Basics of Biology
Professor Doctor Vishal Trivedi
Department of Biosciences and Bioengineering
Indian Institute of Technology, Guwahati
Lecture 15
Nucleic acids

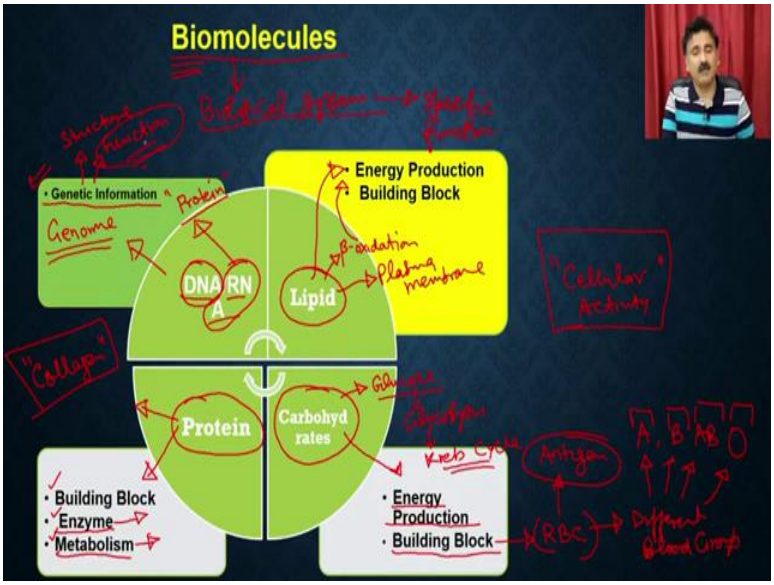
Hello everyone. This is Doctor Vishal Trivedi from Department of Biosciences and Bioengineering IIT Guwahati. And what we were discussing? We were discussing about the living organisms. So, so far what we have discussed in this course is we have discuss about the classification of the living organisms and then we also discussed about how these living organisms are related to each other and in that context we have discussed about the many types of evidences what we people have put together to explain that the higher organisms are been evolved from the lower organisms and so on. And then subsequent to that we have also discussed about the theories which are explaining the evolutions and so on.

And in the previous lecture we have also discussed about the cellular structures. We have discussed about the prokaryotic structure as well as the eukaryotic structures. And within the prokaryotic cells we have discussed about the genomic DNA. We have discussed about the cell wall. We have discussed about the many features of the prokaryotic cells. And when we were discussing about the eukaryotic cells we have discussed about the different types of membrane-bound organelles what are present in that eukaryotic cells. So we discussed about the nucleus. We discussed about the mitochondria, chloroplast, Golgi bodies, the organelles which are important for the vesicular trafficking such as the endoplasmic reticulums, Golgi bodies, lysosomes. And then we also discussed about the plasma membrane.

And we were discussing about these organelles, we have not only discussed about the structures of these organelles. We have also discussed about the function of these organelles for the cell. But while we were discussing about these functions we have been always talking about the different types of biomolecules which are participating into the different types of functions. For example, when we are talking about the vesicular trafficking we have said that the vesicular trafficking is being done by the various types of the tags or signals. So all these tags and signals are nothing but the biomolecules. So in this module and as well as in the subsequent module we

are going to start discussing about the biomolecules. So let us start about discussing about the biomolecules.

(Refer Slide Time: 3:27)



So when we say about the biomolecules, the biomolecules, the name suggest that these are the molecules which are going to be present into the biological system and they are actually going to have the specific functions. So these are the molecules which are present in the biological system and they are actually going to have some specific functions. So when we say biomolecules we are only going to talk about the some of the macromolecules or some of the important biomolecules.

So if you see about the organisms or the living cell they are actually having the four different types of biomolecules which are actually been participating into the different types of reactions. So we have the DNA or the RNA. And the function of the DNA and RNA is that they are actually going to carry the genetic information. Mostly the DNA is present in the form of the genomic DNA or the genome of that particular organism.

There are exceptions where instead of DNA we have the RNA as well as genomic DNA. And then the RNA is actually a molecule which actually carries the information from the DNA and that is how it helps in the protein synthesis. That anyway we are going to discuss when we are going to discuss about the some of the processes, biological processes such as the replication,

transcription and translations. So that time we are going to discuss in detail about the role of DNA and RNA.

Similarly, we have the lipid molecules, and if you remember in the previous lecture when we were talking about the cellular media, we said that the lipids are required for the energy production because the lipids are been oxidized under the beta oxidations, and lipids are the molecules which are formed by the summation of the glycerol and the fatty acids. And then they go under, the fatty acids actually undergo the beta oxidations and that is how they are actually going to produce the large quantity of energy. Apart from that lipids are also been an integral part of the plasma membrane. So they are also been considered as the building blocks.

Then, we have also discussed about the carbohydrates. So carbohydrates when we were talking about the preparation of the media we have also discussed about the carbohydrates. We said that the media also has the carbohydrates such as the glucose or the sugar. And the carbohydrates main job is that it is actually been functional as the energy metabolism or to the energy production. So glucose is been metabolized within the glycolysis and as well as the Krebs cycle. And that is how it is actually going to produce large quantity of the ATP. And that ATP is going to be utilized by the cell for several types of functions.

Apart from that the carbohydrates are also been present as a signal molecule in some of the cells. For example, the carbohydrates are present on to the RBC. And these carbohydrate molecules are actually giving, forming the different types of antigens, and that is how the RBCs could actually adapt to the different types of blood groups. So you can have the different types of blood groups because of the sugar modifications.

So what are different blood groups? We have the four different types of blood groups. We have the A. We have the B. So we have the AB, and we have the O antigens. So in the A we held A antigens in the B we have the B antigens. In the AB we have the A and B both antigens. And in the O we do not have any type of antigens. So the difference between these antigens is but it has a protein which is been modified by the carbohydrate molecules.

Then, we come back to the proteins. So in the proteins, we have the proteins are made up of the constituents amino acids. And proteins have a very, very… is having a huge role in terms of as

the building block. So proteins are actually been present in the plasma membrane. Proteins are the structural proteins. Like for example in the humans, humans are actually been able to walk or stand because they have a protein which is called as the collagen. So because of these collagen fibers the humans could be able to stand and walk because most of the bones are made up of the collagen fibers.

Then it also has the enzymes. So these enzymes are actually participating into the different kinds of reactions. Whether these reactions are detoxification reactions, what is happening in the liver, such as the enzymes which are involved in the hydroxy oxidase pathway or there are enzymes which are important for the neutralization of some of the toxic molecules.

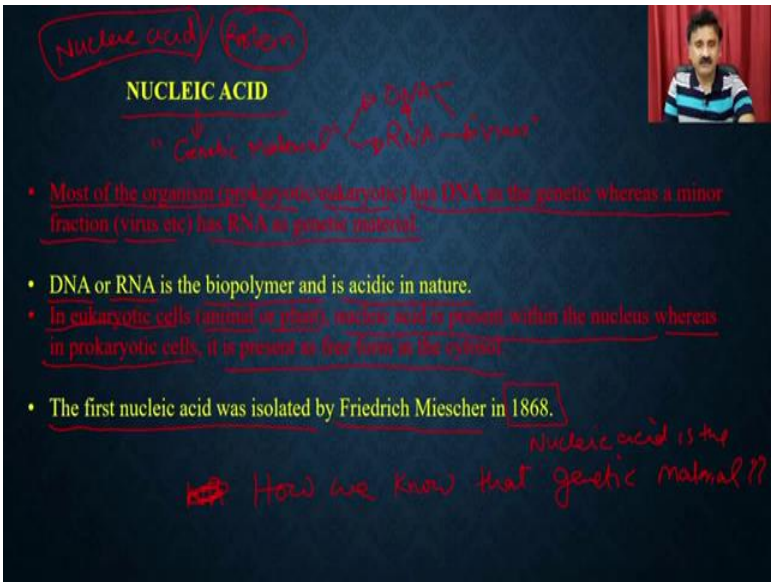
Apart from that, the enzymes are also been utilized in catalyzing the conversion of the substrate into the product and that is how they are actually going to shuttle the different types of constituents between the different pathways. And enzymes are also been participating into the different types of metabolic reactions. If you remember we have talked about the glycolysis and Krebs cycle which are actually important for the energy production, especially, and also the beta oxidations. All these oxidative pathways we are actually been using the enzyme as the driving forces, and that is how the proteins are important for as the building block. Proteins are important as an enzyme, and protein is important for the metabolism.

And that is why if you want to understand the cellular activities or if you want to understand the cellular functions it is important to understand the structure as well as the functions of these important biomolecules.

So today we are going to start with the nucleic acids. And with the nucleic acids we are going to understand about its structure, and we are also going to understand about the important functions. And because as I said, in our subsequent modules we are actually going to discuss about the functions in detail where we are going to talk about out the replication, transcription and translation, so those detail functions we are going to take up and the mechanism as well into those particular modules.

So let us start our discussion and we are going to start with the nucleic acids.

