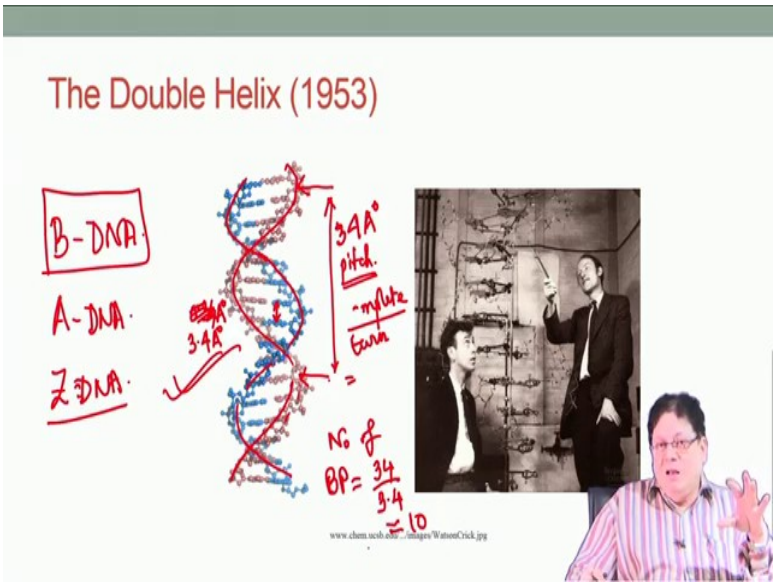
Organic Chemistry In Biology And Drug Development
Prof. Amit Basak
Department of Chemistry
Indian Institute of Technology, Kharagpur

Lecture - 23
Nucleic Acid (Contd.)

Welcome back, in the last session, we have given you some basic concepts about the structure of nucleic acids mainly, the deoxyribonucleic acid (DNA).

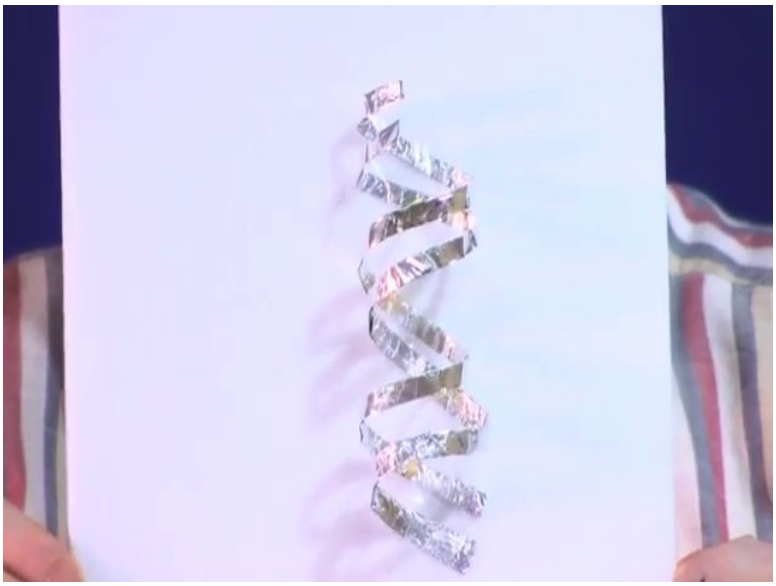
(Refer Slide Time: 00:33)



We know that these two scientist who are shown here their picture (Francis Crick and James Watson) proposed this double helical model for DNA. The most beautiful aspect of this double helical model is that the presence of hydrogen bonding interaction between the bases. So, the bases are stacked inside. G forms hydrogen bond with C and the number of hydrogen bonds within G-C are 3, similarly the number of hydrogen bonds between A and T are 2.

So, whenever one strand has A, in the opposite strand, the other complimentary base T will be present in the other stran. If it is C here, that will be G there. So, all complimentary bases will be paired with each other. Now, while doing that they adopt a helical conformation and that is why it is called double helix. Now the double helix structure is shown here. I have made a kind of a model to make it little bit clear, to show you a double helix, which is made up of just aluminum foil, you can see that.

(Refer Slide Time: 02:13)



This is a small piece of the double helix. You have to notice something. First of all these two helices are right handed helix. I told you about this concept in case of the proteins, what is the right handed helix? That means, the way a right handed screw goes, that is called right handed. So, there are two strands, interestingly both the strands are right handed.

And so, what is meant by the anti-parallel strand? That means, on the top this if this strand is 5'; that means, the other strand, this is if this is end point, that will be 3' and on bottom exactly the reverse will happen.

