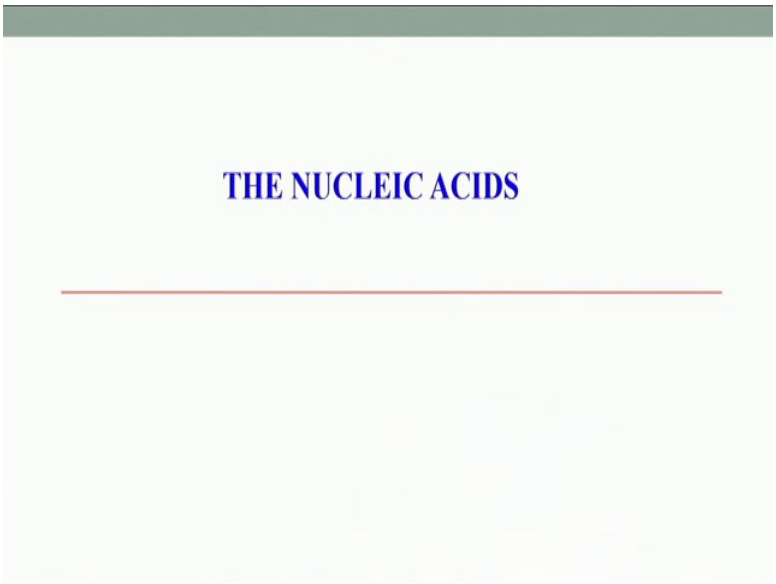
Organic Chemistry In Biology And Drug Development
Prof. Amit Basak
Department of Chemistry
Indian Institute of Technology, Kharagpur

Lecture - 22
Nucleic Acid

Welcome back to this course on Organic Chemistry In Biology and Drug Design.

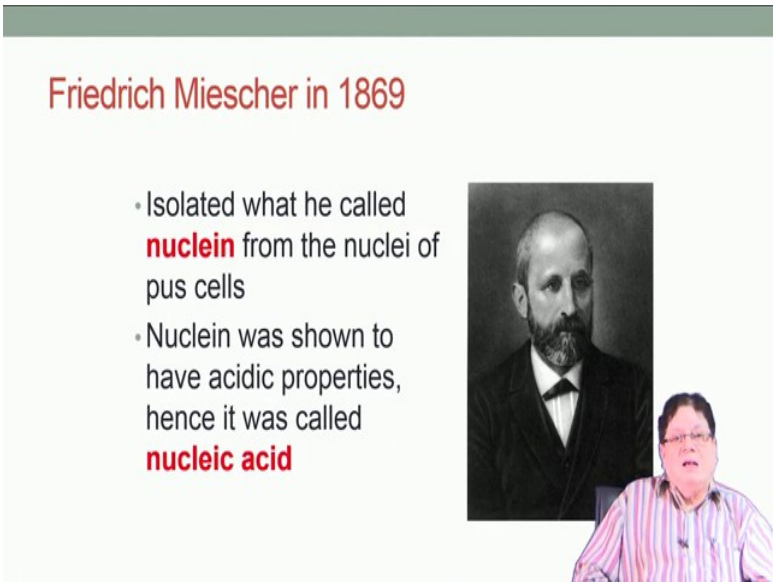
(Refer Slide Time: 00:31)



So far, we have covered various aspects of an important biomolecule like the proteins; we discussed their functions, and we have concentrated mainly on their catalytic activity(enzymes). We covered the structure of proteins, various hierarchical structures, then sequence determination, protein separation, enzyme kinetics and their inhibition profile. Then we went to one aspect of synthetic biology, and we covered the catalytic antibodies or abzymes; we discussed that how to design them so that they can be used to catalyze different organic reactions. We have just given few examples.

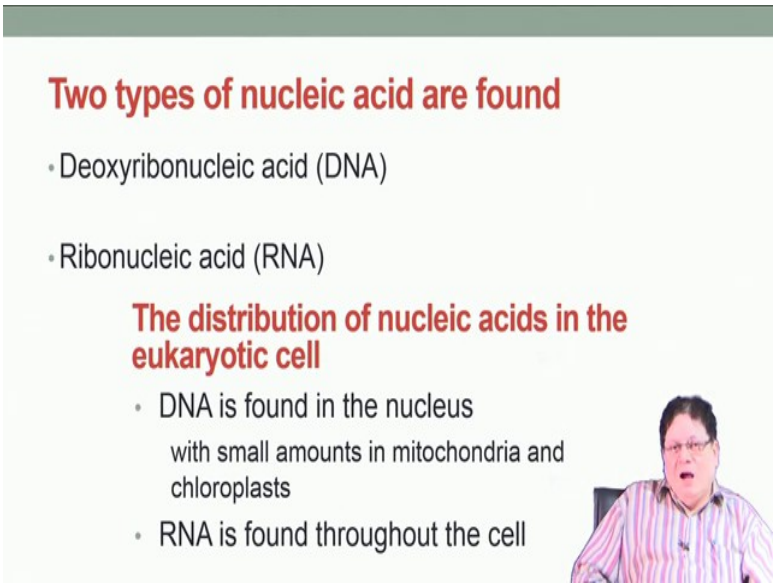
Now, today we will switch the gear and change to another very important topic, that is the Nucleic Acids. The nucleic acids are constituted of of sugar, phosphate, and a nucleobase. This is the monomeric unit which then joins together one by one and finally, give the nucleic acids.

