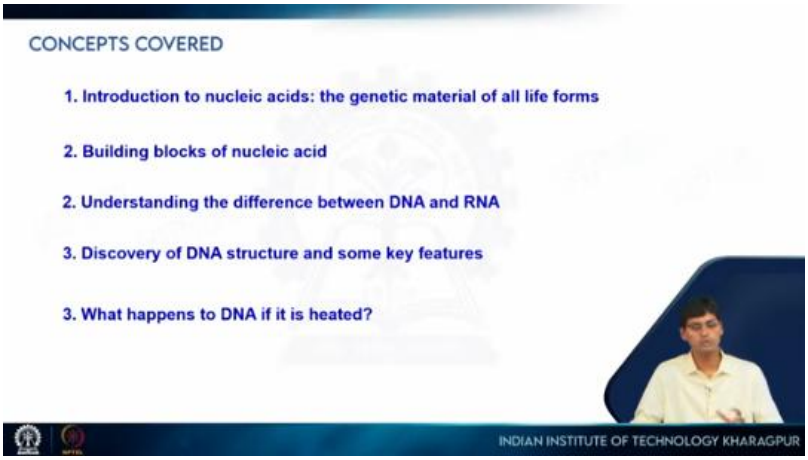
Introduction to Complex Biological Systems
Professor Dibyendu Samanta and Professor Soumya De
Department of Bioscience and Biotechnology
Indian Institute of Technology, Kharagpur
Lecture 3
Chemical and Physical Properties of Nucleic Acids

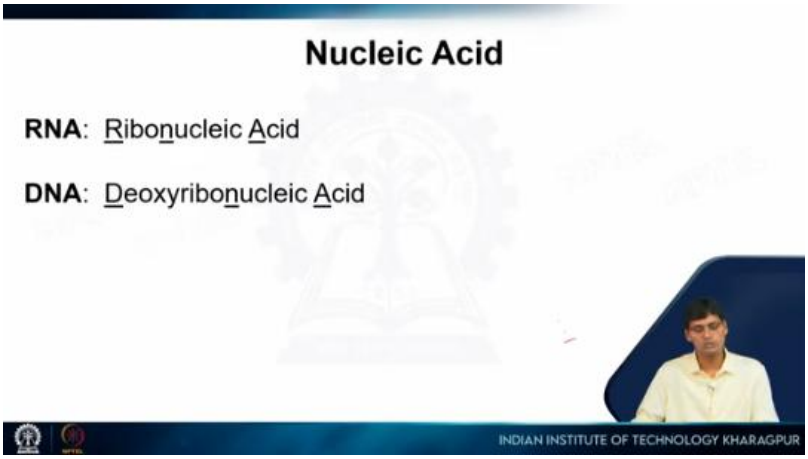
Hello everyone, I am Dibyendu Samanta from IIT Kharagpur. So this is our third lecture on the chemical and physical properties of nucleic acid. During the last lecture, we discussed genetic material, particularly three classic experiments that proved that DNA is our genetic material, namely Griffith's experiment that proved the transforming principle or concept of transforming principle, followed by Avery, MacLeod and McCarty's experiment that clearly showed that DNA is our genetic material. Even more, Hershey and Chase experiment convincingly proved that DNA is our genetic material.



Today, we will mostly discuss DNA or nucleic acid. So here, these are the major points I would like to cover. The introduction to nucleic acids, which is our genetic material, and then building blocks of nucleic acid, followed by understanding the difference between two major nucleic acids, that is DNA and RNA, and then discovery of DNA structure and some key features, followed by some physical properties of DNA, particularly what happens to DNA if it is heated.

So nucleic acids, all of us are aware about gene. So gene is, for example, if one kid is doing very well in academics and there are many people in the same family. They are also very good in academics. Many times we say this is in their genes. So genes are very common to all of us.

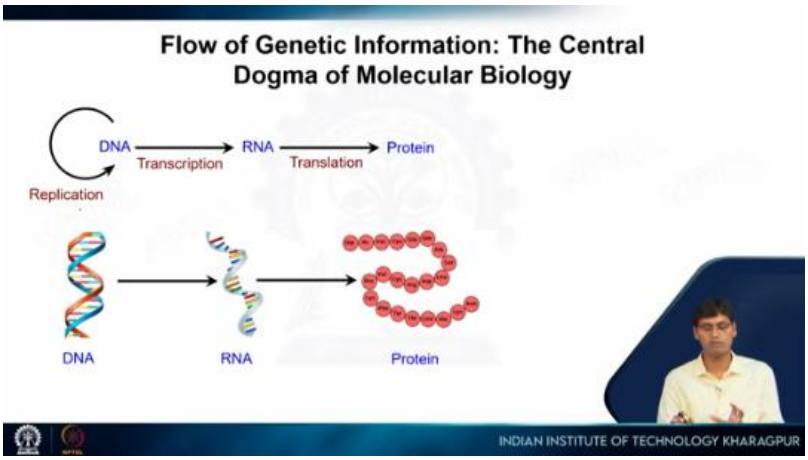
But what are the chemical properties of genes that were not known before. So from those three classic experiments, we came to know that gene means DNA in this case. Now, here in particular you can see that RNA and DNA. So, RNA stands for ribonucleic acid and DNA stands for deoxyribonucleic acid. So, what is nucleic acid?



So as I said that gene was known a long time ago but nucleic acid also before establishing that DNA is our genetic material was known almost in 1970's or 80's. During that time scientists discovered some acidic substances present in the nucleus. So that's why this name came, the nucleic acid, some acidic substances present in the nucleus. So then many other scientists, they actually did a lot of experiments and finally, they came to know what is present inside nucleic acid.