Now, we will discuss some of the important features of DNA. What is the pitch of this double helix? Pitch means again, if you take one strand and you make a complete rotation, if you traverse from one point and go to a complimentary point in the same strand, then that the rise in the height is called the pitch of the helix.

Suppose, I take this point and so the pitch will be from this point; take the same strand and go bellow and take the other point. So, the distance between these two will determine the pitch of the screw, the pitch of the helix, in this case. According to the diagram, this is your one strand that is going like this and this is the other strand that is going like this.

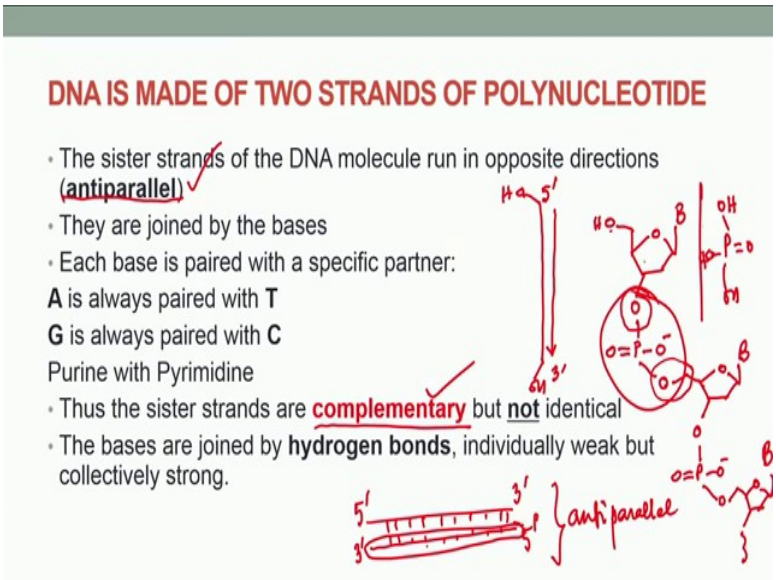
So, what will be the pitch here now? If you take this red strand here, follow this red strands and go to a similar point, similar point means this one. So, this distance is basically the pitch.

If you take a complete turn, then the rise in height or if you start from here, what is the lowering of the height? That gives the pitch.

If you start from here going to there; rise in height, if you come from here to there is lowering of a height. This is what a complete turn. Watson and Crick. proposed the structure for B-DNA. There are again different kinds of DNA; one is a B-DNA; another is A-DNA; another is Z-DNA, but the DNA that Watson Crick used was B-DNA, it is the most common DNA that is encountered.

So, we will concentrate on the B-DNA. In the B-DNA, this one complete turn is 34 Å so; that means, from here to there you cross 34 Å in length; if you go to the next base pair, that is actually 3.4. So, if that be the case; that means, if I ask that how many base pairs are there in one complete turn? So, number of this pairs in one complete turn will be 34 divided by 3.4. So, it is approximately 10. So, there are 10 base pairs in one complete turn. In this A-DNA there are some DNAs, which are more compact than this. There may be around 11 base pairs, in A-DNA; but as I said, most common DNA is the B-DNA and Z-DNA is little different, it has left handed double helix. A-DNA is the one which is more compact. Remember, the DNA is an acid it is a nucleic acid. There are these phosphates, which connect two nucleotides.

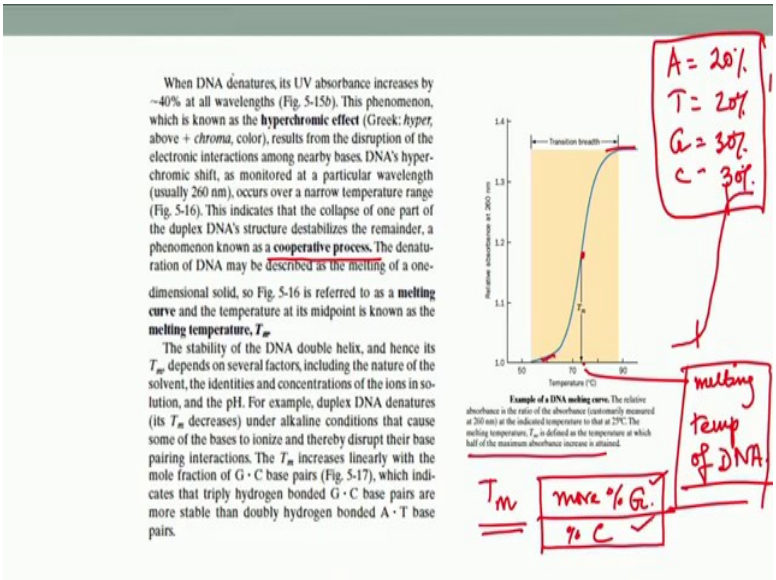
(Refer Slide Time: 08:05)