(Refer Slide Time: 02:13)



Let us now briefly discuss the history related to the nucleic acids. In 1869, Friedrich Miescher, isolated something that he called "nuclein" from the nuclei of pus cells. Pus cells are the cells where we have infections; pus cells are excreted. So, from these pus cells, he isolated a product which we he called a nuclein. And later on nuclein was shown to have acidic properties and hence it was called nucleic acid. Thus the term changed to nucleic acid when it was found that it has got acidic groups. It is basically an acid, with acidic properties.

(Refer Slide Time: 03:07)

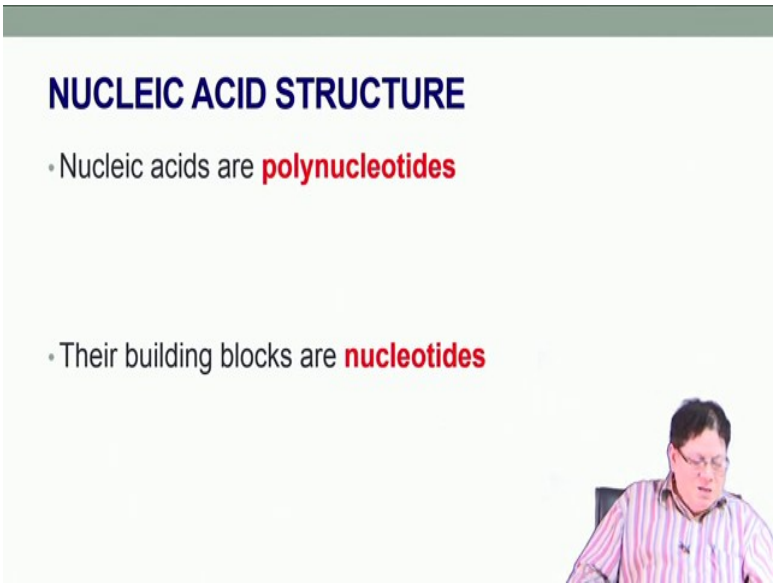


Two types of nucleic acids are known to us; one is deoxyribonucleic acid that is the DNA; that is usually the compound which dictates or which stores the information of making the required molecules to perform certain functions. To be more specific, deoxyribonucleic acid stores the information that via the ribonucleic acids gets transmitted; this information is utilized to make the required proteins that are required to maintain the living system.

You know that the cells are broadly classified as eukaryotic cells (that is present in the higher organisms) and the prokaryotic cells (that is present in bacteria). So, DNA is mostly found in eukaryotic cell in the nucleus; there is a separate zone (nucleus) which has got its own cover and within that the DNA is found. Small amounts of DNA are also found in mitochondria and chloroplasts; but mainly in the nucleus for the eukaryotic cells.

RNA on the other hand is found throughout the cell. So, throughout the cytosol, RNA can be found. So, the is one of the differences between the DNA and RNA. DNA is very localized whereas, RNA is not localized. In case of your prokaryotic cells, you do not have any particular nucleus. So, the cell has no inner compartment; thus it is not compartmentalized. So, DNA, RNA and protein are all together in the cytosol. Such scenario arises in case of the lower level of organisms.

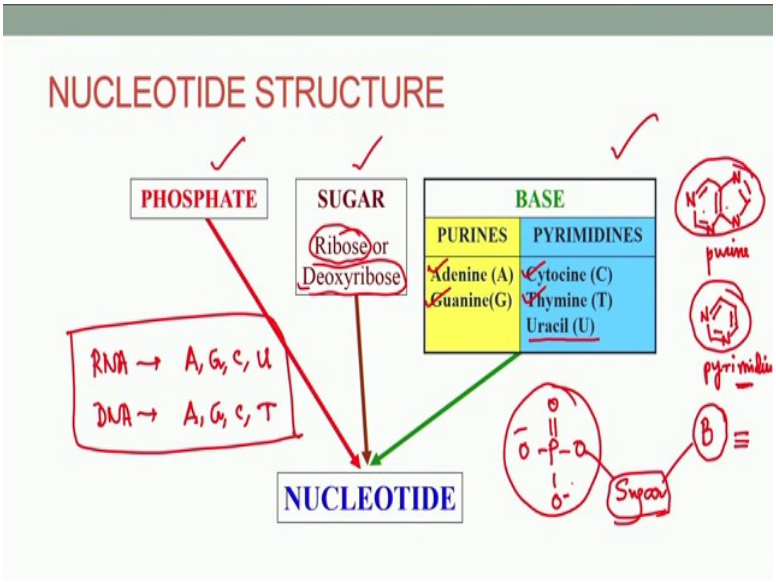
(Refer Slide Time: 05:32)



Now, the nucleic acids are of two types. Nucleic acids are poly-nucleotides. What are the monomeric units of this nucleic acids (poly-nucleotides)? Their monomer is the nucleotide

and the nucleotides are joined together. The nucleotides are joined one after another to form the poly-nucleotides, and these are called the nucleic acids.