We know that this is made up of nucleotides. But this way, nucleic acid is first discovered, then ribonucleic acid and deoxyribonucleic acid. Although we discuss that DNA is our genetic material, I should tell here that RNA can be genetic material in a few cases.

For example, some viruses have RNA as their genetic material. So all of us know during the COVID period that whatever we are talking about, this SARS-CoV virus or Coronavirus, so they have RNA as their genetic material. But this is not frequent. Only a few viruses have RNA as their genetic material. Now, we are going to discuss a little bit about the central dogma of molecular biology.



This is very important theme in molecular biology that I also explained before that DNA goes to DNA. That means it can replicate itself and that is called replication followed by transcription.

So whatever is written inside the DNA, it can actually form other polymers of nucleotides. That is called RNA and the process is called transcription. So, here transcription is done by the major enzyme called RNA polymerase and replication is done by DNA polymerase.

Many others enzymes and factors are also involved, but those are the special enzyme that carry out this thing. And now from RNA finally we get protein. So, information is present in DNA and proteins are actually the molecular machines which work for us. So, these are known as molecular machines. They actually work for us, but the instructions are present in DNA.

And RNA is present in between DNA and protein. So, it helps to get protein. So, this process from where we are getting the protein from the RNA. Particularly, I would say in this case, we should say this is mRNA.

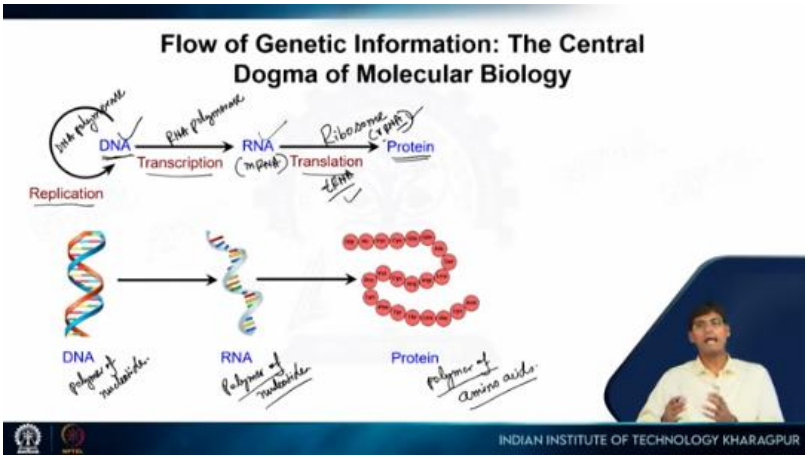
This is messenger RNA and it gets translated and it forms protein. So, this step is guided by two major things. One is ribosome ribosome. So ribosome is made up of many different types of proteins as well as nucleic acid particularly rRNA ribosomal RNA and also another important nucleic acid that is tRNA or transfer RNA that also play an important role in this process. So, now if we see here in the ribosome, we have rRNA ribosomal RNA. So, if we see, although proteins are the molecular machine that works for us, but the thing is DNA and RNA which is here I am saying mRNA or messenger RNA followed by ribosomal RNA and tRNA all of them are involved to get the protein. So nucleic acids actually play an important role in this in getting the protein out of the information. So it's like a huge thing in our body and we should know about nucleic acids in detail, like what are the major properties of nucleic acid.

So I'm going to discuss the very basic and fundamental thing about nucleic acid that will be applicable both for DNA as well as RNA. Wherever some differences are there, I will also explain that. Here, I should also mention why those terms are like replication, transcription and translation. So replication means just to make its own copy, the same thing.

So that's why it is called replication, the replication of DNA. Now, transcription means that some information is present somewhere and then you are just copying that information again. So, you are writing that information in some other places. That is all. For example, we sometimes say to get the certificate or mark sheet, we say the transcript. So, that is the same thing. So, here DNA is also a polymer of nucleotides, RNA is also a polymer of nucleotides. So, basically their language is the same but only we are rewriting the information which is present in DNA in a different form so that it would help to make the protein.

So, after DNA to RNA that means when the transcription is happening, three major RNA we get one is mRNA or the messenger RNA, which is the template for protein synthesis and then we have two different major groups of RNA also. One is ribosomal RNA and the other one is transfer RNA. So, all these three RNAs are formed from DNA but all of them are polymer of nucleotides.

But now from RNA we will get the protein. So, proteins are the polymers. So, they are polymers of amino acids. So now you see the language has completely changed here, until now DNA and RNA are polymers of nucleotides but now this is polymer of amino acid for example. If some of us say for example that I am going to France but I don't know the French language then what do I need? I need a translator. So that person will actually help me to translate the language.



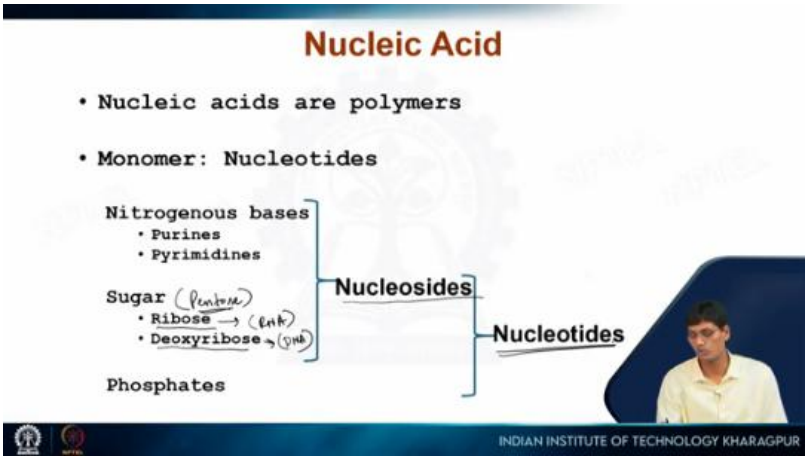
So similarly here, at the end, whatever we are getting the protein, it is not made up of nucleotides. The language has been completely changed. This is the polymer of amino acids and that's why this is called translation.

