaphase. And the organization of this is just beginning to be understood; it perhaps happens by formation of loops and their further folding in association with non-histone proteins as well.

(Video Ends: 25:52)

(refer time: 25:53)

We have discussed that base pairing is specific. However, there are no restrictions on the type of sequence for the successive or adjacent bases next to each other on one strand. And this can be seen here as the output of a sequencing reaction. You can have, you know, different variations of DNA sequence. So, there can be any combination of successive bases forming the sequence, and this sequence of bases actually corresponds to the genetic information.

(refer time: 25:53)

Here is a list of some of the key references for this lecture that will be useful to read up. Stay tuned for the next lecture on the decoding of genetic information.