(Refer Slide Time: 06:08)



What is nucleotide? The nucleotide contains a sugar, a base and a phosphate. So, you have a base here and you have a sugar (a particular type of sugar) and a phosphate linkage, together they make a nucleotide. Now, the question is what are these bases in the nucleotides? What is the sugar in the nucleotide? The sugars that are present are ribose or deoxy-ribose. Ribose is an aldopentose (five carbon containing monosaccharide with aldehyde group in the acyclic form); in deoxyribose; one hydroxy group is removed from some position. So, the sugar is either ribose or deoxyribose and then the bases are basically of 4 types. These bases are broadly classified as purine bases or pyrimidine bases depending on the type of hetero cycle. This purine and pyrimidine are the names of the heterocyclic system that are present in the bases. What is purine? This is called pyrimidine and when you add a an imidazole, then it becomes purine. So, this is your purine. 1,3 imidazole fused pyrimidine is purine; and the other one is only pyrimidine. So, this is the skeleton this is called pyrimidine.

So, these are the nucleobases, two of the bases we will have this heterocyclic ring and the other two bases have this heterocyclic ring. Now, you will notice there are 4 out of 5 bases which are present in the nucleic acids; and that is because there is a difference in one of the bases in RNA as compared to the DNA; that means, in deoxynucleotide, the purine bases are adenine and guanine. And other pyrimidine bases are cytosine and thymine. They are written

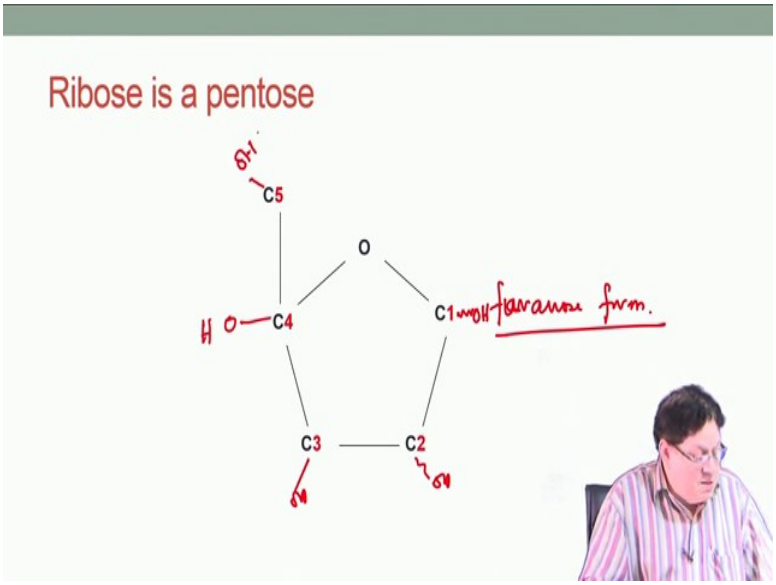
as A, G, C, T; whereas, in RNA, instead of thymine you have uracil. So, RNA has the bases A, G, C, T; whereas DNA has bases A, G, C, T.

Now, I will show the structure of the bases. Interestingly, there are only 4 bases in a particular type of nucleic acid. The arrangement of these 4 bases, their sequence actually dictates that how the information is stored in the DNA; in other words, how the information is stored, that is dependent on the sequence of these bases. It is similar to the concept of binary numbers in case of computers..

The binary digits are 0 and 1. You can arrange this 0 and 1 in different possible ways and that gives different types of signals. So, it is the same thing in case of these 4 bases; you arrange all these bases in different ways and that gives different types of information stored on the that DNA. So, nature did not use many bases, it used only these 4 bases to store the information.

Now, what about the sugar? The sugar is basically same; ribose in case of RNA and deoxyribose in case of DNA. So, this information that is stored that depends on the sequence of the bases because the sugar is constant. The phosphate part is also same. So, what changes is the base. This sequence of bases when they are connected in this poly-nucleotide decides the information that is going to be stored.

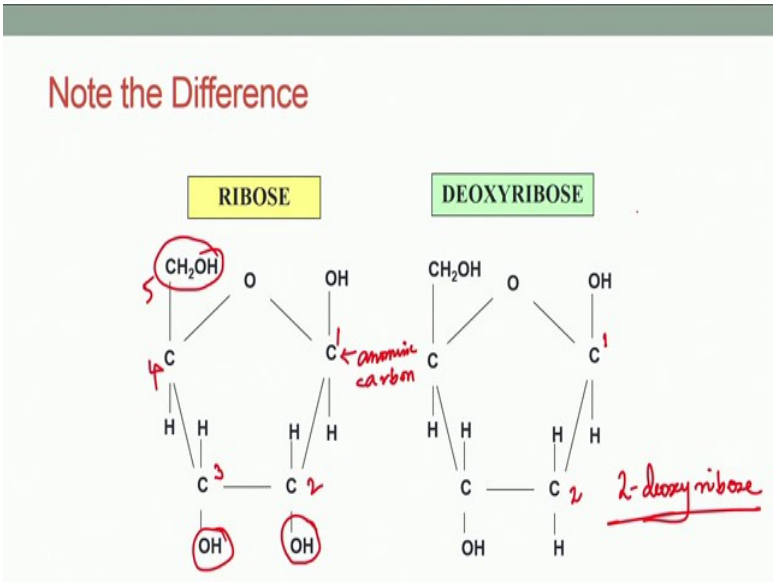
(Refer Slide Time: 11:57)



Now, let us see the structures of these bases; and the structure of the sugar. The sugar as I said is a ribose, it is specifically an aldopentose. So, that will have 5 carbons and it is present in the furanose form. We have to remember that there are two forms that are possible: furanose and pyranose form. The cyclic form has a hemiacetal linkage with an anomeric OH.