So now. So we just discussed the basic, central dogma of molecular biology but there are some updates and a little bit of addition here and I will discuss that in the right place. Now, I would mostly concentrate on nucleic acids. So those are the polymers of nucleotides. Now, if we see what is present in nucleic acids, the first thing is nucleotides.

But what is present in nucleotides? So in nucleotides, we have three major things. One is nitrogenous bases and nitrogenous bases can be of different type. It can be purines. It can be pyrimidines.

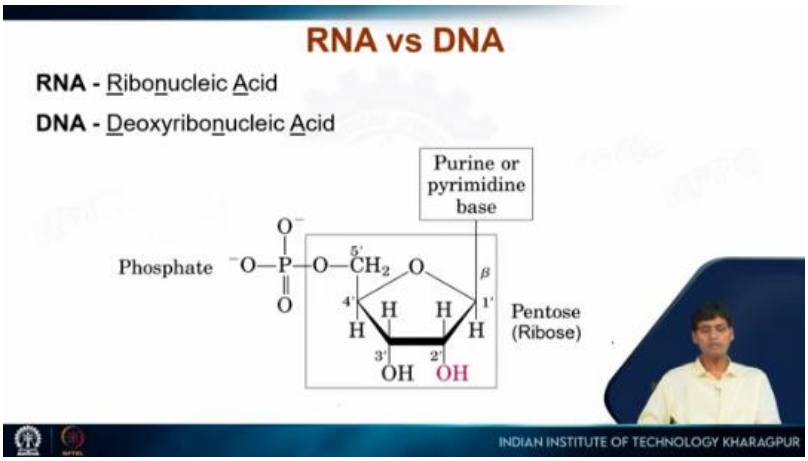
I will explain soon. So like pentose sugar is present in nucleic acid. So this sugar and nitrogenous base together forms nucleosides. So pentose sugar means 5 carbon sugar present in nucleic acids. So it can be ribose or it can be deoxyribose. Ribose sugar is present in RNA or ribonucleic acid and deoxyribo sugar is present in DNA and nitrogenous base and sugar together we say this is nucleoside.

Now at the beginning I told that nucleic acids are the polymers of nucleotides. So, now if we add a phosphate group with the nucleosides then it will be nucleotides. So, this is nucleotides, this is the monomer of nucleic acid.



So, we will discuss more about this now and now about RNA versus DNA. What are the major compositions or differences in composition in case of DNA and RNA? So, now, the

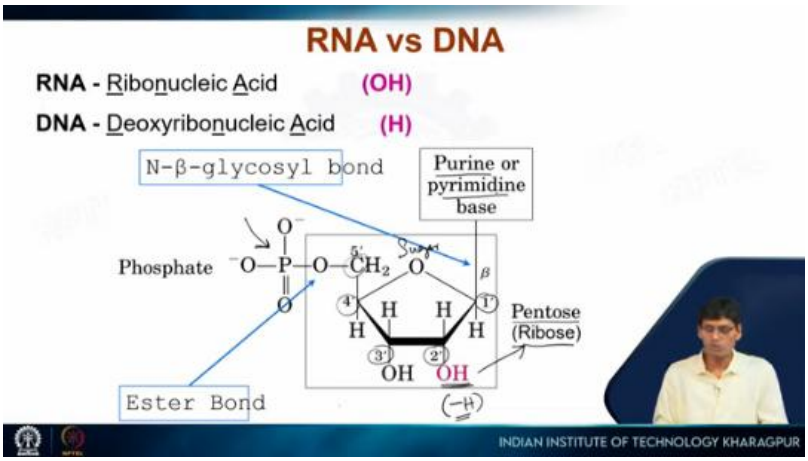
major thing, the major difference is the sugar group, like whatever the sugar moiety present. As I already mentioned that pentose sugar or the 5-carbon sugar is present in both DNA and RNA. But in the case of ribonucleic acid, it is ribose.



So, what is the difference between ribose and deoxyribose? So, the name itself suggests that deoxy. So, somewhere they lack oxygen. So, if you see here, this is pentose sugar and this is carbon number 1, 1 prime carbon, 2 prime carbon, 3 prime, 4 prime and 5 prime. So, this is 5 carbon sugar.

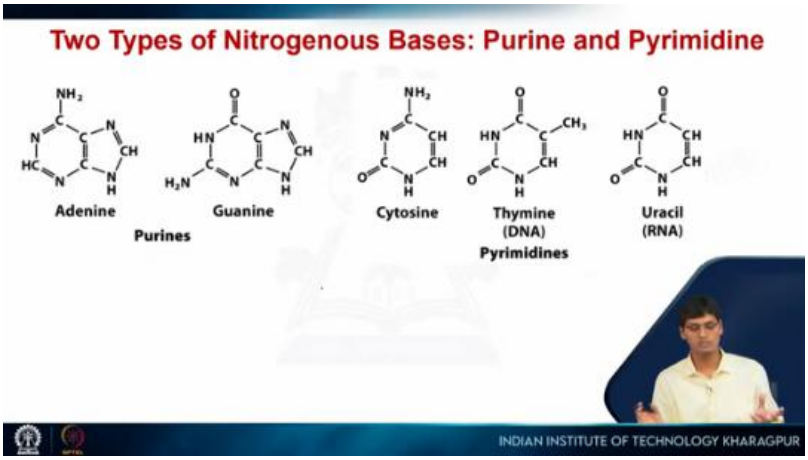
And now in the two prime positions, you can see here OH is present. So OH is present in case of ribose sugar, which is present in RNA. But in the case of DNA, instead of OH, it should be just H. So oxygen should not be there. So that's why deoxyribonucleic acid. So that's the difference.