(Refer Slide Time: 10:32)



Nucleic acid, the most of the organisms like whether it is the prokaryotic organism or the eukaryotic organisms has DNA as the genetic material, whereas the minor fractions such as the viruses has RNA as the genetic material. So nucleic acid is important because it is a genetic material and what is present in the cell. Mostly the genetic material is DNA, whether it is the prokaryotic or eukaryotic cell.

But in some cases like for example the viruses you have the RNA as a genetic material. But even then, when the viruses have the RNA as a genetic material, that RNA get converted into the DNA when the virus is entering into the cell, and that is how it actually propagates within the host as a DNA molecule. So in most of the cases the DNA is going to be the genetic material, and in some cases the RNA is also going to be a genetic material. DNA or the RNA, whether it is DNA or the RNA is a biopolymer and it is acidic in nature. It is acidic in nature because it has the groups like that.

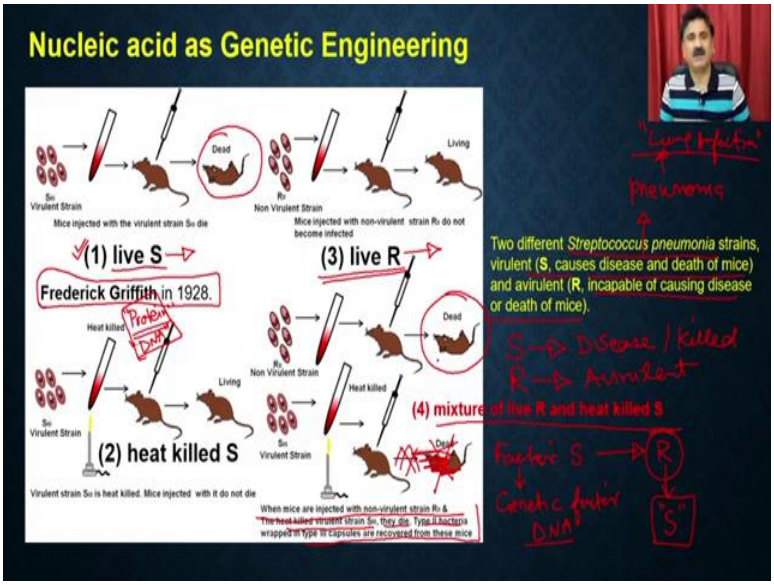
In eukaryotic cells, whether it is the animal cell or the plant cell, the nucleic acid is present within the nucleus. That we have already discussed the DNA is present inside the nucleus. Nucleus is a double membrane structure which has the nuclear membrane and other kinds of structures, whereas in the prokaryotic cells it is present into the freeform into the cytosol.

The first nucleic acid was isolated by Frederick Miescher in the year of 1868, and since then the people know about the nucleic acids. Now the first question comes is how we know that genetic material is, that nucleic acid is the genetic material? So how we know? Because we have the 3 candidates. We have the nucleic acid; we are the nucleic acid which can be a genetic material, or it can be a protein because earlier people were thinking that protein could be a genetic material rather than the nucleic acid because protein is actually a functional molecule.

So it can do a lot of functions. So people were under the impression that the protein could be the functional molecule. So that could be actually carrying the information from one generation to another generation. And that is how it could be, that could be the genetic material, because the purpose of the genetic material is that it is actually going to carry the information from the one generation to the next generations.

So then people have started doing the classical experiments. There are many experiments what the scientists have done we have a classical experiment which is been done by the Hershey or Chase. And we have also experiment by the Avery. And we also have the experiment which is done by the Griffith.

(Refer Slide Time: 14:00)



So what I am going to talk about the Griffith experiments. So what Griffith has done, so Griffith was a scientist and he has performed an experiment in the year of 1928, where what he has done

is he has taken a mice, and he has taken the two strains of the Streptococcus pneumonia. So that Streptococcus pneumonia is causing a disease which is called as pneumonia. So pneumonia is a lung infection. So it is a lung infection disease and it is actually lethal if it is not been treated because it is going to destroy the lungs, and that is how the person will not be able to take the respirations.

So what he has done is he has taken the two different strains of this Streptococcus pneumonia strain. One is strain which is called as the S strain which causes the disease and the death of the mice. So he has taken the two strains, one the S strain which is actually causing the disease and it is actually going to kill the mice. So that is a virulent strain.

Then, he has taken another strain which is called R strain, and R strain is a avirulent strain and it is not been able to cause the disease. So it will not be going to cause the death of the mice. So avirulent strain, R is avirulent strain. So what he has done is initially he has tested whether this fact is true or not. So what he has done is he has taken the four set of mices. In the first set of mice he has injected the live strain. So what he has done is he has cultured these strains into a culture vial, and then he has injected those strains in into the mice. And what he could found is that the mice are dead, because they might have developed the pneumonia and that is how the mices were dead.

Then, what he has done is he also tested whether the vial, R strain is also good or not. So what he has done is he has taken the R strain. He cultured the those strain and then he injected the R strain into the mice. And what he could find is the mice could not be able to develop the, disease and he was alive. So that is what he has tested. And by this two experiments he tested the fact that the S strain is actually virulent strain and the R strain is a avirulent strain.

Then, what he has done is he has actually cultured the virulent strain, and then before injecting the bacteria to the mice he has heat killed them. So once you heat kill them you are actually going to kill the protein part. So you are actually going to destroy the proteins. You are actually going to destroy the protein.

So when you destroy the proteins you could not be able to cause the disease. So it could not be able to cause the disease because the genomic DNA was … so it could be able to kill the protein

and as well as the DNA. So it could not cause the disease. So when he heat killed the bacteria, the bacteria could not be able to transfer the genetic material or could not be able to grow. And that's how he could be able to cause the … he could not be able to cause the disease.

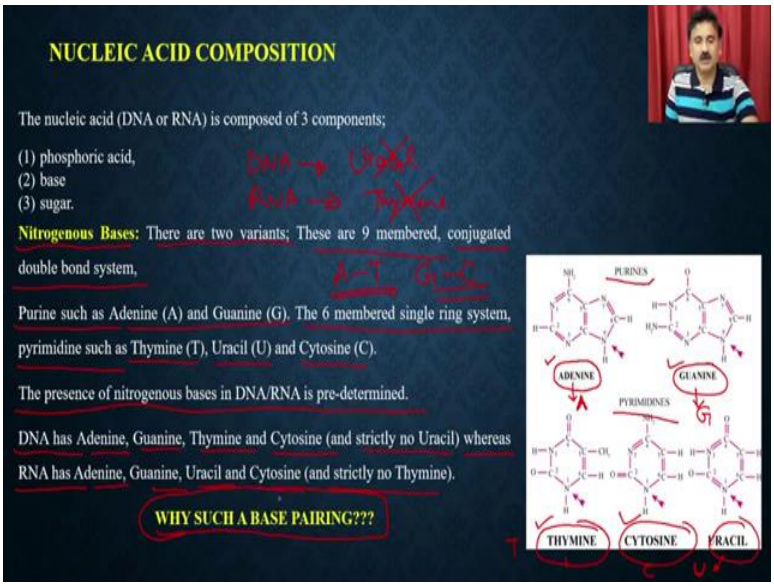
Now, what he has done is he has first heat killed the virulent strain. So he got the heat killed virulent strain. And then he mixed those killed virulent strain, and he has he has injected; he has mixed that with the avirulent strain. And when he has injected those mixture, when he injected the mixture of the virulent strain and the heat killed S strain, what he could found is that he could found the death of the mice.

So what he could be able to conclude that is a factor which was present in the S strain, and that factor was transferred to the R stream and that is how the R strain got converted into S strain, and that is how when the R strain, he is injecting the R strain on its own he could not be able to kill the mice. But when he heat killed the virulent strain that is not causing the death.

But when he has mixed this heat killed mice and he could inject that into the mice he could find that the mice are dead. So this you can ignore. This is just a wrong drawing actually. So when mice are injected with the non virulent strain and the heat killed virulent strain, they die. Type 2 bacteria are recovered from these mices.

So, by doing this experiment Griffith has concluded that there is a factor which is genetic factor, and that genetic factor is nothing but the DNA what was being transferred from the S strain to the R strain and that is how the R strain has also acquired the phenotype of the S strain, and that is how when he has injected the mixture it could be able to lead to the death of the mice. So by doing this experiment, and we have couple of more experiments which people have done, like Hershey-Chase experiments and Avery’s experiment, that also proves that the DNA is the genetic material.

(Refer Slide Time: 19:49)



So what is the composition of the nucleic acids? So nucleic acid is composed of the 3 components. You have the phosphoric acid and because of that only the nucleic acid is having the acid-like properties. Then you also require the base. So these are the nitrogenous base. And then you also require the sugar. The phosphoric acid provides the backbone to the polymer whereas sugar work as anchoring point for the nitrogenous base. The 9-membered nitrogenous base gives the diversity into the sequence of the nucleic acids.

