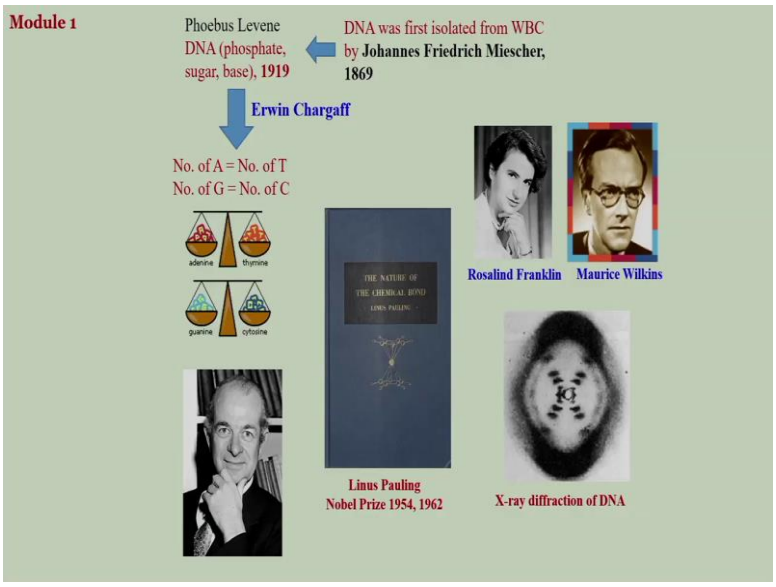
Essentials of Biomolecules:
Nucleic Acids, Peptides and Carbohydrates
Dr. Lal Mohan Kundu
Department of Chemistry
Indian Institute of Technology dash Guwahati

Lecture 02
DNA Double Helix Chemical Parameters

Hello everybody and welcome back to my second lecture of the of this course in on biomolecules. So, in the last lecture I have given a short introduction about the different biomolecules that are present in the living organisms and I also have tried to show the importance to study the chemistry and the biology behind those molecules. So, and after that we have started with the first biomolecule that I picked up as DNA, nucleic acids in deoxyribose nucleic acids or RNA also later on we will go on.

So, what we have seen is that the background story of it that with the help of the certain information's and that were available during that time.

(Refer Slide Time: 01:26)



That the DNA consists of the 3 major components and the one is phosphate otherwise is sugar and another is of course the nucleus bases. So, that is one piece of information. Second is of that number of equivalents of adenine is the same as that of thymine and the number of guanine is equivalent to the number of cytosine. And with the help of the experiment actual experiment and x-ray diffraction pattern of DNA that has been obtained by Rosalind Franklin and Maurice Wilkins.

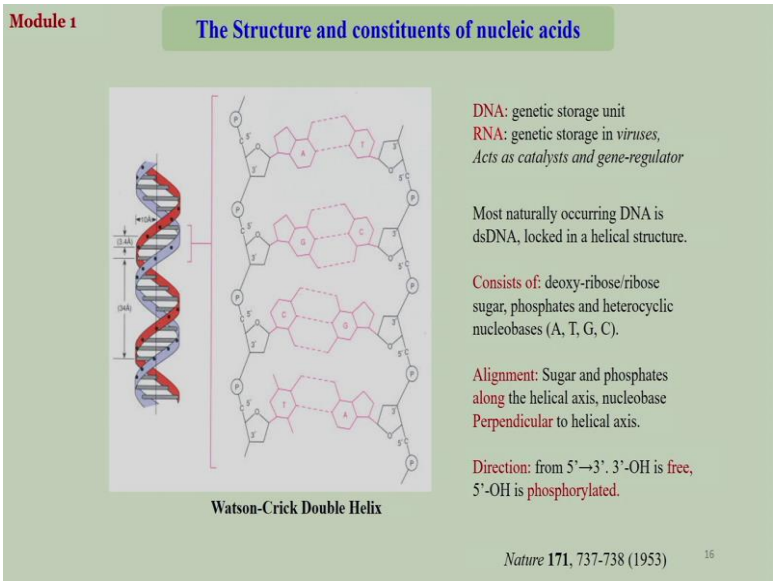
Watson and Crick finally came up with the structure of the DNA that is the double helical structures and B DNA. So, and that structure is obviously unique in nature unique in the sense that this is the first molecule or this is one molecule in one biomolecule that is not a single entity it is actually a combination of two molecules and they are oriented in such a way that they are and they have certain relations.