And now this is just sugar, but Now in terms of nucleotides if you see we have here sugar and then here the nitrogenous bases it can be purine or pyrimidine and also it has the phosphate group. So together it will be nucleotides. Now, I should also explain one, I know, small thing here, but which is very important. When we are discussing DNA or RNA structure, that is why I am saying that 1 prime, 2 prime, 3 prime, why not 1, 2, 3, 4, 5, or something like that. The major reason is here, in the case of purine and pyrimidine, in that case also it has multiple carbon atoms. So just to distinguish between these two from nitrogenous bases i.e. from purine, pyrimidine to sugar, we are saying that 1 prime, 2 prime, 3 prime so that we can discuss in a better way. So the carbon number in purine and pyrimidine should be just plain 1, 2, 3, 4, 5 something like that.



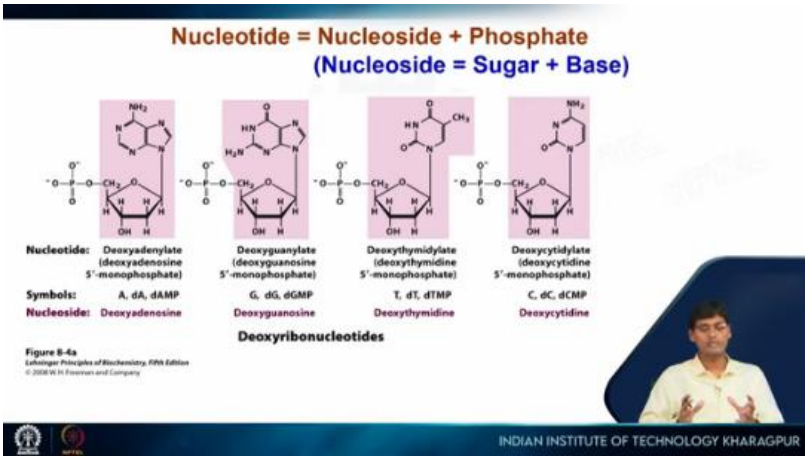
Let us see in the next step what is there. If you see pyrimidine and purine, they are attached to the sugar by a glycosyl bond. So, this is because the carbohydrate is attached to these nitrogenous bases. Some glycosidic group is attached to that. That is why this is called glycosyl bond and the phosphate is attached to the sugar by ester bond.

Now about the nitrogenous bases, which are the major component apart from sugar, this is also major component present in nucleic acid. So in purine there are two flavors of nucleotides present in nucleic acid. So the one flavor is purine. They are much bigger like adenine and guanine.



And pyrimidine, so we have cytosine and thymine. So, this adenine, guanine, cytosine and thymine, all these four nitrogenous bases present in DNA. So, pyrimidines are smaller compared to the purine. But in the case of RNA, we have uracil instead of thymine. So, in case of DNA, we have thymine, but in the case of RNA instead of thymine, we have uracil here.

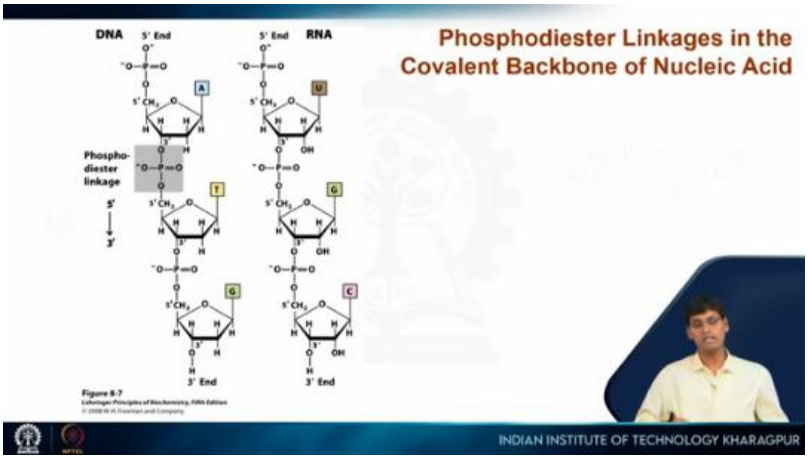
But they are more or less their structure is similar, only this CH3 is missing in case of uracil, that is all. Now, if we try to see about the nucleotide, how it looks, i.e. full chemical structure. So here you can see all these four structures. What we are showing here is deoxyribonucleotides.



So, those are the nucleotides, but these nucleotides are present in DNA. Why? Because here we do not have the OH group. So, that is why this is deoxyribonucleotides, which are the major building block of DNA. Similarly, if you see in this case, this is also nucleotides, but

this is ribonucleotides. So, here you have an OH group. So, this is ribonucleotides, which is present in RNA and as I already mentioned, again, I am clarifying that this portion, if I just include this portion here, just sugar and the nitrogenous base, so that should be our nucleoside and whenever we are including phosphate so together I would say this is now nucleotide building block of nucleic acid.

Now, we should go into the polymeric form of nucleic acid. So, we are just discussing the monomer nucleotides. Now, if we see DNA or RNA, their basic architecture is the same as how they are getting polymerized. So, if you see here in this case, this is one phosphate group here. This is sugar and this is the nitrogenous base we are just writing A or adenine just you know for simplicity otherwise again and again we have to write very big structure.



So, all these three together is one nucleotide, but now if you see this is one nucleotide I am just mentioning this is adenine, it is attached to sugar as well as phosphate group here. Now, it is next attached to another nucleotide that is thymine. So how are those two nucleotides joined together? So that is the question now. So this is the covalent linkage between two nucleotides.