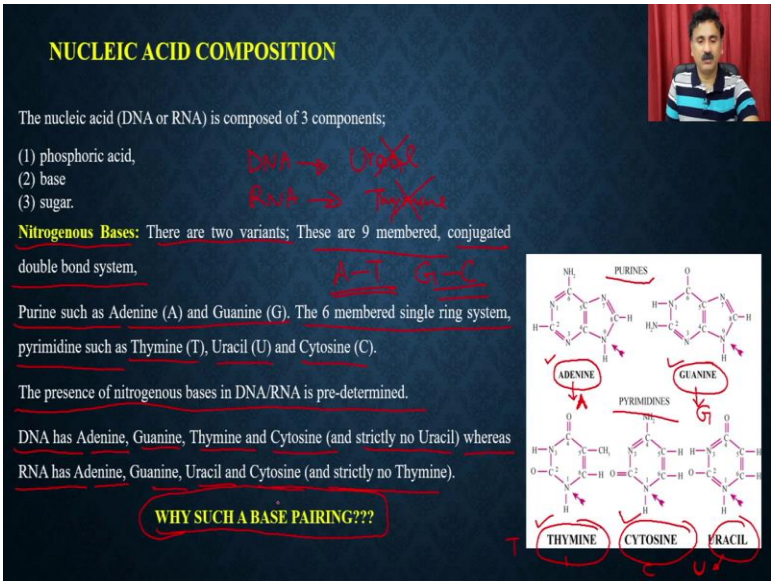
So we have the phosphate backbone. Phosphoric acid serves as the backbone of the molecule. So this is the phosphoric acid what you see. It is actually serving as a backbone. So what you see here is the four points and these points are actually mixing it with the sugar, and that is how it is actually anchoring.

And then the sugar is working as an anchorage point for the nitrogenous bases. And then we have the sugar. We have the 5 membered cyclic reducing sugar present in the nucleic acid. There are two different variants. The sugar molecules which contain the hydroxyl group at the 3 prime are known as the ribose sugar. And it is present… it is known… or otherwise it is known as deoxyribose. So you can have the two different types of sugar. You can have the ribose sugar or the deoxyribose sugar.

So what is the difference? In the ribose sugar, at the 3 prime, at the 3 prime as well as at the 2 prime you are going to have the hydroxyl groups. So, and because of that this is called as the ribose sugar, or the… and whereas in the case of deoxyribose sugar that 2 prime OH is actually missing. And that's how it is called as the deoxyribose sugar. The ribose sugar is present in the RNA whereas the deoxyribose sugar is present in DNA.

Based on the sugar nucleic acid is classified as the RNA or DNA. The ribose sugar is present in RNA whereas the deoxyribose is present in DNA. The purpose of the sugar in the nucleic acid is to provide the attachment point for the nitrogenous bases. So whether it is ribose sugar or the deoxyribose sugar the purpose of the sugar molecule is that it is actually going to provide the attachment point for the nitrogenous bases. So nitrogenous bases are actually going to attach at this point. This OH is actually going to be used in the sugar for attaching the nitrogenous bases.

(Refer Slide Time: 22:41)



Now**,** the third component is the nitrogenous bases. So the nitrogenous bases are, there are two different variants, 9 membered rings, 9 membered conjugated double bonds system which are called as the purines such as the Adenine and Guanine. So you have the two different type of sugar, two different types of nitrogenous bases, the 9 membered conjugated double bond purines such as the Adenine and Guanine.

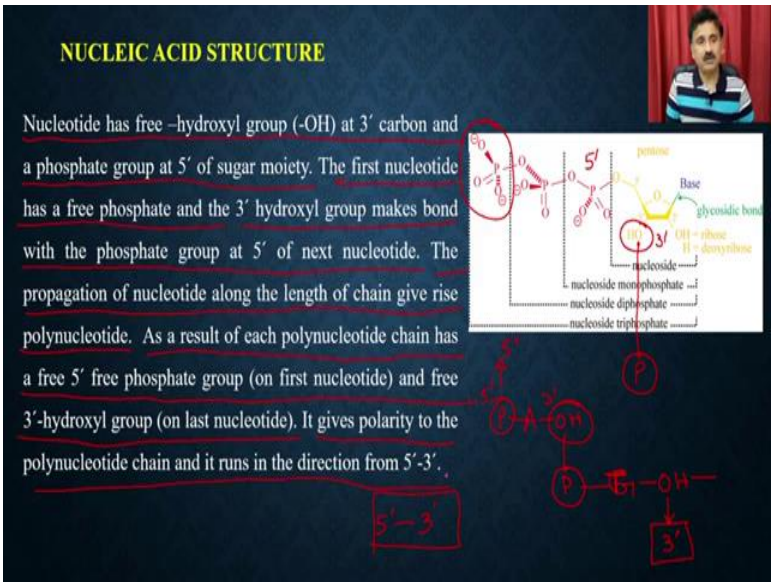
So what you see here is the structure of the Adenine and Guanine, and these are the purines. Similarly you have the 6 membered single ring system, pyrimidines such as the Thymine, Uracil and Cytosine. So this is the 6 membered ring, single ring structure. These are the double ring structures. And here you have, these are the pyrimidines and you have the three pyrimidine molecule, the bases. You have the Thymine Cytosine and Uracil. Remember that Adenine is actually been labeled or written as A, Guanine is written in G, Thymine is written as T, Cytosine is written as C and Uracil is written as U.

The presence of the nitrogenous bases is predetermined, which means it is already been known that which nitrogenous bases are going to be present in the genomic DNA or to the DNA. The DNA has the four nucleotides or the four bases. It has the Adenine it has Guanine it has Thymine and it has a Cytosine. So Guanine, it has the Adenine and Guanine, and it has Thymine and Cytosine, but it does not contain the Uracil.

Whereas, the RNA has the Adenine, Guanine and Uracil and cytosine, and it is strictly no Thymine. So remember that in DNA that you are going to have no Uracil, and in the RNA you are not going to have Thymine. So that you should remember rest all are actually going to be present.

Now these bases are actually making a pair with each other, and the A is making a pair with T whereas G is always making a pair with C. So why such a base pair? That question is actually very important and that question is going to answer when we are going to talk about the base pairing within the A and T and G and C.

(Refer Slide Time: 25:22)



Now let us understand the structure and that also going to give you the idea about the base pairing. So nucleic acid structures, so the DNA is double stranded whereas RNA is single stranded in most of the cases. There are exceptions where you have the double stranded RNA but that is very few cases where you have the double stranded RNA. Mostly the DNA is double stranded whereas RNA is single stranded.

In individual monomer responsible for making DNA or RNA is called as the nucleotide, and as a result the DNA or the RNA can be considered as the polynucleotide. So just like we have the polymeric sugars, we have the fats, polysaturated fats and something like that, similarly you can actually call the nucleic acid, whether DNA or RNA as the polynucleotide because it is actually the summation of the nucleotide molecule.

So what is nucleotide? The individual nucleotide is a nucleoside attached to the one or more phosphate groups and can… so what is nucleoside? So when, you can see here, when the sugar molecule is attached to the base with the help of the glycosidic bond it is actually going to give rise to a compound which is known as the nucleoside. So this is the structure. If you have the sugar attached to the base then that is called as the nucleoside.

If it is going to have the phosphate, so if you have the phosphate and if that attaches with the sugar and it attaches with the base then that is called as the nucleotide. In some cases you can

also call it as the nucleoside mono phosphate. So if it is a 1 phosphate molecule you can call it as nucleoside monophosphate, if it is 2 phosphate then it will be called as the nucleoside diphosphate. And if it is a 3 phosphate then it is going to be called as the nucleoside triphosphate. You can also call it as the nucleotide. So nucleotide could be, whether you have the 1 phosphate group, the 2 phosphate groups or the 3 phosphate groups.

Each nucleoside is composed of the nitrogenous bases attached to the sugar through the glycosidic bond. So this bond is the glycosidic bond which actually links the base to the sugar. Now, once the nucleotides are formed, these nucleotides are also going to be arranged in different fashion. So nucleotides are having free hydroxyl group at the 3 prime end.

So you see the free hydroxyl group which is at the 3 prime end, whereas it has a phosphate group at the 5 prime of the phosphate. So this is a phosphate group at the 5 prime end. The first nucleotide has a free phosphate group, and 3 hydroxyl groups make the bond with the phosphate group at the 5 prime of the next nucleotides, which means you use this particular group. So this is the 5 prime and this is the 3 prime end.