Now, so this O minus makes it an acid actually, this is OH, phosphoric acid derivative. So, , if we compare this part with phosphoric acid. Phosphoric acid is nothing, but P double bond O, three OH groups; it is a tribasic acid. Out of these 3 OHs, one of the OH is forming an ester with the 5' OH, another OH is forming an ester with 3' OH. So, this linkage is called phosphodiester linkage. So, if someone asks you that what is the linkage between the nucleotides? It is the phosphodiester linkage.

So, this is an acid, because one ionizable hydrogen is still there and at the biological pH, it will be O minus. Definitely there will be some water, because there is water in the biological medium. So that can be hydrated but, if it is put in a very anhydrous condition, then the DNA becomes more compact, because not many of the water molecules are there. So, that is why A-DNA has around 11 base pairs per complete turn, but B-DNA (that is our normal DNA) under normal condition, has got 10 base pairs per complete turn. So, these are some statics about this DNA structure.

(Refer Slide Time: 09:55)



In DNA, two strands are connected in anti-parallel fashion by hydrogen bonds. Can you break this hydrogen bonds? Yes, you can break these hydrogen bonds, because after all these are not covalent bonds, even covalent bonds can be broken, if you raise the temperature, molecules will dissociate..

DNA is the double helix. So, at a very narrow temperature range, it has been found that the DNA becomes separated, the two strands of the DNA become separate and that happens within a very narrow range. Why narrow range? It is not that as I raise the temperature from 30 to 35°C, 5 percent of the hydrogen bonds are broken, then way into 35 to 40°C, another 5 percent, not like that.

The way to follow this breakage of hydrogen bond is to study the absorbance at the λ_{max} of the DNA. See DNA has bases which are aromatic in nature. So, they absorb at 260 nm; around 260 nm is their maximum absorption (λ_{max}).

So, what you do? You take the double stranded DNA and measure the absorbance at 260 nm; you start at temperature say 25°C. Now, you start heating the DNA. So, you will see that the absorbance initially increases very slowly and finally, at a certain narrow range of temperature, it just increases rapidly and then becomes stationary. So, this will be the absorbance increment curve. Remember, we are measuring the absorbance at different temperatures at 260 nm.

Now, the question is that why it occurs with such a narrow range of temperature, Are all these hydrogen bonds broken at the same time? This is what is called the cooperative effect, means somewhere, if some breakage starts, immediately the other hydrogen base pairs follow that; it is like, if you have a stack of bricks and if you push one brick, each brick starts pushing the other in a cascade of events. So, then all the bricks ultimately collapse from the upright position into the horizontal position. So, that is called cooperative effect. Although, you have only disturbed one brick, but the whole brick system falls.

So, here also when you break certain number of hydrogen bonds initially, the whole thing now collapses, it is a cooperative effect. So, this cooperative process happens within a very narrow range of temperature.

You plot the relative absorbance at 260 nm. Relative absorbance is measure with respect to some absorbance So, what you do? You first measure the absorbance at 25°C and then you raise the temperature. So, measure at 30°C, what is the absorbance, then at 35 and then what ever absorbance values you are getting you actually divide that by the absorbance at 25 multiply it by 100.

So, you can express it in percentage or you can express it in actual value. But what is important here, is that the absorbance increases. See not only the hydrogen bond is broken within a narrow range of temperature by a cooperative process, but the absorbance also increases and then becomes stationery; that happens, because the DNA bases are initially stacked inside of the double helix.

So, there will be less absorption of light at that time when the hydrogen bonds are intact; and when the hydrogen bonds are broken, then the bases are now, more exposed to the UV light. So, there will be more absorption. As you break it, the bases are free to absorb otherwise, earlier, they are all stacked together inside congested narrow pocket inside the DNA. So, as it becomes open it absorbs more. So, there is an increase of the absorption. Now, in this sigmoidal curve, if you take the middle point of this rise, this middle point; that means, where I can say half of the hydrogen bonds (50 percents of the hydrogen bonds) are broken at the middle point.