There are 5 OH groups attached to the furanose form as shown. In a particular configuration (with respect to the chiral centres), that becomes ribose. So, that is present, in nucleic acids.

(Refer Slide Time: 13:00)



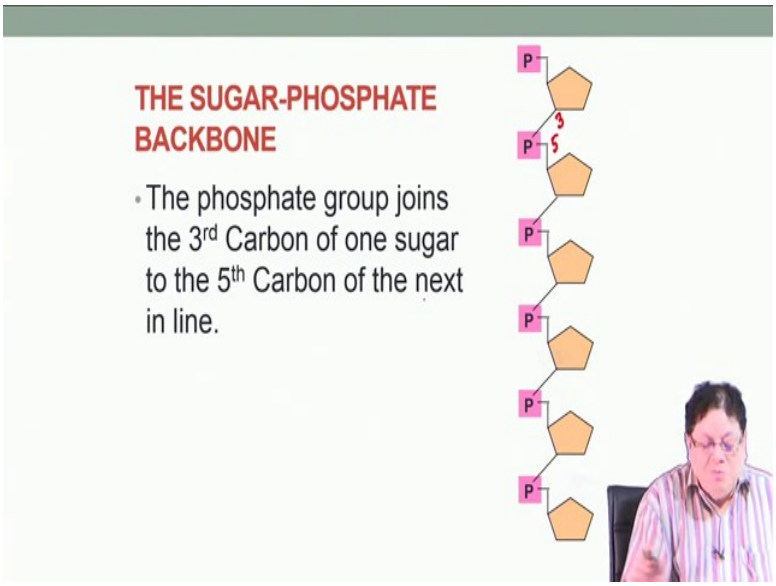
So, now, if we put the hydroxy in the proper perspective; that means, with proper stereochemistry, you get ribose where this is the anomeric carbon. Through the anomeric carbon, you make the glycosides. If you replace the OH and place a carbon, that is called a C-glycoside; if you put a nitrogen here that will called be a N-glycoside, O-glycoside. So, all these can be possible, all linkages occur through this anomeric carbon.

On the other hand, we should notice that this OH is α, α means it is below the plane of the ring and this is also α and this CH_2OH is β. So, that makes it ribose. If you change the stereochemistry, that is no longer ribose. So, nature uses this sugar to make the backbone of the nucleic acids. And deoxyribose is the one which will lack one hydroxy group and it is found that this 2-hydroxy group; if you name the ribose separately then this will be your 1, that will be your 2, that would be your 3, 4 and 5. So, it is the 2-hydroxy that is gone. So, this

is 2-deoxyribose because there are many deoxyriboses, but the ribose that is used to make the backbone of the DNA is 2-deoxyribose. So, that is all about ribose.

But regarding this numbering system, one thing I should mention here; the base has lot of carbons and nitrogens. So, the question is after you attach the base with sugar, whether this same numbering system will be retained or it will change? There is a slight change because base takes the numbering as 1, 2, 3, 4 and the sugar takes a numbering 1', 2', 3', , 4' and 5'.

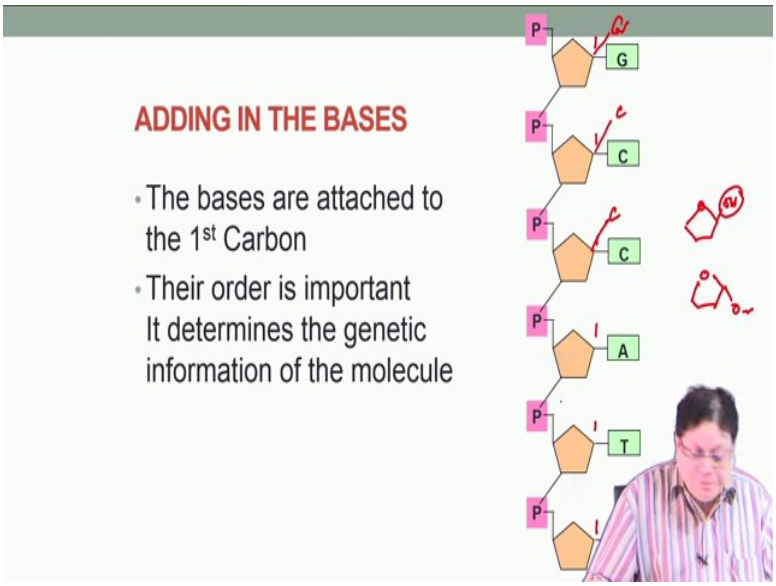
(Refer Slide Time: 15:18)



Let us now discuss the sugar-phosphate backbone. We know that there is a sugar moiety and that sugar moiety has C5 OH which will ultimately become 5' when the base is attached. The phosphate linkage connects these sugars one after another. You see this is the connection when it makes the nucleic acid backbone. And if you inspect it carefully then you will see that the phosphate group joins the third carbon (C-3') of one ribose sugar with the fifth carbon (C-5') of the next sugar.

So, when the connection is made, when these sugar units are joined by a phosphate. So, you see the 3' hydroxy of one sugar unit is joining with the 5' hydroxy via a phosphate linkage. We will also check this phosphate linkage. We have just said in a block like P, but there has to be oxygen on the phosphorous so, we will do that. So, the 3rd carbon of one sugar is connected to the 5th carbon of the next in line.