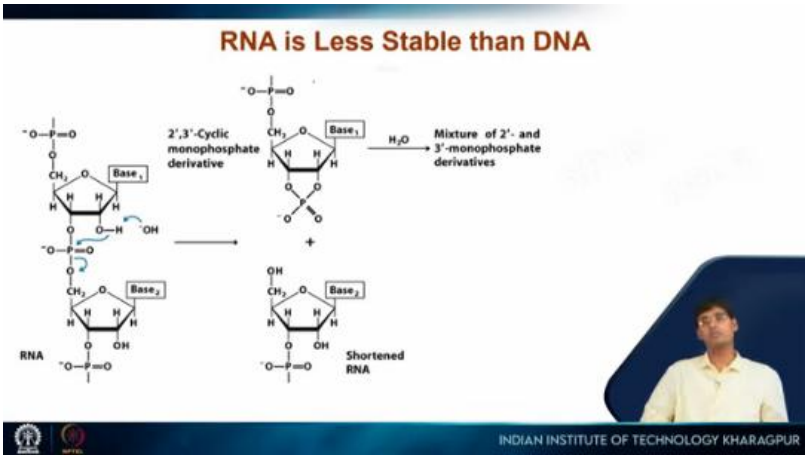
The reason is DNA is a polymer of nucleotides. So if you see, in between these two nucleotides, we have one covalent linkage here that is phosphodiester bond. So, if you see this is one ester bond and another ester bond and it is with the phosphate group. So, that is why the term is called phosphodiester, two ester bonds present here.

So, nucleotides are attached by phosphodiester bonds to make the polymer. In addition to that, we should clarify one important thing here that whenever we are writing, any nucleic acid sequence, so for example, I'm writing here ATGCAGC, something like that, just nucleotide sequence present in some gene or DNA. Now this is not a complete sequence. Always we refer by some direction, I would say this is convention on left hand side, we are writing 5 prime and this is 3 prime, 5 prime to 3 prime direction this is the sequence 5 prime ATGCAGC so then what is 5 prime and 3 prime? If you see in this DNA sequence here, at the 5 prime end, that means in the top end, the 5 prime carbon here, 5 prime carbon of this pentose sugar is attached to the phosphate group.

So DNA can extend only in this direction from 5 prime to 3 prime direction as you can see here. So at this, at the bottom end if you see the 3 prime OH is free here. So again, another nucleotide can be attached here by some reaction, which we will discuss later. So again, here another phosphodiester bond can form and another nucleotide can be attached.

So the 5 prime end, it is capped with the phosphate group, which is attached to the 5 prime carbon of the ribose sugar or in this case deoxyribose sugar and the 3 prime end is free with 3 prime OH group that's all. Whatever I discuss about this DNA this is true here in case of RNA also. Everything is the same only at 2 prime position, you will see OH is present, the hydroxyl group is present here and another difference is in the case of RNA, we have uracil instead of thymine. That's all. Otherwise, the basic skeleton is the same in case of DNA or RNA.

So now I just told the basic difference, right? In the case of DNA and RNA, although I told the basic skeleton is the same, the polymerization strategy is same, but at the two prime position of the sugar there is a major difference. The OH group present in RNA and H group is present in case of DNA and another one is the uracil present in case of RNA and thymine is present in DNA. But this thymine uracil will not create a lot of differences between DNA and RNA.



The major difference is the presence of the OH group at 2 prime position of the sugar. So, that is why RNA is less stable than DNA. So I always told that DNA is our genetic material. So, DNA is much more stable also compared to RNA. If we see RNA here, then this is a segment of RNA we are showing here.

So, we are just showing two nucleotides here, ribonucleotides particularly. So, now we have this OH group present here and in our body, in our cellular environment, the pH is a little bit basic. So pH is around 7.2 to 7.4. So in that condition, we have excess OH^- ion in the surrounding.

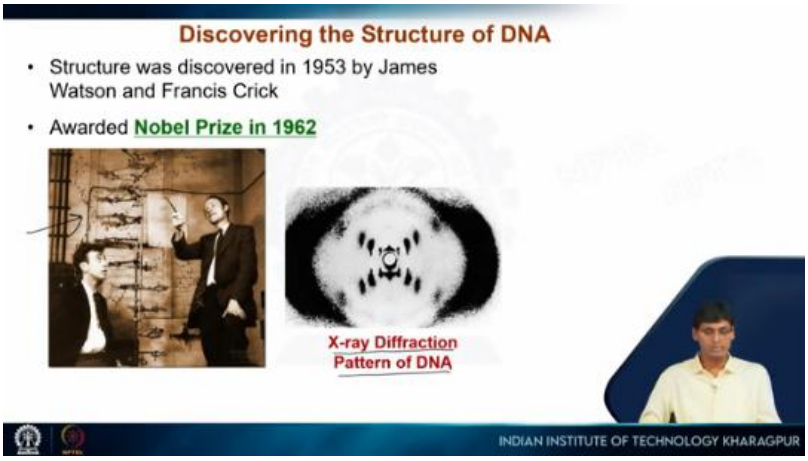
And from there, it will start a reaction and that will finally hydrolyse this phosphodiester bond. So as a result of that, it is very easy that the RNA will be broken down into small segments and that is the reason why RNA is little less stable compared to DNA and biologically it is very important.

Now about the discovery of the structure of DNA, this is a very important thing although at the beginning, I started with this, the discovery of nucleic acid. We thought that DNA was discovered by Watson and Crick. But that is not true. At the beginning, I told, almost 70 years before the discovery of DNA structure, Friedrich, who discovered this nucleic acid at the beginning. So as a result of that, what Watson and Crick did, they discovered the three dimensional structure of DNA.