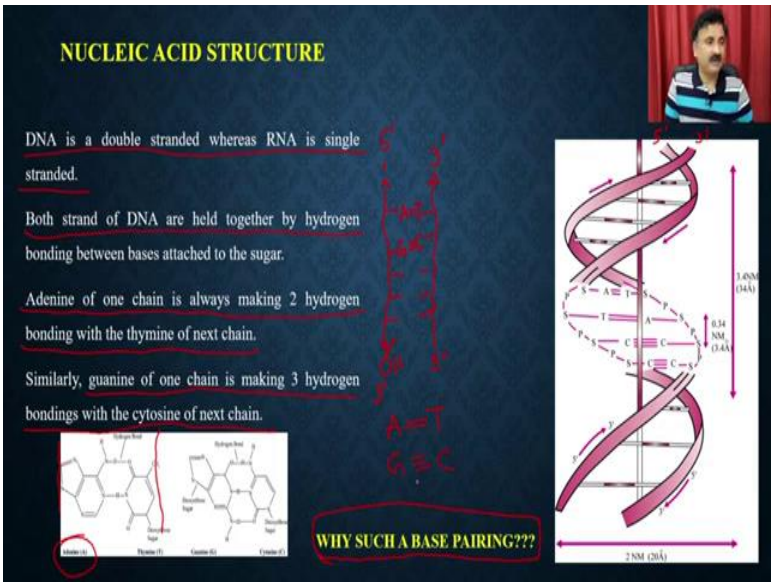
So if it is a first nucleotide it is actually going to use this nucleotide the 3 prime OH group and it is actually going to mix with the phosphate molecules of the next. And that's how there will be a propagation. So you can imagine that if you have the A, the A is actually going to have the OH. This is the 3 prime OH. And then it is actually going to mix with the 5 prime of the next base.

So this is like, for example, you have the T, sorry, you have the G. So that is how it is actually going to propagate. Again here you are going to have OH. So this is the 3 prime end. And this one also is going to have the 5 prime. So it is going to have a 5 prime phosphate. And the 3 prime phosphate, 5 prime, 3 prime OH is actually going to make a bond with the 5 prime of the next strand and that is how that will continue and that's how it is actually going to make a linear chain.

The propagation of the nucleotide along the length of the chain gives rise to the polynucleotide. As a result of each polynucleotide chain has a free 5 prime phosphate group and a free 3 prime hydroxyl group. It gives the polarity to the nucleotide chain and runs in direction of the 5 prime to 3 prime. So DNA is always been called that it is runs in the 5 prime to 3 prime, because you

can have the first nucleotide as the first, this nucleotide, and then it actually makes the bond with the subsequent nucleotide. And that is how the top nucleotide is actually going to be a 5 prime end, and last nucleotide is actually going to have the 3 prime end. And that is why, for conventional purposes the DNA is going to be called that it is running in the direction of 5 prime to 3 prime.

(Refer Slide Time: 30:34)



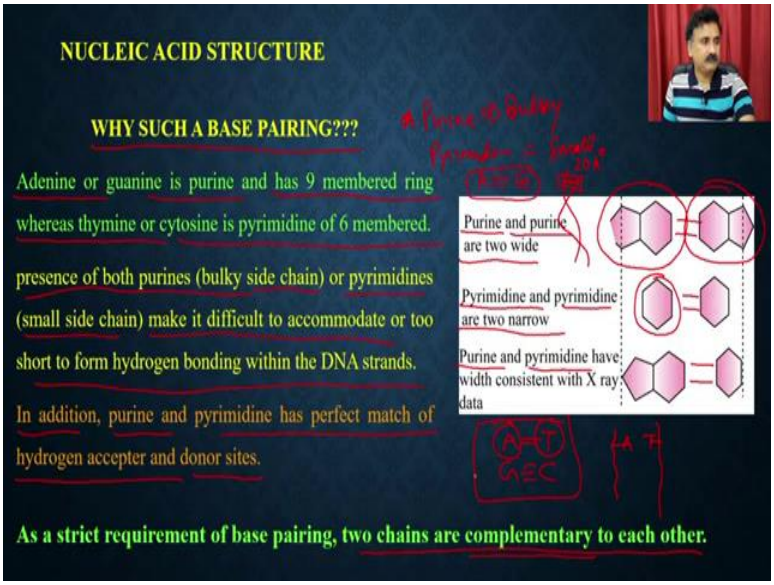
Now, DNA is a double stranded whereas RNA is a single stranded. Both strands of DNA are held together. So you are going to have the two strands in the DNA. So you are going to have a 5 prime end, then you are going to have bases like that, and then you are going to have the OH which is going to be 3 prime end of that last phase. Similarly, you are going to have another strand which runs in the reverse direction. So on this side you are going to have the 5 prime end, and then it runs in this direction and it is actually going to have the 3 prime end. And that is how there will be a base pairing between the basis what are present on to these particular strands.

So if you have the A it is actually going to make the strands. So the Adenine of the one chain is always making the two hydrogen bonding with the Thymine of the next chain. So this is what it is showing. So this is the Adenine what is been attached to one chain. This is the one chain. This is the second chain. And the Adenine what is attached to the first polymer polynucleotide is making a base pairing with the Guanine Thymine what is present on to the second chain. And that is how it is actually making a two hydrogen bonding with the Thymine of the next chain.

So you can imagine that it is actually going to make the double bond, double hydrogen bond with the Thymine. Similarly, if it is having a Guanine then it is going to make the 3 pairs with the Cytosine. So similarly the Guanine of one chain is making the 3 hydrogen bonding with the Cytosine of the next chain. So this is what you see here, that the DNA is having a double helix structure. So it has a double stranded structure, and where you are actually running one strand in the 5 prime, so both the strands are running in the 5 prime to 3 prime.

So one strand is running in the 5 prime to 3 prime direction, and other strand is running in the 5 prime to 3 prime. So for example, this one is actually having a 5 prime on this side and the 3 prime on this side. So this actually it runs like that.

(Refer Slide Time: 32:57)



Now, again we are going to ask the same question? Why such a base pair, why there is a strict rule that A is actually making a pair with T, and G is always making a pair with C. The answer is that since we understand now the structure of the DNA, we could understand there are two strands, two polynucleotide strands which are making a base pairing, and the structure is also very clear.

You have a very double helical structure where you have the width of the structure and as the 20 Angstrom width, and in one round it actually goes from the 34 Angstrom. So that is why the DNA is actually having a very strict structure in terms of the width and in terms of the bases or

the distances between the strand, the distances between the bases. So because of this particular type of restrictions we have a strict base pairing between the A and T.

You can see that the Adenine or the Guanine is purine, and it has a 9 membered ring which means it is actually going to be bulky compared to the 6 membered ring. So whereas Thymine or the Cytosine is a pyrimidine and the 6 membered ring. So Adenine like the purines are actually going to be little bulky, which means they are actually going to be big. You see this. It has two rings. So it has two rings whereas the pyrimidine is actually going to be smaller, going to be small compared to the purines.

So if you keep the two purine molecules, for example, if you keep A and if you keep the G, the purines and purines are actually going to be very wide. So they will actually going to have the steric hindrance is between the two molecules, because you know that the DNA runs as a double helical structure. So width of this strand is already been fixed and the width is 20 angstrom. So it cannot accommodate any kind of distortion.

So if you have the two bases which is purine and purine, the spaces between them is actually going to so narrow so they will not going to perform the bonds and they also not going to accommodate into that. And that is why it has been decided that you can have one which is slightly bigger and one is smaller.

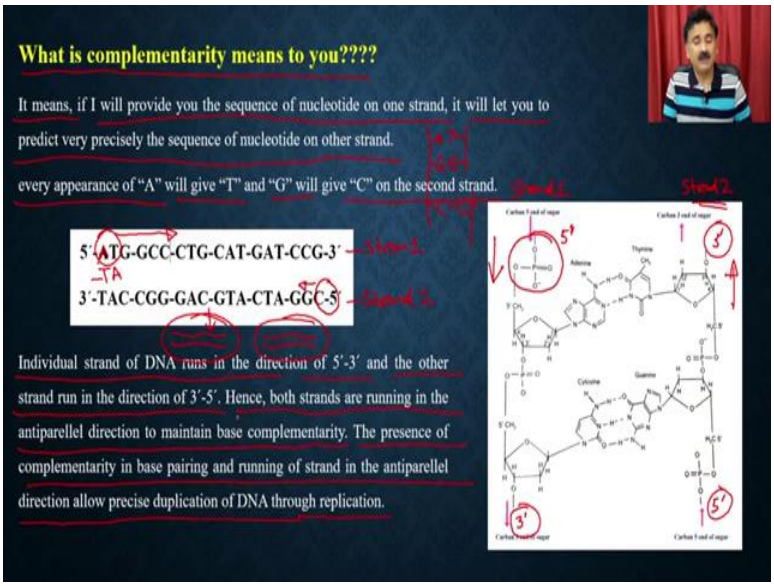
Same is true for if you have the two pyrimidines. If you have two pyrimidine they are actually going to have the very narrow. So if you have the two pyrimidines they are going to be so far away that there will be no hydrogen bonding between them.

So because of that we have a consistent base pairing like where, one side you can have the purine, the other side you can have the pyrimidine. Now presence of the both purine which are actually having the bulky side chains or the pyrimidine which are going to be small side chain makes it difficult to accommodate or too short to form the hydrogen bonding within the DNA strands.

In addition, the purine and pyrimidine has the perfect match of the hydrogen acceptor and the donor sites, which means if you have the purine and the pyrimidine as the bases, not only there would be definite geometry so that they can be able to accommodate into the DNA structure, they are also having the perfect stereochemistry so that they can have the perfect base pairing.