So, that is what is called the melting temperature (T_m) of DNA; melting temperature of DNA. T_m is defined as the temperature, at which half of the maximum absorption increase is attained; that is one way of saying it, but you are not telling about the structure. By giving this definition, you tell about the structure, if you say that 50 percent of the hydrogen bonds are broken at the melting temperature. It is not melting of organic compounds. So, remember it is not going from solid to liquid; it is actually, the separation of the hydrogen bond. So, we are talking about percentage of hydrogen bonds that have been broken.

So, we have a number now, 50 percent; the temperature, at which 50 percent of the hydrogen bonds are broken is called the melting temperature of the DNA. Now, this melting temperature is very important, because if a species has a low DNA melting temperature, then what will happen? As the environmental temperature increases, so for that species the DNA will become single stranded. If DNA becomes single stranded, all the subsequent processes like the flow of information, which are called replication, transcription, all these processes will fall apart. DNA has to be double stranded in order to transmit all the information.

So, nature has created these DNAs in such a way that the melting temperature is way above the environmental temperature for most of the species. For most of the species, it will be above 70 or above 80, but temperature will never reach that mark; the maximum temperature that would be around 50°C in Rajasthan somewhere.

The body temperature is around 37°C, the melting temperature for human DNA is so high that it remains in the double strands. So, maintaining a double strand is very important. Now, the reverse is also true. See when you heat the double stranded DNA. So, it becomes single stranded.

Now, if we think that I will cool the solution. So, if I cool the solution, what is going to happen? Now, these strands will slowly come together and try to find the complementarity in the other strand. So, at some point of time, they will also join together.

So, ultimately as you cool, it will become a double stranded; the rate of strands separation and rate of this strands annelation (which is also called as hybridization that when two strands find the complimentarity and become a double strand) may vary. The rate of hybridization will on the concentration that you are using, because after all it is the concentration that how close are these molecules are, that will decide the rate; remember the point that you can again bring back the double helix by cooling the same solution.

So, if you heat, DNA forms the single strand and then as you cool it becomes the double strand again. This is also very important when you read the molecular biology. Now, the question is what are the factors on which melting temperature depends? Obviously, if you have a DNA, where there is more percentage of G that means there is more percentage of C; that means, there are more percentage of this 3 hydrogen bonds scenario that are between G and C. So, if you have more G content; that means, you have more C content so; that means, your hydrogen bonds are more strong now.

Now, the two strands are connected by a larger number of hydrogen bonds, because G is more, hence C is more. So, your T_m will increase, because it will require more energy now to separate the strands. So, the first thing that T_m depends on is the percentage of GC content. You should also remember that for DNA analysis, suppose, you want to know the percentage of G in the DNA; if you know the percentage of any one of the bases, you can calculate the percentage of the other bases.

Suppose, in a DNA you have percentage of A is 20 percent. So, that will give you the percentage of all the other three, because if A is 20 percent, T is 20 percent and so, that will consume 40 percent. So, out of the rest 60, G is 30 and C is 30. So, you do not have to analyze all, you can analyze only 1 and get the percentage for the other three. T_m depends on the length of the DNA. The greater the length of the DNA, more the number of hydrogen

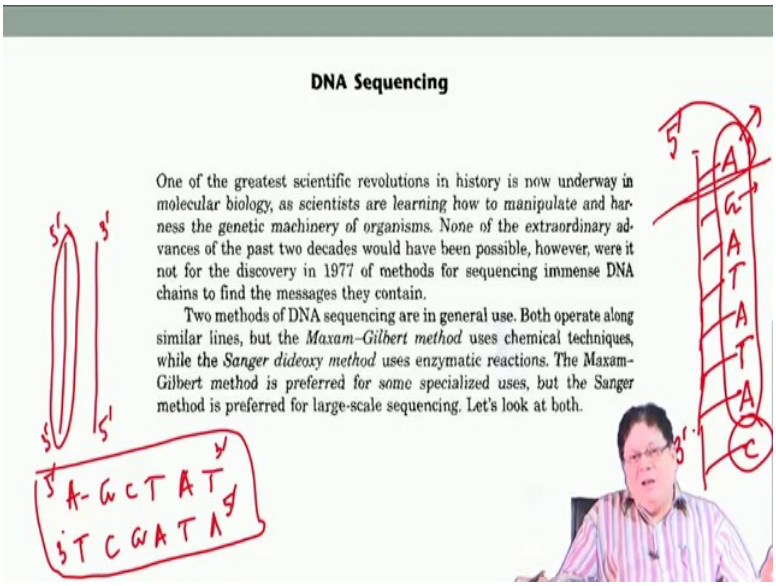
bonds, that means, you require more energy to separate them into the single strand. So, for shorter DNA, the T_m will be less, if it is a larger DNA T_m will be more. Then T_m also depends on presence of a molecule which disrupts hydrogen bond, like urea.