Mostly from there we will know different physical properties of DNA. So, as you can see here, in 1953, James Watson and Francis Crick, published this paper in Nature. So, describing the structure of DNA and here you can see their model of DNA, just a hand-built model how it looks like and because of their discovery they got Nobel Prize in 1962 which is very important in terms of our biomedical science because it helps us to understand many aspects regarding genetics genes and biochemistry, molecular biology many things. Although Watson and Crick got the Nobel Prize because of this, they proposed this structure and it was absolutely cool.

But the thing is the X-ray diffraction data, which was actually obtained by Maurice Wilkins and Rosalind Franklin and based on that data, they understood what is the pattern in DNA, that some helical pattern is present, some basic repetitions are there. So finally, they came up with this model. I should mention here that during the same time, there are many other scientists, who were trying to decode what DNA looks like, including Linus Pauling.

But James Watson and Francis Crick, they came up with this model which proved to be true also. To get more details about the discovery, this particular book, the double helix written by James Watson, which is a fantastic book and now I would like to focus on some key features of DNA. We know that DNA is double standard, DNA is double helix all those things but it was really difficult to discover from that X-ray diffraction so Watson and Crick proposed that because of some pre-existing data.



So, one important thing is Chargaff's rule. So, Chargaff was a big chemist. He contributed a lot in this DNA understanding on DNA structure also in that way. So, what is Chargaff's rule? He found that in nucleic acid, particularly in the case of DNA, in case of nucleic acid, we have nitrogenous bases, purine and pyrimidine.

What Chargaff found that for a particular species, for a particular cell, if you analyze DNA, he observed that the ratio of purine to pyrimidine is 1 and the ratio of pyrimidine to pyrimidine is 1. So this is a very important thing and as a result of that I mentioned that pyrimidines are bigger structures, bigger nitrogenous bases and pyrimidine are the smaller ones.

Sometimes I forget also so the very easy thing to remember is the short name. Purine is a short name, but the structure is big. On the other hand, pyrimidine is the bigger name, but the structure is short. So they found that the purine and pyrimidine ratio is always one.

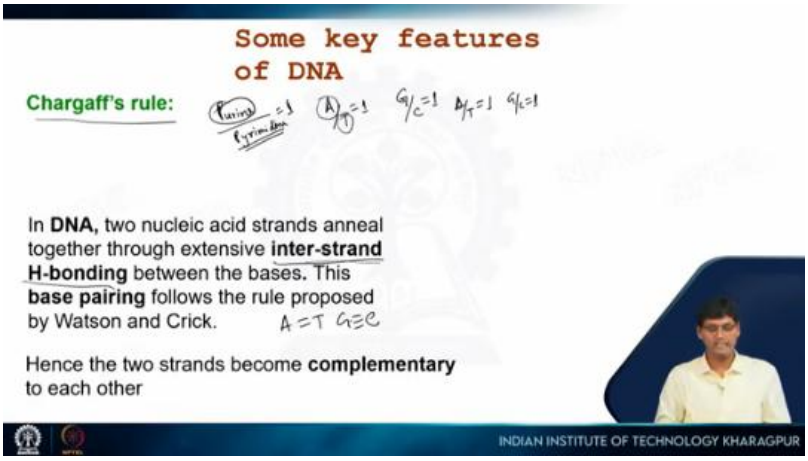
So that means there is some association between purine and pyrimidine. They stay together. Not only that, if we analyze even in more details, what was known that in DNA, $\frac{A}{T}$=1. So, A is purine and T is pyrimidine.

Similarly, $\frac{G}{C} = 1$. So, G is guanosine and C is cytosine. So, as a result of that, I can say that $\frac{A}{T} = 1$ and $\frac{G}{C} = 1$ in case of DNA. So, this actually helps Watson and Crick also to understand the DNA structure quickly.

So that means some specific base pairing, something is present there that is very important. I would say covalent structure of DNA, it was known long before. Covalent structure means that the nucleotides are attached by phosphodiester bonds and the sugar is attached to nucleotide bases. So those are the covalent structure of DNA. But this non-covalent structure is very important.

That was not known during that time. Watson and Crick came up with the right model. So in DNA two nucleic acid strands they are annealed or I would say they are somehow interacting together through inter strand hydrogen bonding, this is the non-covalent structure and this base pairing follows the rule proposed by Watson and Crick that A base pair with T, so here A and T they make two hydrogen bonds in between them and G and C they make three hydrogen bond in between them. So this is called Watson and Crick besepairing.

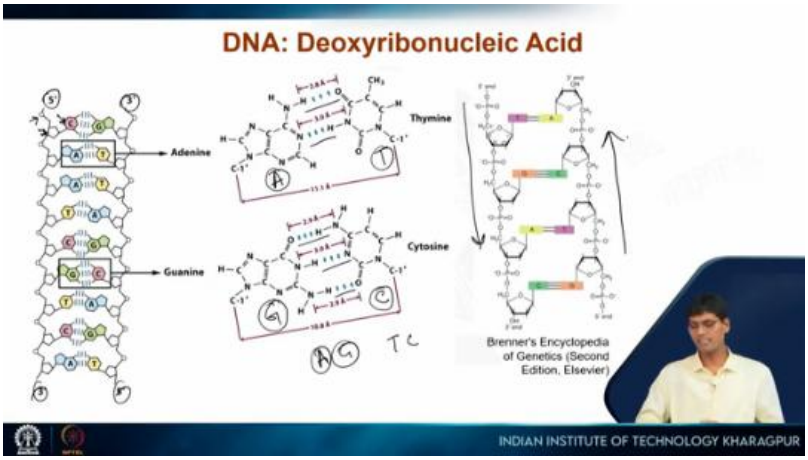
And this is evident from even Chargaff's rule also that the ratio of purine to pyrimidine is equal to one and hence the two strands become complementary to each other so as a result of that what I would like to say for example, if I say that this is the sequence of DNA ATCGCGTA then automatically I can derive the sequence of the other strand. So the other strand will be TAGCGCAT. Why? Because A is always base pair with T, C is always base pair with G. So they are complementary to each other. But one important thing is their direction will be different.