(Refer Slide Time: 17:03)

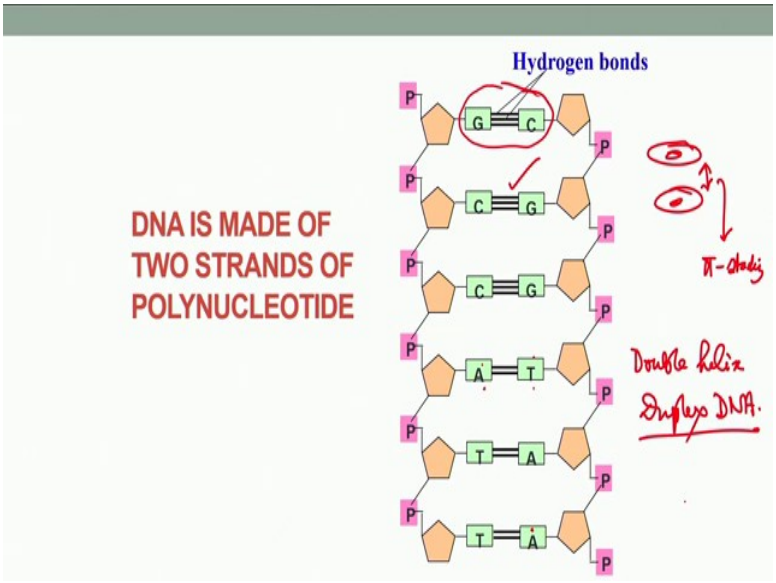


Then you add to the base to make it a complete nucleic acid. So, now, the bases are added. Where the base is added? The bases are added at the anomeric carbon; that is the C-1 carbon.

Another point to note is that whenever a sugar is there, the anomeric OH can take a β orientation that is the β glycoside; it can take an α position also. So, that is the α glycoside. But in this case, remember these are all connected like this G, this is C, this is C. So, they are all βglycoside, but these are N-glycosides because this is not connected to the carbon atom, this is connected to the nitrogen. I told you glycosides can be classified according to the atom which is connected to the anomeric carbon. So, if it is carbon, that is carbon C-glycoside, if it is oxygen that is O-glycoside and if it is nitrogen that is N-glycoside.

Now, so this is the arrangement of the nucleic acids when they make the polymer.

(Refer Slide Time: 18:48)



Let us now discuss the seminal works done by Watson and Crick. When the DNA was isolated, people tried to know the actual 3 dimensional structure of the DNA. There were a lot of efforts that went on simultaneously. It was one group at London, and then there were other groups at Cambridge University where Watson and Crick were working; and then there were other groups at California, where Linus Pauling was trying to make the structure of DNA. So, there was a fierce competition between all these groups.

Now, to know the 3 dimensional structure of this molecule, we needed a X-ray diffraction diagram. That means, you had to isolate pure DNA. Actually, DNA is not isolated as crystals, but it is like a fiber and taking the X-ray data of the fiber was not easy, but it was solved ultimately. And it was Watson and Crick who announced in a landmark paper in 1953 that DNA has a helical structure, there are two helix helical strands which are joined together by weak forces; at that time, they also said that the guanine (one of the purine bases) is hydrogen bonded to a pyrimidine base; specifically G forms hydrogen bond with C and A forms hydrogen bond with T. These lines are basically hydrogen bonds.

So, what happens that when one strand has all these sequences G, C, A, T, T, then the opposite strand that is called the complementary strand, that will have bases which can form hydrogen bond with the base that is present in the in the original primary strand. So, now, the sequence of the complementary strand you can write C, G, T, A, T; because there is T

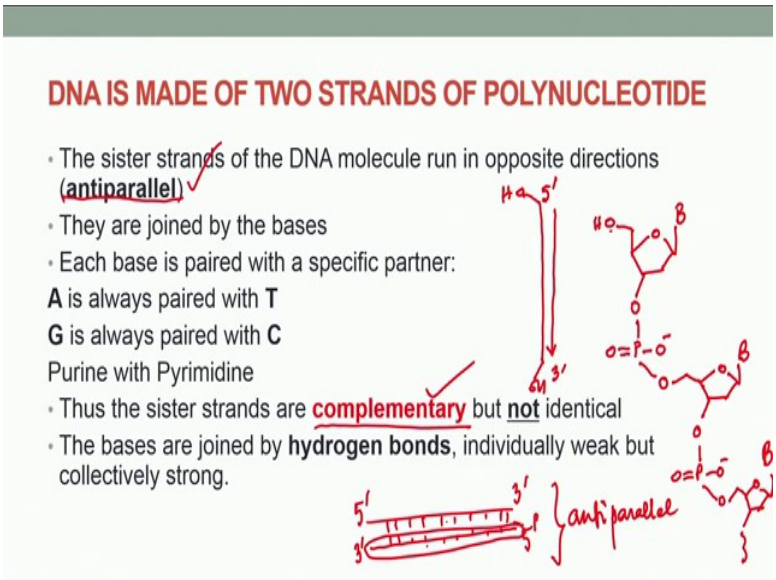
recognizes A and forms two hydrogen bonds; similarly G recognizes C and horms 3 hydrogen bons. So, in that way, the two strands are attached.

What is the force by which these two strands of DNA are attached to form a double helix? It is not only the hydrogen bonds that attach these two strands, but also π-stacking. These rings are primarily aromatic in nature. And when the aromatic rings lie one after another in a parallel mode, then there is some stabilizing interaction which is called π-stacking interaction.