Urea is a famous molecule, which disrupts hydrogen bonds in protein as well as in DNA. So, it is a hydrogen bond disruptor. So, if you try to determine T_m in presence of a hydrogen bond disruptor, your T_m will fall, because that hydrogen bond disruptor will assist in destroying the hydrogen bonds. If you add compounds like urea, that will also lower the T_m, because it helps in breakage of the hydrogen bonds. A fourth point is also important. If you take an ionic compound suppose, if you are taking a DNA in the sodium chloride solution and then trying to determine the melting temperature in sodium chloride.

Now, you got a melting temperature of say 60 °C, then you increase the concentration of the sodium chloride and the question is what will happen to the T_m? If you increase the concentration of the sodium chloride, what will happen? Because of the presence of the excess chloride ion now the DNA phosphate will repel these chloride ions and so, the two strands will be more intimately connected. They do not want to separate, because they are all surrounded by the chloride ionis. So, they want to push this DNA and make it stronger; if it does that; that means, higher concentration of sodium chloride will enhance the melting temperature of the DNA.

So, these are the four factors; one is GC content, the other is the length of the DNA, the third is the hydrogen bond disruptor and fourth is the presence of ionic compounds like sodium chloride.

(Refer Slide Time: 24:09)



So, now the next thing is that, when we start the protein chemistry or the peptide chemistry, we came to know that there is hierarchical structure in proteins: primary, secondary, tertiary and quaternary. DNA is already known to be double helix. So, more or less, the three dimension geometry is known. What is important here is to know the sequence of the bases, that means how the bases are arranged one after another in a piece of DNA.

So, there could be different types of sequences that are possible in a DNA, but these sequences are so important, because that will decide that what type of protein will ultimately be made. So, when you talk about protein sequencing, primary sequencing means when you want to know the covalent primary structure means; where covalent bond is involved in joining of the of the amino acid units. In DNA structure analysis means, the DNA base sequencing; that means, what is the sequence of bases, because other things remain same. Other things means the sugar is same all the time; in DNA it is deoxyribose and the connecting bound is phosphor diester linkage.

So, the difference between two sets of DNA is basically, the sequence of bases that are there in the DNA. So, it is a challenge, but until this was solved, the DNA chemistry was difficult to take forward further, because what was needed is the knowledge about the sequence of DNA; and as I said, this is important, because that differentiates one DNA from the other DNA. Now, what are the problems of doing DNA sequencing? Remember, in proteins, there is a technique called Edman degradation, which involves the reaction with the N terminus

and then breaks the peptide bond selectively; the first amino acid comes out and the remaining peptide could be isolated and we can again do the Edman degradation.

So, it is a repetitive process. Suppose, the DNA is having this type of sequence (first base be A and the second base be G) and suppose, this is the 5' end and this is the 3' end. If you want to know the first nucleobase present, then you should develop a technique, where only the first base is taken out and then your remaining part can be isolated. Thus you should develop a technique which must be different, because the second one is G. The problem is how to differentiate between A and G since both are purine bases.

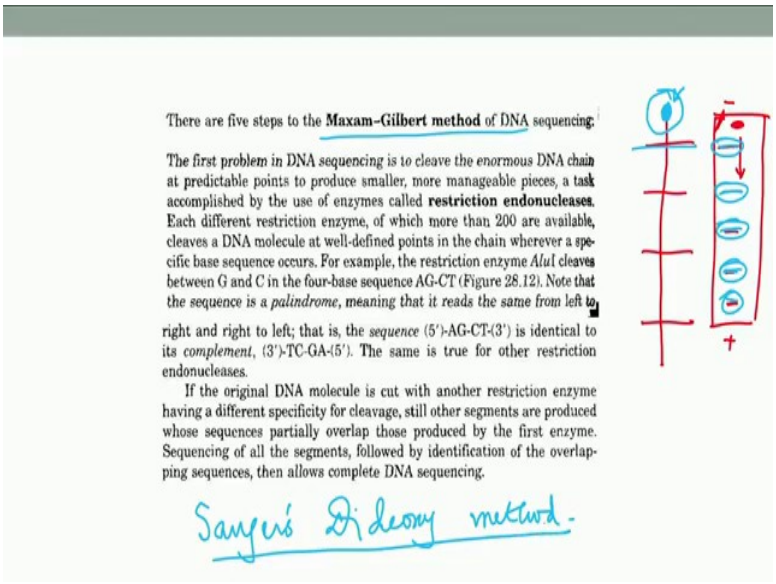
So, it will be very hard to find a condition, where only A will be cleaved; that means, the A containing the sugar phosphate will be cleaved and you will get a DNA, which is only devoid of A, because A and G are very similar in their structure and there is slight difference in their reactivity; obviously, because one has got only NH_2 6-amino purine, the other has a carbonyl at the 6 position and there is a another NH_2 there and NH_2 there. So, the problem is to find out very selective reagents, which will cleave at either A or G or C or T.