I am going to discuss it soon. So this will be 3 prime to 5 prime. So DNA, those two strands, are running in opposite directions. Now, this is the structure of double-stranded DNA as you can see here that 5 prime to 3 prime and 3 prime to 5 prime.

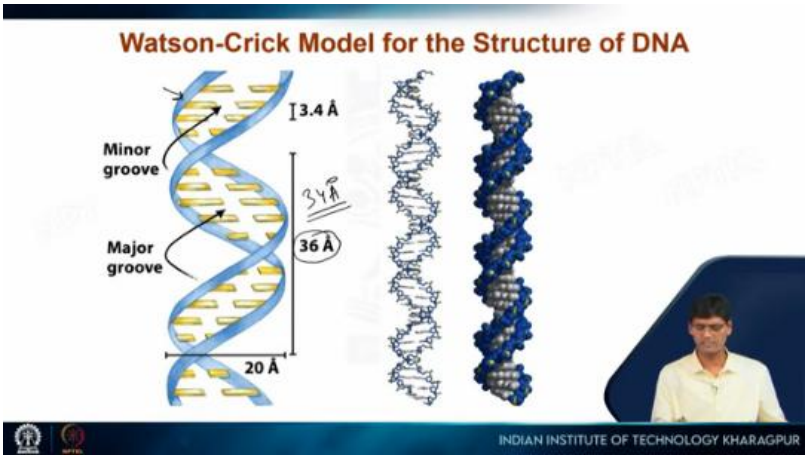
So, they are anti-parallel and now if you see here, this is the phosphate group, this is sugar and this is the nitrogenous bases. So that means that this phosphate sugar is the backbone, sugar phosphate backbone and nitrogenous bases are attached with the sugar group.

And now these bases are forming some hydrogen bond, specific hydrogen bonding. So, as I already told that adenine and thymine form hydrogen bonding. So, they have two hydrogen bonds here and in the case of G and C we have three hydrogen bonds as you can see in this figure. Now this is also very important so DNA is double-stranded and helical structure so as you can see A and G they are much bigger. So when DNA, when I will explain their structure, you will see that two strands are going in their dimension, like two strands are parallel.



Now, as a result of that, the width of the DNA should be fixed so one big adenine and one thymine, one big and one small. Similarly, here guanine is big and the cytosine is small. so as a result of that A and G together for example then the width will be more, both are much bigger and similarly T and C both will be much smaller so also there will be problem in hydrogen bonding and now in terms of antiparallel if you see in this figure, you can see this is going in this direction 5 prime to 3 prime. If you see in this way, then you will be saying that this ATGC, they are properly placed and they can form the hydrogen bonding and during replication I will explain that what is the chemical basis of this anti-parallel nature of this DNA. You know that DNA is a double-stranded structure and here, the helix, which was proposed by Watson and Crick, you can see here is a double helix. So, DNA is a double-stranded double helix. What kind of helix?

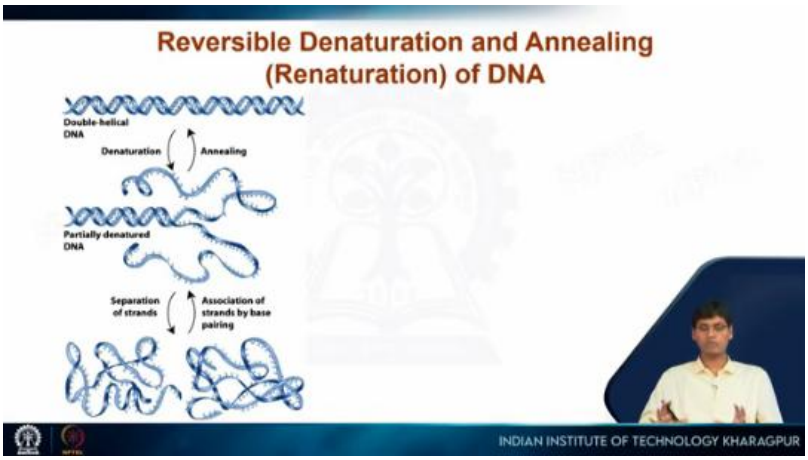
Right-handed helix means it is very simple. For example, if we take our right hand, like this and that thumb this is in upward direction and this is the imaginary axis. Now if we take these two fingers so these are the two strands of DNA and now if I just put these two fingers like this they are rotating, this will be the right-handed helix right so DNA is the right-handed helix of these two strands. So as you can see here, whatever we are showing here in blue color this is nothing but sugar phosphate backbone and in between them those nitrogenous bases they are forming this hydrogen bonding. So this is very important and also from their structure, they found lot of parameters for example in one turn what Watson and Crick found that in one turn we have that the distance is 34 angstrom, which Watson and Crick observed but later on a little bit different measurement, which is a more correct that it is around 36 angstrom and in between in complete one turn. Now we say that 10.5 residues are present and those are some specific details. Without knowing those, it will not be a problem, like the major, the structural property, the chemical property are very important things to understand about DNA and DNA structure, how those nucleotides are present and those are the instructions of life, how they are organized.



Now if we see this is a ball-and-stick and space filled model so this DNA this is whatever common DNA, this is the B form of DNA this is called B DNA. In the textbook you might get A DNA and Z DNA. So, they have some additional features that are a little bit different in terms of the width of the DNA and how many nucleotide pairs present per turn, little difference are there, but those are not important to understand the fundamentals of nucleic acids.