So as a strict requirement of the base pairing two chains are complementary to each other, which means A is always going to make the pair with T, and G is always making a pair with C. And because of that if there is a A it is actually always going to be T. So that is how these kind of phenomena is called as the complementary, which means if you, in one strand you have A, there is no way that you can have any other residue stress. So it can be T only on this side. So that is why this phenomena that you have one A, so it has to be T on the other side, and that kind of feature is called as the complementary to each other.

(Refer Slide Time: 37:17)



Now, what is complimentarity means to you? What is meant by complementarity? It means if I provide you a sequence of nucleotides on the one strand it will let you to predict very precisely the sequence of the nucleotide on the other strand. As I said, if the one strand you have A, the another strand you don't have to tell. Like you know that it is actually going to be T. If you have the G it is actually going to say that it is going to be C. If you say it is going to be C then it is going to be G. It cannot be A, because just now we have discussed that there is a definite base pairing between the A and T. So because of that this is the advantage of being a complementary DNA, complementary sequences.

Now every appearances of A will actually going to give you the T and every G is actually going to give you the C on the second strand. You can see the DNA sequence. So I have shown you a

sequence, like so this is the strand 1 and this is the strand 2. Now even if I do not it give you the strand 2 you can be able to generate the sequence of the strand 2, because everywhere you have the A you can put just put the T. If you have the T you can just put the A. So that is how you can be able to generate the complementary sequence. And that is the advantage of having the DNA structure as the complementary structures.

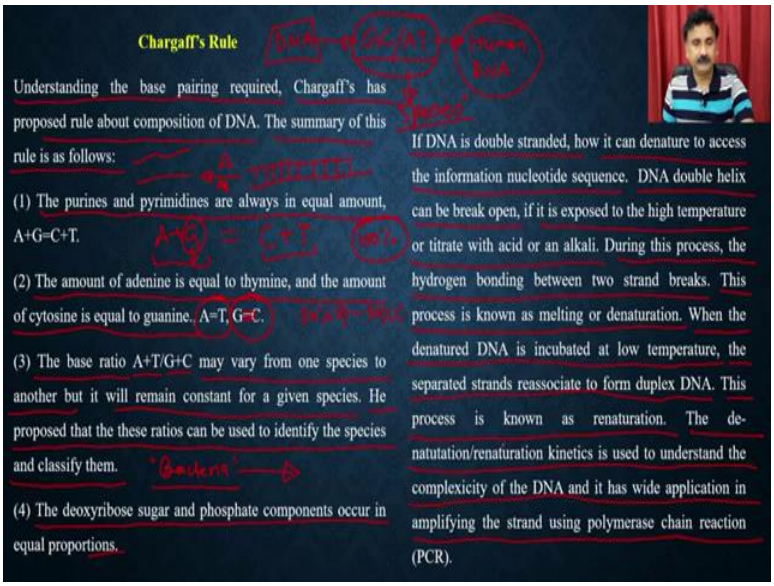
The individual strands of the DNA runs in the direction of the 5 prime to 3 prime, and others strands are going to be run in the direction of the 3 prime to 5 prime, which means if this is the strand 1 this is actually going to start with the 5 prime of the phosphate. So it is going to start with the 5 prime phosphate group and then the first base is actually going to use this 3 prime hydroxyl group and it is actually going to make the bonds would the subsequent bases. And that is how the last nucleotide is actually going to be end up at the 3 prime hydroxyl group.

Same is true for the strand 2, but the strand 2 it is actually going to… strand 2 is actually going to start in the reverse direction. So this one is running in this direction, and this one is running in the this direction. So you can see the 5 prime group is actually having bases with the… it is starting from the first base and then from here it is actually making the bases with the subsequent thing. So that is why you see this is the 5 prime end, this is the 3 prime end. So that is how the DNAs are actually, are running in the complementary sequences, and it also are anti-parallel in directions. Hence both the strands are running in the anti-parallel direction to maintain the base complementarity.

The presence of the complementarity in base pairing and the running of the strand in anti parallel direction allows the precise duplication of the DNA through the replications. So it has an advantage that the genome of any organism can be duplicated can be multiplied without going through with the information on to the second strand.

For example, if this is strand, so we can actually put the two machineries on to this. We can put the two machinery, one machinery on this side and another machinery on this side. And that is how this actually can make one copy, and this also can make one copy. And that is it is actually going to give you the true copy of DNA. And that is the advantage of having the complementary as well as anti-parallel directionality within the DNA structures.

(Refer Slide Time: 41:02)



Now, the additional advantage is that you can actually be able to use this particular type of base pairing rules and you can be able to calculate even the composition of the DNA. So understanding the base pairing required Chargaff has proposed the rule about the composition of the DNA.

The summary of this rule is as follows. The purines and the pyrimidines are always in a equal quantity, which means A plus G is equal to C plus T. So if I give you the amount of G you can be able to calculate the amount of all other properties. The amount of Adenine is equal to the Thymine and the amount of Cytosine is equal to the Guanine, which means, and why is it so? Because A is actually going to make the base pairing with T and G is actually going to make a base pairing with C. So this is 3 bonds actually.

So that is why, as I said, if I tell you, a organism has 30 percent G. That means the organism has 30 percent of C actually. So if it has 30 percent of the C and you know that the total is 100 percent, then this is actually equal to this. This means if I just put these values into this formula I could be able to calculate the A as well as the T as well. The base pair ratio that A plus T by the G plus C may vary from one species to another, but it will remain constant for the given species.

So if you are going to talk about, for example one species, for example if we talk about the humans, so human species is homo sapiens. So this AT ratio, AT by GC ratio could vary from

one organism to another organism because it depends on the complexity of that particular genome. But it could not be varying within the species.

So, for example if I have isolated a DNA and if I do not know the source of this DNA what I can do is I can just calculate the GC by AT ratio, and I can be able to tell that whether it is a human DNA or not, because I can be actually able to match that with the GC by AT ratio for the human samples. Same is true for the other organisms also.

Suppose I got a sample which is from the dog or from the cow, because this ratio is not going to vary within the spices. It could vary from the species to species. So it is actually a clear identifier for that particular species. And that is how the people are using this kind of analysis in the DNA-based identification processes.

He proposed that these ratios can be used to identify the species and classify them. That is why, if you see like, for example the people are identifying a new bacteria. What they are doing is they are doing these kind of analysis to know whether this bacteria belongs to a existing bacteria what they have in the system or not. If it is not then they will say, oh, they have identified a new bacteria. And then they can be able to classify them, and according to some more tools they can be able to put them under the appropriate groups and appropriate class actually.

The deoxyribose sugar and the phosphate component occurs in the equal proportion. So the G and C only varies, but the sugar and the phosphate component which are actually the building blocks of the DNA or RNA is not varying between the two species.

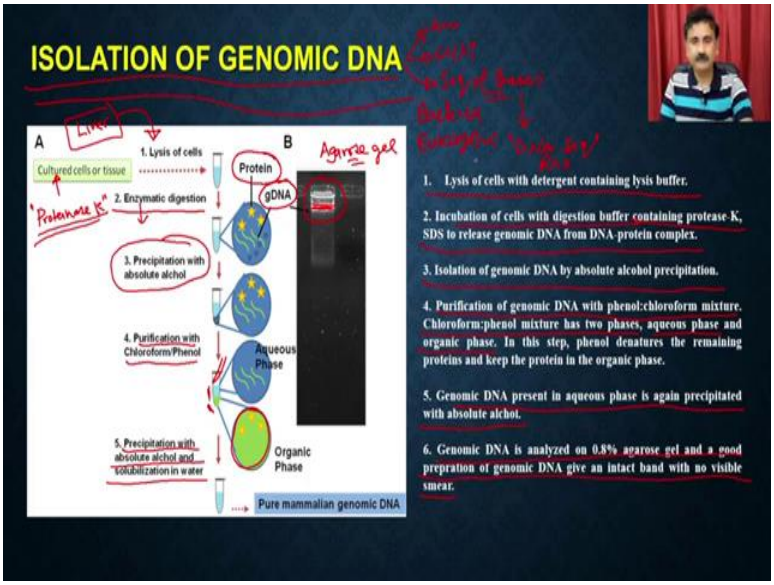
If the DNA is double stranded how it can be denatured to access the information of the nucleotide sequence? The DNA double helix can be break open. So do you know that DNA is double helix? So the question is how actually you can break this DNA so that you can be able to extract this information. DNA can be break open if it is exposed to the high temperature or titrate with the acid or alkali. So there are definite treatments which you can do actually.

So if you take this DNA and if you heat it you can be able to break the strands. If you add the acid or if you add the alkali then also you can be able to break because if you put the acid and alkali it is actually going to destroy the hydrogen bonding between the bases, and that is how it is

actually going to get broken. During this process, the hydrogen bonding between the two strands breaks. This process is known as the melting or the denaturations.

When the denatured DNA is incubated at lower temperature the separated strand reassociate to form the duplex DNA. This process is known as the renaturation. So you have the denaturations. If you heat up the DNA then it is actually going to denature, and if you lower down the temperature it is actually going to come back because you know that A is having a complementarity with the T, and G is actually going to make the base pair with C. So that's how they will come together. Denaturation and renaturation kinetics is used to understand the complexity of the DNA and it has the wide applications in amplification of the strand using the polymerase chain reactions.