So, that was the challenge, how to do that? The other thing is that you have a double stranded DNA, anti-parallel 5' 3' and 3' 5'. So, initially you need to separate these strands, because you cannot do the sequencing for both the strands at the same time, in the same solution.

That is not possible, but one thing is that if you know the sequence of one strand, you can immediately write the sequence of the other strand, because you know that suppose if 5' to 3' is like this A G C T A T; and I ask you to write the complementary strand, you can immediately write 3' to 5' T C G A T A.

So, that will be the other strand. So, you do not need to know the sequence of both the strands, just one strand is enough. So, the initial thing that you have to do is to isolate only one strand. That you can do by melting the DNA and then before it can hybridize back, you separate these two strands, that is possible. You can separate these strands from a solution.

(Refer Slide Time: 30:01)



But before that of course, there is something some other important point which you have to consider is that, when the DNAs are cut , we have to cut at selected points of the DNA and then and then we have to try to locate these species.

Now, the question is; is DNA visible to your eyes? No, you cannot; it is not visible to eyes. Is protein visibly to eye? Yeah, there are some green fluorescence proteins, which are visible if you shine light, but DNAs are not visible. In case of protein, you know, there is a dye which was called coomassie blue.

So, coomassie blue goes and binds to the protein and then you can wash out the remaining coomassie blue and wherever there are proteins, only blue bands can be seen. Similarly, in DNA you have to have something, which will show you the presence of the DNA at a particular location.

So, basically the principle is as following: The DNA is broken into pieces at precise points and then these pieces are put in an electrophoresis gel. You know what electrophoresis is. In electrophoresis, you put the material; you apply voltage, then depending on the charge, it moves. Now, DNA is always negatively charged, because it has phosphate, it is a phosphoric acid; it is a phosphodiester; that means, one of the ionization OH is still there. So, it is an acid.

So, in biological pH, it is always negatively charged. So, the DNA will definitely move in this direction towards the anode. Now, if there are different DNA pieces of different sizes, what will happen? You will have a distribution like this; the smaller size will move fastest and the larger size will move slower, but again the issue is that you are not able to see the DNA here, because the DNAs are invisible. Electrophoresis is done in a matrix; in protein, you use a polyacrylamide gel; in DNA, there is another carbohydrate based material, which is called agarose.

So, it is called agarose gel, in case of DNA, but the principle is same, that you apply the voltage and then the DNA being negatively charged, moves towards the anode. In case of protein, which has an isoelectric point, you have to take care of the isoelectric point and depending on that it will move either to the cathode or the anode. In DNA, that is not the problem. It will always move to the anode (the positive electrode); and depending on the size, the smaller size moves faster and the larger size moves slower.

Now, to identify where are the DNAs, what you do? You have to put some label in the DNA so that suppose I put a label in the DNA, I put a blue tag and then I break here. So, I when I shine light, I see the the blue tag. So, all these DNAs will have this blue tag.

So, I should have a tag or a fluorescent reporter system, which will allow me to visualize this DNA. There are different techniques that are available, initially it is only the radioactive technique that was used; that means, instead of fluorescence, you put a radioactive tag here. So, wherever there are DNAs, those areas will become radioactive, because of the presence of the radioactive tag, if you now take a photographic plate and put on the top of this, that photographic plate will be affected by only those zones from where the radiation is coming, because you have radioactive material in each of these DNAs.

These are some of the fundamental principles; you can do it by fluorescence, but fluorescence came much later. Initially, it was done by your radioactive phosphorus. In the next session, I will discuss that, but broadly speaking, there are two methods which were simultaneously discovered at the same time; one is called the Maxam Gilbert method, this is a chemical method and the other is called Sanger's Dideoxy method and in the next session, we will discuss them in details.

Thank you.