Since all the bases are parallel to each other, so they have π-stacking interactions. So, this π-stacking also stabilizes the DNA. DNA double helix is sometimes known as duplex DNA.

So, we know that the duplex DNA is formed because of these hydrogen bonds and π-stacking; and nature has devised this mechanism which actually ensures that these bases are all inside because these are the storehouse of our information. These phosphates are actually exposed outside, but the bases are stacked together inwards. That is basically protecting the DNA from outside agents to damage these structures.

(Refer Slide Time: 23:27)



We have already said A pairs with T, G always pairs with C, so that means, a purine pairs with a pyrimidine. There are two important points here. The most important point here is the complementarity of nucleobases. The second important point is the anti-parallel nature. Suppose you join a sugar unit. Now, let us clarify what is this phosphate and how they are linked together. This is oxygen and then you have a phosphorus, that is connected to double

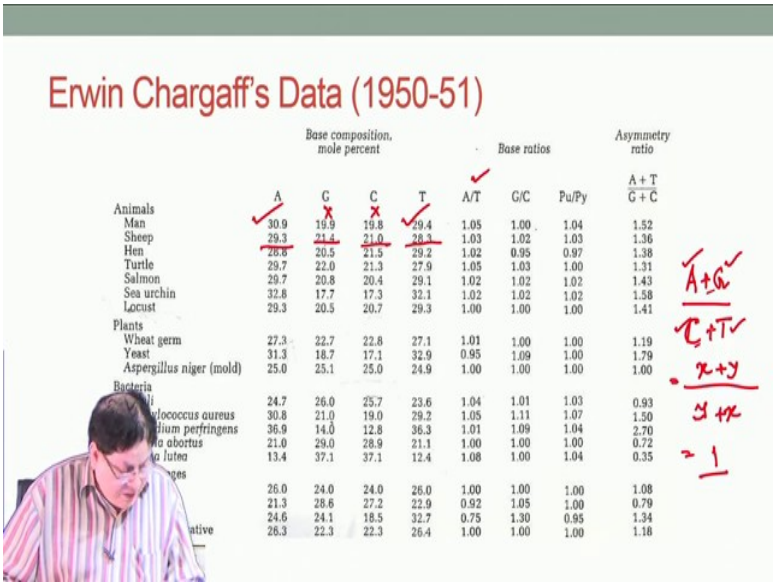
bond O and O minus and then this is the oxygen from the 5’ position and this is the base and then again another one comes like this. So, you have the phosphate, and O, this is your sugar unit. So, it goes like that, that is another base.

Now, there are few things to notice. We are writing the DNA strand from top to bottom, the strand is extending like this. So, at the top, you have the DNA strand, on one side (top) your 5’ OH is free. And on the other side (bottom), 3’OH will be free because the length has to terminate at some point of time.

The DNA has a defined length, so at one end you have a 5’ end and the other end is called the 3’ end. What is 5’? In a peptide, you have N terminus and C terminus. Those are free. In DNA, which hydroxy groups are free? On one end the 5’ hydroxy is free and at the other end, the 3’ hydroxy is free. So, we say that this runs from the 5’ to 3’ direction. So, whenever we write a piece of DNA, we have to mention. For protein it is usually understood that this is the N terminus that is the C terminus, but in DNA usually it is better that you mention, there is no rule that the left side is always 5’ and the right side is 3’; it could be the other way around.

I showed that the stabilizing forces responsible for this attachment of the two strands are all weak forces; all the Gs are paired up with the C via 3 hydrogen bonds. Although, each hydrogen bond is weak, but the sheer number of hydrogen bonds, the number of weak interactions is so enormous that DNA becomes very stable. But also remember that there is another stability factor which is called the π-stacking interaction between these underlying base pairs.

Now, so basically there are two strands whenever I talk about a double helix DNA; and there will be hydrogen bond in between the complementary nucleobases. What is anti-parallel? That is because one strand runs from 5' to 3' in one direction. And the other strand runs from 3' to 5'. (Refer Slide Time: 28:22)

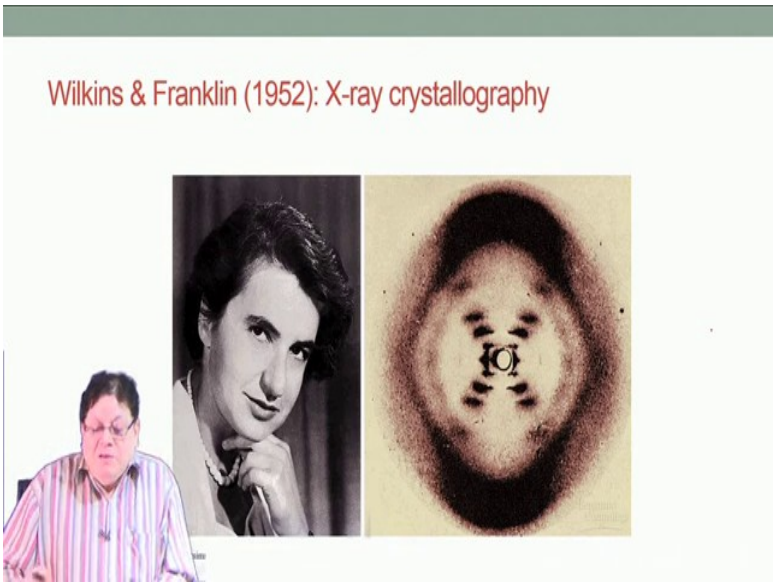


Before the structure of DNA was known, there was another scientist whose name is Chargaff. He took DNA samples from different sources. He was studying the percentage of adenine, percentage of guanine, percentage of cytosine and percentage of thymine. These are the 4 bases. He was measuring the percentages of these bases.