(Refer Slide Time: 46:53)



Now, once you have understood about genomic DNA we can also understand about the isolation of the genomic DNA, because then you can be able to understand how you can be able to utilize this genomic DNA for the subsequent reactions. Just now we discussed you can be able to calculate the GC by AT ratio and you can be able to identify the species. For that purpose you first have to identify the genome, first you have to isolate the genomic DNA from that particular organisms. So if you want to isolate the genomic DNA, whether it is from the bacteria or whether it is from the eukaryotic cells, first step is that you have to break the cell. So you have to lyse the cells with the detergent containing the lysis buffer.

So what you have to do is you have to take the cells and you have to incubate it with the lysis buffer. And that what is going to do is lysis buffer is actually going to lyse the cells. If you are going to start with the tissue, like for example if you are starting with the liver then you might have to digest that liver with also with the lysis buffer. Then you incubate, incubation of the cell with the digestion buffer containing the protease K SDS to release the genomic DNA from the DNA protein complex.

Then you incubate that with the enzymatic digestions. That enzymatic digestion is actually having the composition where you have an enzyme which is called as the protease K. That protease K is actually going to chew up the protein part. And that's how it is actually going to release the genomic DNA from the DNA protein complex.

Then you isolate the genomic DNA by the absolute alcohol precipitation. So then what you are going to do, is you are going to .precipitate the genomic DNA with the absolute alcohol and that is actually going to give you the DNA which is still be having a contamination. So what are the contamination you are going to have? You are going to have the contamination of the proteins. You are also going to have the genomic DNA.

Now next topic is, or next task is that we should remove the protein part because we want to have the pure genomic DNA from the cell. So then what we are going to do is we are going to incubate that with the chloroform phenol and isoamyl alcohol mixture. So the purification of genomic DNA with the phenol chloroform mixture, so phenol chloroform mixture has two phases. It has the aqueous phase and it has the organic phase. So when you extract this DNA which contains the protein with the phenol chloroform mixture it is actually going to give you the two phases. You are going to see the aqueous phase, and then you are going to see the organic phase.

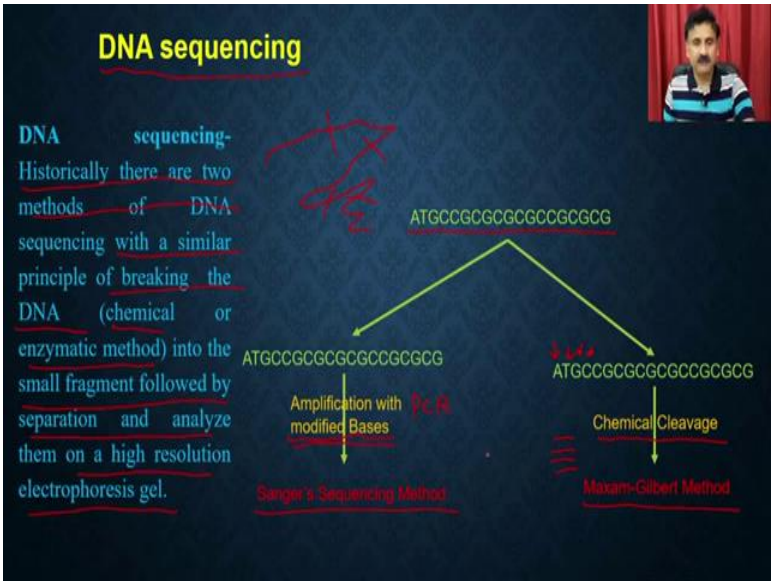
So what is there in the organic phase? All the proteins are actually going to become in the organic phase because they will get precipitated whereas the DNA is actually going to be present into the aqueous phase. So now what you have to do is, very carefully with the help of the pipette you have to collect the aqueous phase and you have to put it into the new Eppendorf, and then you have to use the precipitation.

So in the step 5 you have to precipitate that with the absolute alcohol and you have to solubilize the DNA into the water and that is how you are going to get the pure genomic DNA. So once you start the pure genomic DNA you can be able to run that on to the agarose gel, and then it is actually going to give you a band. So what you see here, this white color stuff is actually the genomic DNA.

One of the classical feature of genomic DNA is that it is very heavy so it will not run very fast. It is actually going to remain very close to the well. So the genomic DNA present in the aqueous phase is again precipitated with absolute alcohol and the genomic DNA is analyzed on the 0.8% agarose gel, and a good preparation of genomic DNA give an intact band with no visible a smear.

Now once you isolated the genomic DNA you can be able to do many things. You can do, as I said, you can do the annealing temperature curve, you can do the GC versus AT ratio calculations you can do even to know the sequence of the bases. How can do the sequence of the bases? How you know that what are the sequence of the nucleotides? So for that you have to put the DNA into the DNA sequencing reactions.

(Refer Slide Time: 51:45)



Let us understand about the DNA sequencing reactions. Now, once if you want to do the DNA sequencing, for example this is the DNA what you have and you want to do the sequencing. You

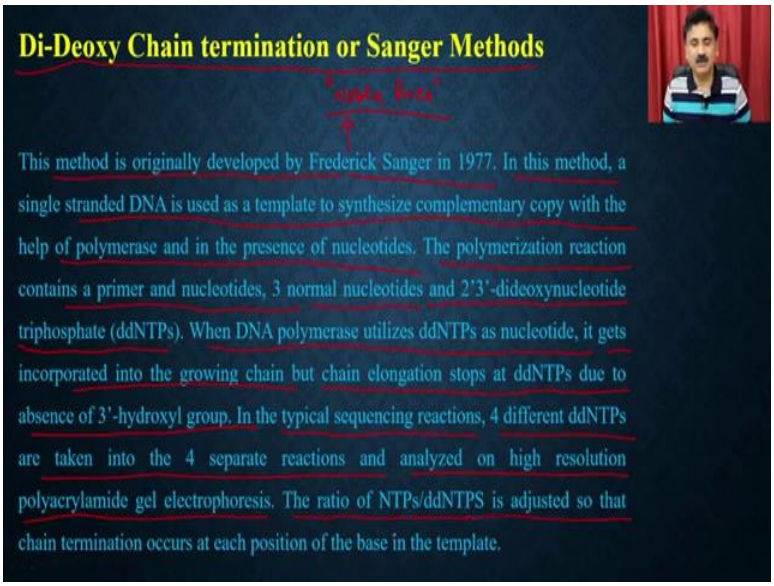
have the two options. So historically there are two methods of the DNA sequencing with a similar principle of the breaking the DNA.

So first you have to do is you have to first, because you know, it is not going to be a single stranded DNA, it is going to be a double stranded DNA. So first you have to break the DNA either by the chemical method or the enzymatic method into the small fragments and followed by the separation and analysis into a high resolution electrophoresis gel.

So you have the two methods. One, you are going to break this into the two parts. So, for example if you have complete genomic DNA what you have to do is, first have to break the DNA. And then you get the small fragments. And then in the method 1 you can amplify this with the help of a PCR and while you are doing the amplification you can use the modified bases. So when you do the modified bases, at those base sites it is actually going to, not going to amplify beyond that. And that's how you are actually going to do the sequencing. And that method is called as the Sanger's sequencing method.

In other method what you can do is you can you can use the base-specific chemicals. So they are going to actually do the cleavage. For example you can have the base specific cleavage for A, T, G and C and therefore it is actually going to break the DNA accordingly. So if you use that when you are actually going to analyze them on a high resolution electrophoresis gel, then you are going to get the many fragments and those fragments, if you do the analysis of those fragments it is actually going to give you the sequence of that particular genome, genomic DNA. That method is called as the Maxam Gilbert method. So if you use the modified bases the method is called as the Sanger’s sequencing method. If you use the chemical cleavage method then it is called as the Maxam Gilbert method.