And in this case the relation is bonding and PI stacking interactions so that makes a single entity but two different set of monomer molecules. So, that is one uniqueness of DNA second is that all I have to already talked about that it is the only molecule in the living cells or even you can find it cannot find out a similar molecule that can self-replicate that can make its own copies. So, because of this discovery this has this has been a tremendous discovery during that time.

Because structure and functions of DNA was not been able to be explained until and unless the correct 3-dimensional structure of DNA was elucidated. So, 1962 Watson and Crick along with Maurice Wilkins they had received the Nobel Prize in the Physiology and medicine. Rosalind Franklin was not included in the list because she died beforehand at a very young age and that is a very that is a quite a sad story.

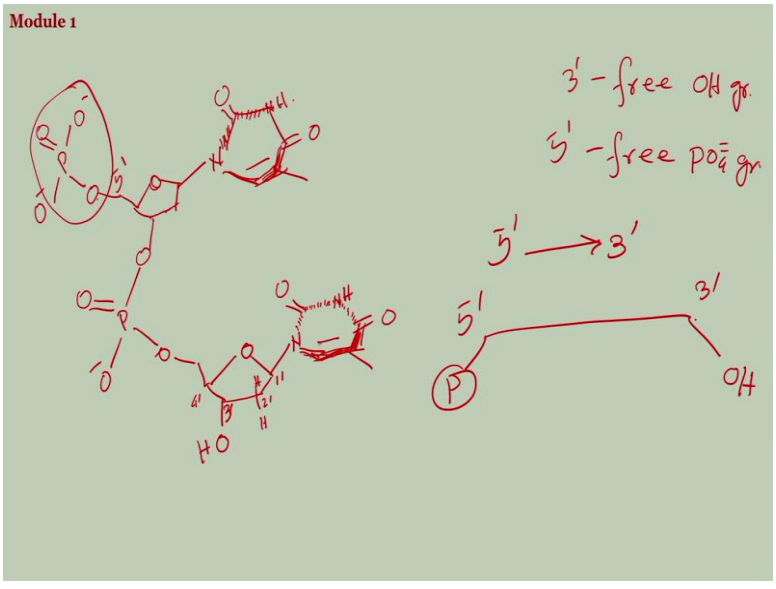
But on the happy end on the happy note Linus Pauling did not return empty-handed also. So, in the same year 1962 Linus Pauling received his second Nobel Prize in peace. He was one of the key persons who were canvassing who had strong voice against the use of nucleus weapons and for that in the same year as James Watson and Francis Crick and Maurice Wilkins received a Nobel Prize for DNA 1962. Linus Pauling received his second Nobel Prize for Peace. So, that brings me to the structure of the DNA.

(Refer Slide Time: 04:37)



So, this is the plane and paper model of DNA that shows the skeleton of the DNA and I think I have explained yesterday also that you have this has a suger unit and the nucleus base is here and then a phosphate unit 3 sugar nucleus and four wit they are oriented in such a way that it forms a double helix. And I have mentioned that the sugar and the phosphate makes the skeleton this is the backbone of the DNA and then the flat aromatic nucleus bases are perpendicular to the sugar or to the helix that makes a scene perpendicular position. So, that the one nucleus base and then the second nucleus base are on top of each other.

(Refer Slide Time: 05:28)



So, if I draw I will try to draw the structure this is the deoxyribose that means their OH here, here there is no OH is only proton and proton this is 3 prime end here you have the nucleus base that is the one prime end 2 prime this is 4 prime and then comes the 5 prime has CH2 and then OH, OH that is connected to a phosphate. So, there is another sugar here and they say OH

so the 3-prime end if I consider this is the end of this DNA 3 prime end always has a H group free hydroxyl group.