This turned out to be a very important experiment. From man, he isolated some DNA; the percentage of A is 30.9 and percentage of T is 29.4. So, within all experimental error, they are same. There is not much different. The percentage of G is equal to C and A is equal to T. So, you take another species like sheep; 29.3 is A, and T is 28.3; these are two (C and G) are almost same 21.4 and 21.0. So, it was after studying the DNA of various species, he came to the conclusion that A/T is nearly 1; and G/C will also be nearly 1; that means, A, percentage of A is equal to percentage of T, percentage of G is equal to percentage of C.

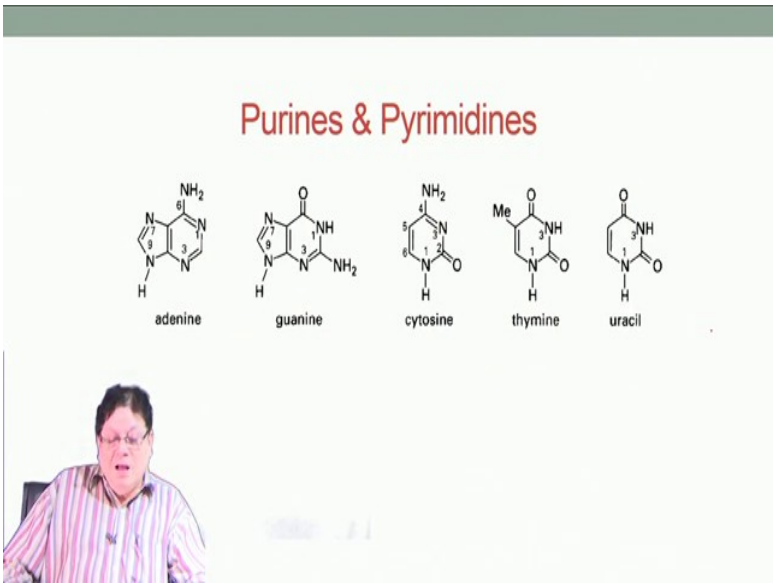
You can also find out what will be the ratio of A + G and C + T. Now, the percentage of A is equal to percentage of T. So, if that is x, so that will be also x. Percentage of G is suppose y, so percentage of C is also y. So, A + G divided by C + T is also equal to 1.

(Refer Slide Time: 30:59)



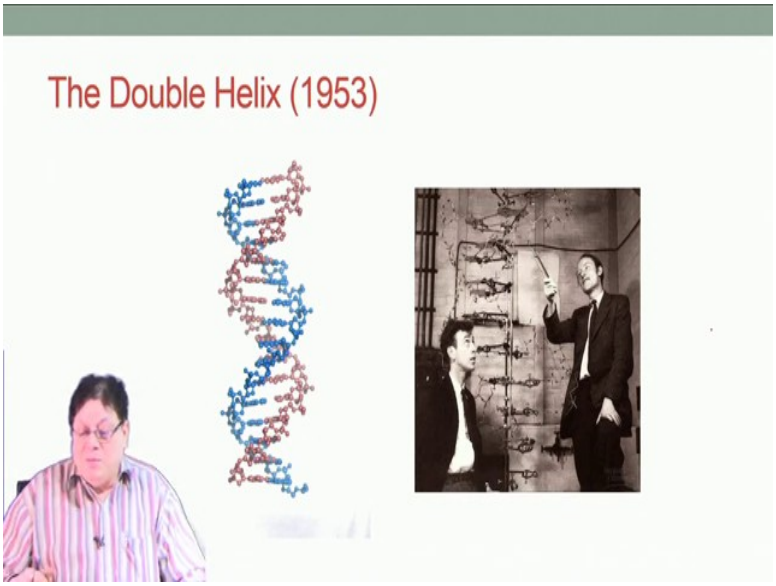
Now, this is the famous X-ray crystallographic picture that was taken by the London group. I was telling that there were 3 groups at that time which where heavily competing with each other. So, there was one lady scientist, hers name was Rosalind Franklin. Franklin, in 1952, took this X-ray picture of DNA. That was the best X-rays picture that one could get at that time. So, Franklin got this picture. Maurice Wilkins was also working along with her, he was a colleague of her. She showed this picture to Maurice Wilkins.

(Refer Slide Time: 32:02)



What is the truth behind it, nobody knows, but somehow Watson and Crick got hold of Franklin's X-ray photograph. Crick was very good in X-ray crystallography. So, after seeing all these, he interpreted that the DNA is helical. Only problem was they could not come out with an explanation that why the percentage of A is equal to percentage of T which was earlier proposed through the Chargaff's rule that percentage of A is equal to percentage of T and percentage of G is equal to percentage of C. Both the both of them worked together to solve the problem.