So, here this is the last concept. I am going to discuss here about the reversible denaturation and annealing of DNA. This is very important which is related to whatever I discussed right now.

So, the thing is DNA is double helical structure as you can see double standard double helix. Now, in a tube for example, here this is a tube and you have a DNA sample here. So, now if you increase the temperature of this sample, you are just heating this tube so something will happen to DNA as you can see here the DNA double strand will be opened up, they will be coming into single standard form. Why? Because those double standard structures are because of the hydrogen bonding. They are non covalent interactions. So whenever we are increasing the temperature of the solution, the hydrogen bonding will loosen and they will be destroyed. So as a result of that, finally, you will get the single stranded DNA. So, this is called denaturation of DNA. So, here you can see partially denatured DNA that means the temperature is not that high yet, but if you increase the temperature then two strands will be completely separated and this is two strands are separated completely here.



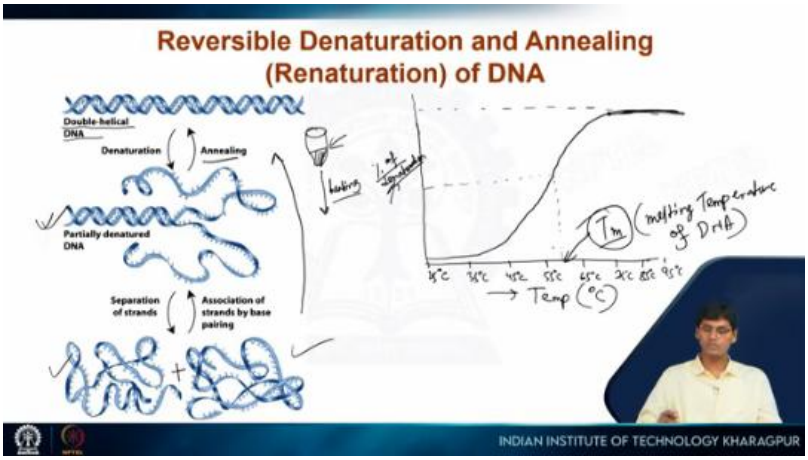
Now if you again start cooling down. So again you are decreasing the temperature. So as a result of that, those two strands again will come to each other and they will form proper Watson-Crick base pairing. So that is called annealing.

It will form a double-stranded and double helical structure. This is very important, like properties of DNA, which makes us to design many different types of experiments, which I am going to discuss sometime later. On this basis, if you see here, if I plot this, here in X-axis, this is temperature.

So, this is I am starting from 25^{o}C, it is 35^{o}C, 45^{o}C, 55^{o}C, 65^{o}C, 75^{o}C, 85^{o}C, 95^{o}C and this is the percentage of denaturation. Now I am taking this double standard DNA sample in this tube and then if I increase the temperature, what will happen? What will be the stage of denaturation? You will see something like this curve. This is because at the beginning at very low temperature like 25^{o}C, 35^{o}C, 45^{o}C, nothing will happen. So, DNA will not be denatured because at that temperature hydrogen bond will not be broken, but at higher temperature hydrogen bond will be completely broken.

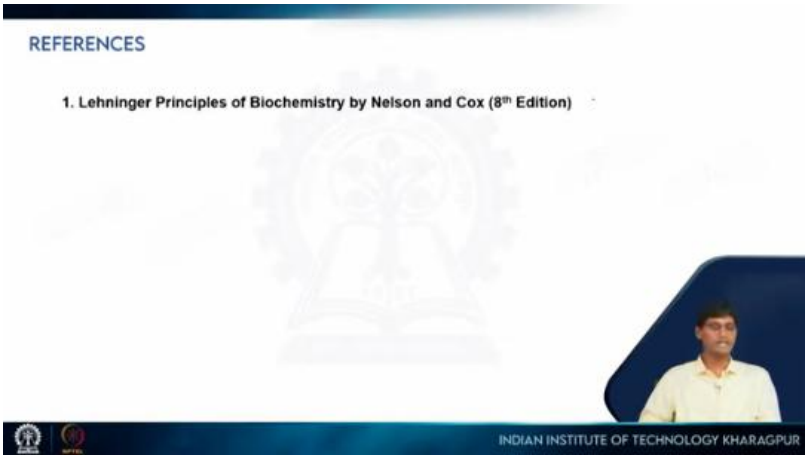
So, here for example, all hydrogen bonds are broken and now DNA, they are in their single standard form. That is why you can see it is not going up anymore so this is the maximum denaturation happened. This is called the denaturation curve. So, here the maximum denaturation and this is the 50 percent denaturation here.

So, whenever the temperature where the DNA is denatured 50 percent that temperature is called Tm or melting temperature of DNA melting temperature of DNA. So from this melting temperature we can comment on DNA, particularly their base composition. This is the last slide where I am showing two data sets, for example we have two tubes and we did the same experiment whatever I explained and from one tube, for example from this tube number A we get this data whichever plotted in blue colour and this one tube B we get the data in this brown colour so what is the difference in terms of base composition nucleotide composition in between these two DNA samples.



So, from this curve we can see that sample B requires more heat to get denatured. So, as you can see the Tm is higher in case of sample B. This is because I would say in case of sample B, we have more GC content. G and C, their proportion is much more compared to A. Because as I already mentioned that G and C, they have three hydrogen bonding.

That means you need more heat. You need more energy to destabilize that hydrogen bonding. As a result of that, in sample B, we have more GC content compared to A. So, that is all. So from this experiment we can talk about their nucleotides component and for further reading you can go through Leninger's principle of biochemistry and this is an excellent textbook and that is all.



Thank you very much.