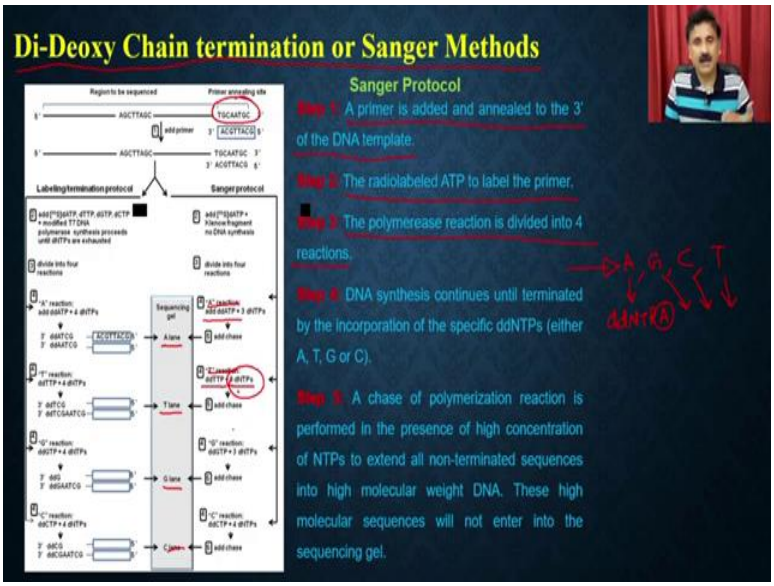
(Refer Slide Time: 53:49)

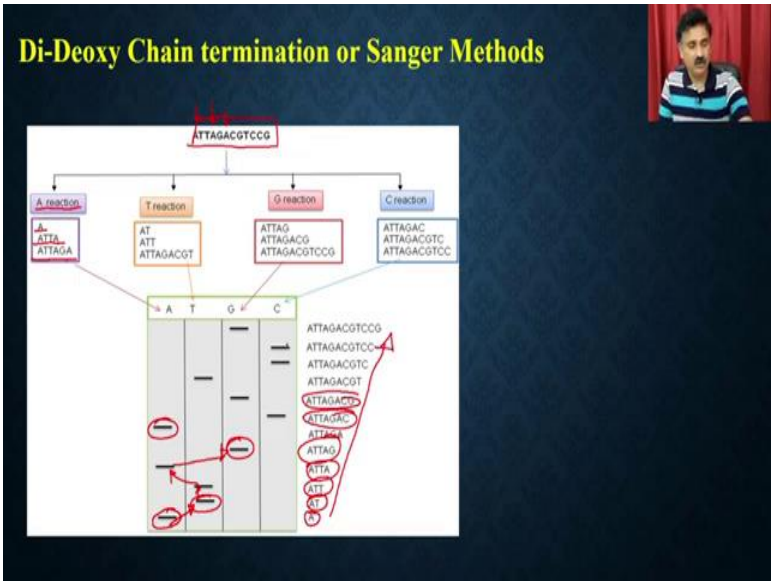


So what is the Sanger's method or the di-deoxy chain termination methods? So this method is originally been developed by Fredrick Sangar in 1977, and for this method Frederick Sanger has got the Nobel Prize for discovering the method to sequence the genomic DNA or the DNA. In this method, a single stranded DNA is used as a template to synthesize the complimentary copy with the help of polymerase in the presence of the nucleotides. The polymerization reaction contains a primer and the nucleotide. The 3 normal nucleotides and another one is a modified nucleotide which is called as the 2 prime 3 prime triphosphate.

When the DNA polymerase utilizes ddNTPs as nucleotide it gets incorporated into the growing chain but the chain elongation stops at ddNTPs due to the absence of the 3 prime hydroxyl group. In the typical sequencing reactions, the four different ddNTPs are taken into the four separate reactions and analyzed on to the high-resolution polyacrylamide gel electrophoresis. The ratio of the NTP and ddNTP is adjusted so that the chain termination occurs at the each position of the base in the template.

(Refer Slide Time: 55:09)





So what you are going to do is if you want to do the Di-deoxy or the Sanger method, first is that you are actually going to take the DNA. And then you are actually going to synthesize the primer. This primer you have to do the radio label. So this primer has to be radiolabeled so that you can be able to monitor, so you can be able to visualize this under the autoradiogram. So a primer is added and annealed to the 3 prime of the DNA template.

So first you add the primer. And then the radiolabeled ATP is there to label the primer. So you are going to use the labeled primer. And then the polymerase reaction is divided into the four

reactions. So you are going to have the four reactions. The reaction for A, reaction for G, reaction for C and reaction for T. When you say A reaction which means all other nucleotides are going to be normal. But it is actually going to have the ddNTP which is actually going to be A. Same is true for G. Same is true for C. Same is for T.

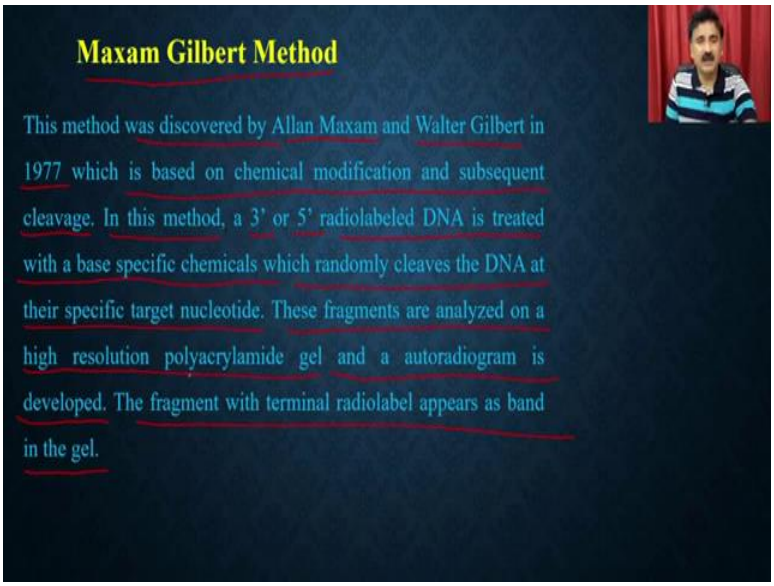
So what you have is you are actually going to divide the reaction into the four reactions, A reactions where all other nucleotides are normal but it is actually going to have the dd, di-deoxy ATP actually. Similarly you are going to have the T reaction. So it is going to have the di-deoxy Thiamine triphosphate. But it is actually going to have the all other 3 nucleotides normal. Same is true for the G reaction, and same is for the C reaction.

So if you perform all these what will happen is that wherever you have the A, it is actually going to stop the growth of that DNA synthesis. Same is true for the T, G and C. And then what you do is you take these reactions and analyze onto the high-resolution gel.

This is what it is going to happen. If this is the sequence what you are analyzing for the sequencing it is actually going to give you the A reactions, T reactions, G reactions and C reactions. So what you see here is that in the A reaction all the places where you have A, like for example, here you have A, you have here A, you have A, here you have, and then you don't have A. So it is actually going to terminate. So it is actually going to give you the fragments like A, ATTA, ATTAGA like that. So if you analyze them onto the polyacrylamide gel or high resolution gel. Similarly this is the T. What will happen is that, these are the A reactions, and these are the T reactions and so on.

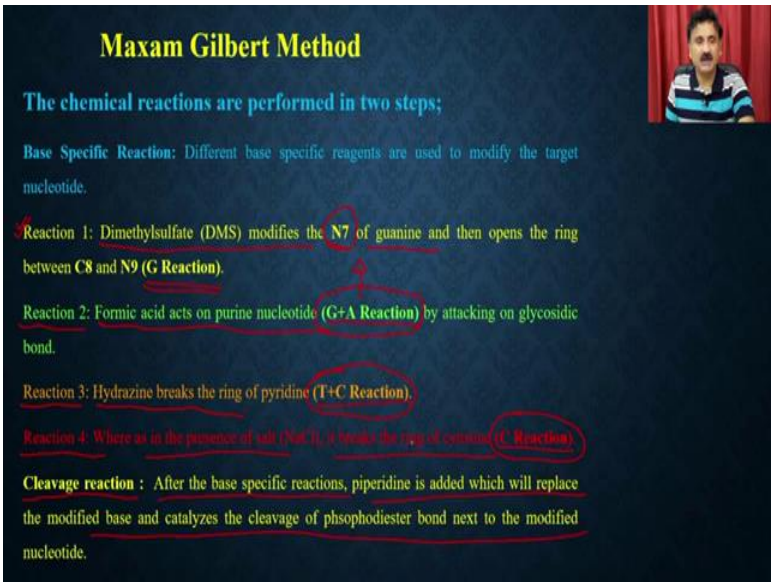
So what you have to do is you have to read these reactions into the reverse order. You have to do like this, then you have to do like this, then you have to go like this, then you have to go like this, then you have to go like this and that. So if you go by this, you go in the reverse direction and it is actually going to give you the fragments. This is the fragment what is happened, so this is the G, it is terminating at G. This is A, it is terminating at A. Then you have C, this is C. Then you have G, and something like that. If you read this in the reverse direction ultimately you are going to get the sequence of that particular genome, the DNA.