(Refer Slide Time: 32:48)



And there is a famous picture I will show you. Yes these are the two gentlemen. This is Francis Crick and this is James Watson. So, they virtually made the DNA with the model building blocks. DNA required numerous number of atoms, they connected those atoms and ultimately found that it has a kind of helical structure; it is like the twisted stairs that you see in the fire exit. So, they worked one whole night and finally, came out with a conclusion and got the structure of the DNA. They proposed that G is pairing with C and A pairing with T. And that solved the pending structural problem.Since DNA is a double helix, that will immediately tell you that the percentage of A will be equal to the percentage of T and percentage of G will be equal to the percentage of C. So, they announced this in the next morning on 28th February. This date is also a very important day as it is celebrated as the national science day. This is because C.V. Raman discovered his Raman scattering on this day. So, it is a very famous day. But they did not publish at that time, they actually

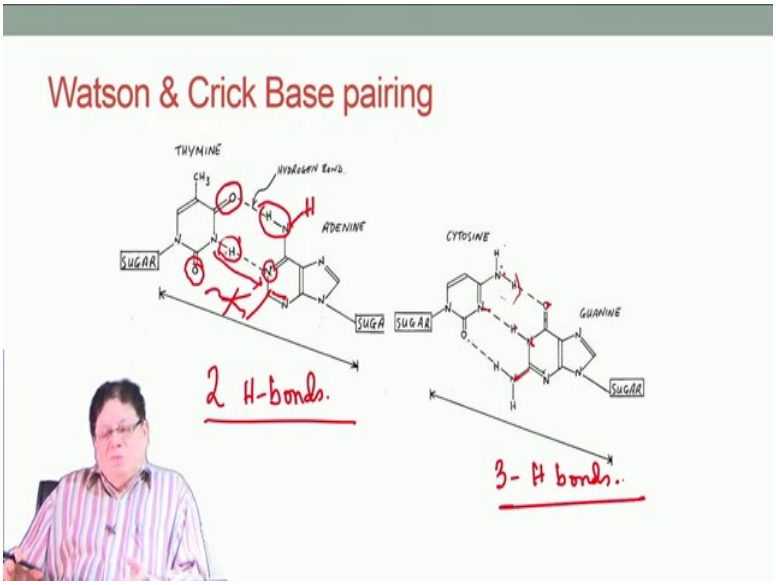
announced it on 28th of February that they solved the mystery of the life because they immediately realized that this discovery of DNA structure will lead to the subsequent happenings in a living system including how the living system functions, how the memory is stored, how from one generation message, is passed to another generation.

So, that was a landmark discovery in biology. And the whole molecular biology area started because we started seeing the biological system at the molecular level. So far, I did not show you the structure of these bases. There are 5 bases, uracil is present in RNA and instead of a uracil, thymine is present in DNA and these are the structures.

So, this is the purine nucleus and then you have an amine. Remember the numbering system, that is also very important. The numbering system starts from here: 1, 2, 3, 4, 5, 6, 7, 8, 9. So, this is 6-amino purine, that is called adenine. So, here it is 1, 2, 3, 4, 5, 6. So, 6 –oxo-2-amino purine is called guanine. Cytosine, thymine, and uracil are called the pyrimidine bases; that means, a type of benzene ring where the nitrogens are in a 1,3 relationship.

So, this is cytosine; again the number starts from here: 1, 2, 3, 4, 5, 6. Now, sometimes it is really difficult, but you should remember this numbering because all the time when some reaction takes place on the bases, we generally say like the C-3 has reacted or N-7 nitrogen is protonated. So, you have to remember this numbering. In pyrimidine, always this nitrogen is one and then it goes to the other nitrogen 1, 2, 3, 4, 5, 6. So, that is same in every pyrimidine (cytosine, thymine and uracil). So, only thing you have to know is that which nitrogen is taken as one.

(Refer Slide Time: 37:20)



In hydrogen bonds, you need a donor and an acceptor. Donor means some electro negative element with a hydrogen and an acceptor is another electro negative element.

Now, if you look at the structure of thymine, this is the structure of thymine. Thymine has a methyl, this is carbonyl and this is a pyrimidine ring. So, it has got a NH, it has got a carbonyl and the carbonyl there. In adenine, actually there is a double bond here, adenine there is this NH_2. So, now, this NH is a donor and thymine carbonyl is well placed to make a hydrogen bond with this NH.

And on the other hand, this is NH on the thymine which becomes a donor. So, this is acceptor, that is a donor which is well placed to make hydrogen bonding. Well placed refers to the distance criteria, the distance between the donor and the acceptor should be ideally less than 3 Å for strong hydrogen bonding. So, this nitrogen forms the hydrogen bond with the NH and this carbonyl forms the hydrogen bond with the NH_2 of adenine. But for the other carbonyl that is an acceptor, unfortunately there is no other group which can donate hydrogen to this carbonyl. So, this has only two hydrogen bonds.

Let us now consider the the G-C pair. If you look at adenine, adenine has one donor atom NH and an acceptor nitrogen, but here nothing is there. That is why no hydrogen bond is formed here. But in guanine you have this is acceptor, this is donor and this is donor. And if you look

at cytosine now, this is your donor; that means, your acceptor donor that combination you need is present. This is your acceptor.

So, again donor acceptor combination is there, but unlike adenine, in guanine you have another functionality here which is another NH_2 which acts a donor. As I said there is nothing here. So, you do not have any hydrogen bond here, but here you have a hydrogen bond because there is the NH_2. So, that is why there are 3 hydrogen bonds formed in the between G and C.

So, basically they are so complementary to each other that they fit electronically, geometrically and then form the hydrogen bonds and that is the basis of their formation of the double helix. Now, these 3 hydrogen bonds versus 2 hydrogen bonds that will be important. And in the next session I will tell the importance of this the G C and the A T and their percentages.

Thank you.