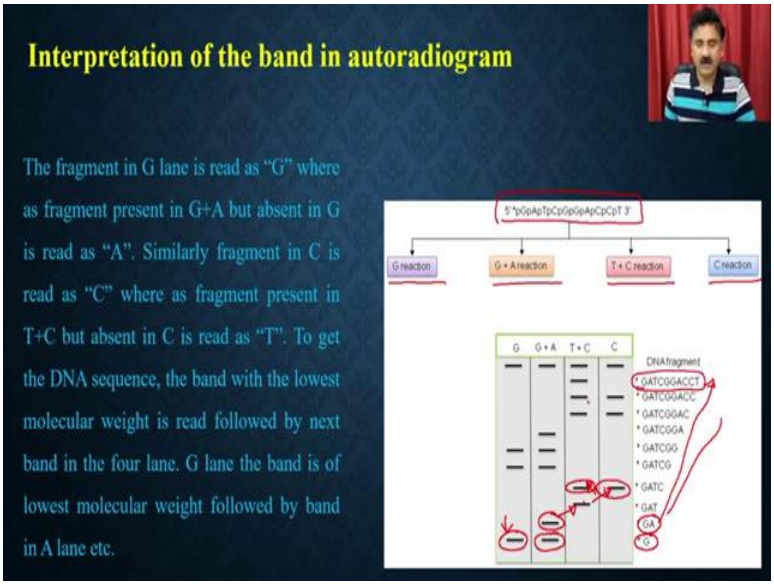
(Refer Slide Time: 58:36)



Now let us move on to the next method. And next is called as the Maxam Gilbert method. This method was developed by the Allan Maxam and Walter Gilbert in the 1977 which is based on the chemical modification and the subsequent cleavage. In this method a 3 prime or the 5 prime labeled DNA is treated with the base specific chemicals which randomly cleave the DNA at their specific target nucleotides. These fragments are analyzed onto a high-resolution polyacrylamide gel and a autoradiogram is developed. The fragment with the terminal radio label appear as a band in the gel. So what you are going to do is you are going to do the reactions for the A, T, G and C.

(Refer Slide Time: 59:22)





So these are the four reactions you are going to do. You are going to do the reaction number 1 where you are actually going to add dimethyl sulfate or DMS, and that is actually going to modify the N7 of the Guanine. So that is, it is going to be called as the G reaction. So this reaction 1 is actually going to modify the Guanine.

Similarly you have the reaction 2 where you are going to have the formic acid and that is actually going to attack the purine nucleotide. So it is actually going to be called as G plus A reactions. Similarly you have the reaction number 3 where you are going to use the hydrazine and that is

actually going to break the pyrimidine and that is how it is actually going to be called as T plus C reactions. And then subsequently you have the reaction 4 where in the presence of the salt NaCl it Breaks the ring of the Cytosine. And that's why it is called as the C reactions.

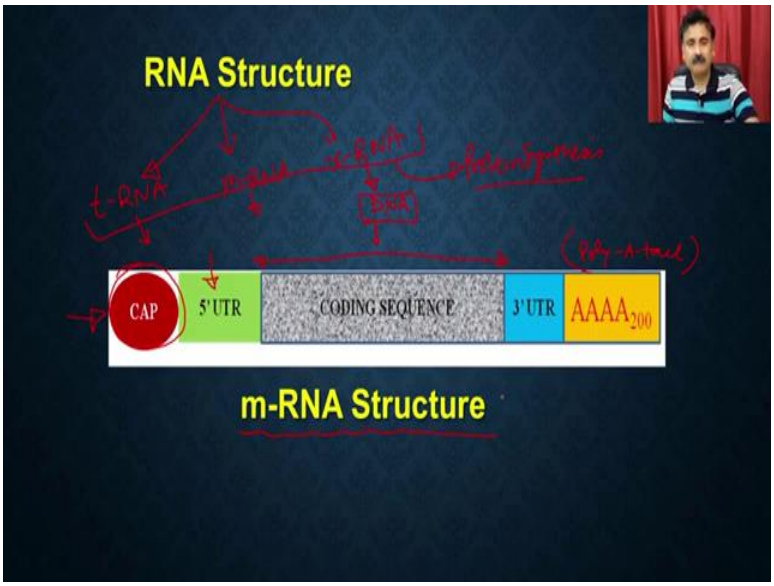
After that you have to do the cleavage reactions. After the base specific reactions piperidine is added which will replace the modified bases and catalyze the cleavage of the phosphodiester bond basis next to the modified bases, and that is how it is actually going to give you the fragments; fragments which contain the G reaction, fragment which contains G plus A reaction, the fragment which contains G plus C reaction and the fragment which contains the C reaction.

So if you analyze them onto a high-resolution polyacrylamide gels the pattern will look like this. So you are going to have the G reactions, G plus A reaction, TC reaction and C reaction. So if you started with this DNA sequence what you are going to get is you are going to get this. So if you have the two bands which are very close to each other or of the same locations then you are actually going to read the G not the G plus A because it is already been present here. So that is how you are going to read the G. The same fragment you would see here is also been present again. So that is why you read the G plus A.

So you go actually in the reverse direction again just like as we have discussed in the Sanger's method. So you go like this and if it is a T plus C and this, so then you are going to read the C first and then you are going to read this. So that is why if you go in this direction, if you go in the reverse direction you are actually going to be end up with the sequence of the DNA fragment which you want to sequence in these particular reactions. So this is all about the sequencing of a DNA.

Now we are not going to discuss about the RNA because RNA is also following the similar kind of structures. It also is made up of the nucleotides. And it is a polynucleotide. The only difference between the RNA from the DNA is that the RNA is single stranded whereas the DNA is double stranded.

(Refer Slide Time: 62:18)

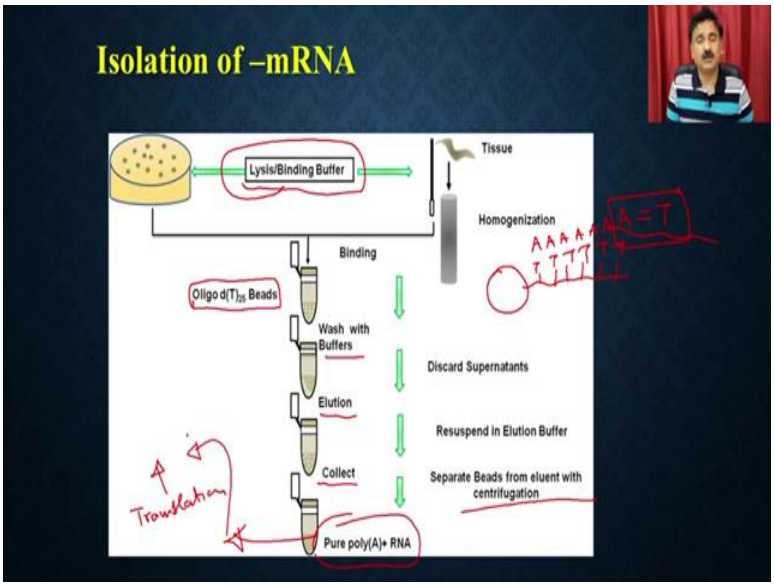


So when we talked about the RNA, the RNA could be of 3 different types. RNA could be the t-RNA. So this is the three different species of RNA. You have you have the t-RNA. You have the messenger RNA. And then you have the ribosomal RNA and all these three RNAs species are extensively been involved into the protein synthesis, where the messenger RNA is actually going to carry the information from DNA, and whereas the t-RNA and the ribosomal RNA are actually going to interpret that information and that's how they are actually going to participate into the DNA synthesis.

So if you see a typical structure of a messenger RNA, what you see here is it has 5 prime cap. This 5 prime cap is actually going to protect the RNA from getting the degradation by the RNases and all other kind of degrading enzymes. Then we have the 5 prime UTRs. This 5 prime

UTR is important for the binding of the factors which are responsible for the protein synthesis. And then you are going to have coding sequence. This is a sequence which is going to be complementary to the DNA sequence. So this is the sequence what you have been synthesized from that DNA. And then you are going to help the 3 prime UTRs and then after that you are going to have the poly A tail. So this is the typical structure of the messenger RNA. And since this has this kind of structure it can be helpful in terms of the isolation of the messenger RNA from the cell.

(Refer Slide Time: 64:03)



How you going to isolate? The first step is that you are going to lyse the cells. We have already discussed this step when we were discussing about the DNA. So we are going to homogenize in the case of the tissue or the hard material. Then what you are going to use is you are going to use a column which actually has the poly T beads. So poly T beads are nothing but a bead which actually has a strand. On this strand you are actually having the T nucleotides. So if you have a T nucleotide and you know that A is always making a base pairing with T, so that is how it is actually going to make the pair with T.

So if you have this kind of strands what will happen is that all the A what are present into the RNA is actually going to bind this, and that is how all the RNA what is present in that particular cell is actually going to bind these beads. And then you have subsequent step. You are going to do the washing with the wash buffer. And then you are going to do the elusions. Once you do the

elusions the RNA is going to be eluted. And then you can collect, separate the beads from the eluent after the centrifugation and that is how it is actually going to give you the pure RNA. Once you got the pure RNA that you can use for many applications. You can use for the using it for the translation studies. You can use that for preparation of the cDNA libraries. You can use that for many other applications.

So this is all about the nucleic acid. What we have discussed? We have discussed about the structure of the DNA. And we have also structure, very briefly we have also discussed about the structure of RNA. Not only that we have also discussed about how you can be able to sequence the nucleic acids. So we have discussed two methods. We discussed about the Sanger's method and we have also discussed about the Maxam Gilbert method.

We have detail lectures and all other aspects which are available on to the MOOCs course. So if you are interested and you would like to discuss or study these processes in detail you can be able to go through some of the other MOOCs courses for the DNA synthesis and other kind of aspects. So with this I would like to conclude my lectures. In our subsequent lecture we are going to discuss the other biomolecules what is present in the cell.

So with this I would like to conclude my lecture here. Thank you.