And the phosphate is this is not a complete phosphate this is O minus so it is basically all phosphate unit PO4 minus unit when it is deprotonated it will be double minus. So, this and then again a phosphate O minus double bond O O minus this is the terminal again so this is the 5 prime end of this. So, 3 prime end has a free hydroxyl group 5 prime end has a free phosphate group 3 prime free OH group 5 prime end has free phosphate grou.

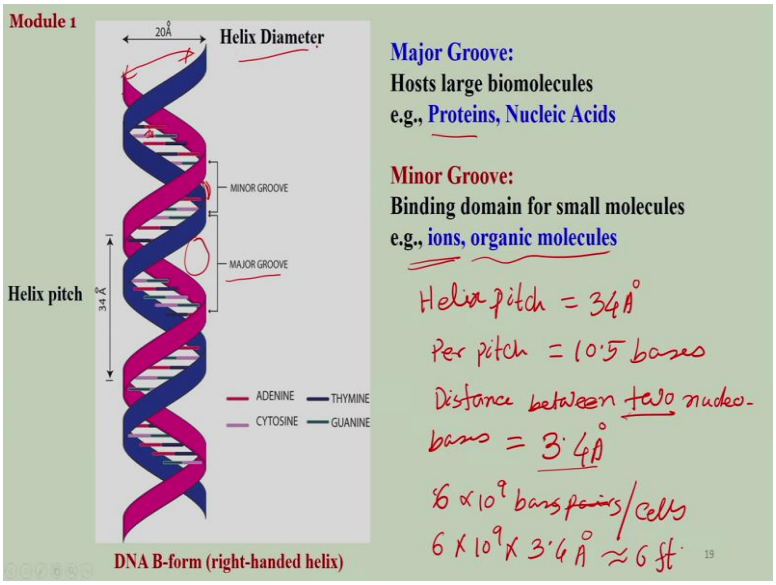
And now coming to the nucleus base if I consider I am just for the sake of simplicity I am considering thymine for both cases. How it would look like? So, if you have the N this will be perking to it like this with it is a 6 membered ring this part I am writing bold so that it means that it faces towards you and the other one is actually behind you or behind the board. So, this is the double-bond methyl NH so it looks like this.

Similarly the other one up the plane and this is the down the plane both are I am right in both methyl so the front face with the double bond methyl is fasces towards you the other side is back so that makes a flatter and the plane perpendicular to the axis. And whenever you write a DNA we usually write it in the direction of 5 prime to 3 prime. So, there is a style of writing the DNA this is usually the practiced style from 5 prime to 3 Prime.

And of course if you have the single-stranded DNA the 5 prime end has our phosphate I am writing P circle as the phosphate PO4 and then 3 prime N has a free hydroxyl group and these are very important actually later on we will see how the presence of the terminal groups actually makes a difference and that we will see during replication process and DNA replication process. So, that is how it comes here so DNA is the genetic storage unit.

Now when it comes to RNA that is the genetic storage in viruses and they are also acts as a catalyst. DNA of course mostly a natural DNA that version exists as the Watson Crick double helix structures we call it double stranded DNA dsDNA and they have helical confirmations which is of B form we call it a B form. Alignment of course as I have talked sugar and phosphates along the helical axies nucleus perpendicular to the helical axis.

(Refer Slide Time: 10:21)



So, this is how the 3-dimensional structure of the DNA would look like this is the DNA B form which is a right handed helix. So, here this way and there are few features that are associated with the structure of this DNA. One is of course that if you look one complete turn of the helix that is starting from here to there up to here these what this distance it is a complete turn of a hillocks this is known as helix peach and that is about it is written here it is written here it is about 34 angtrong and per helix pitch there are approximately 10 base pairs or 10 nucleus bases that are there.

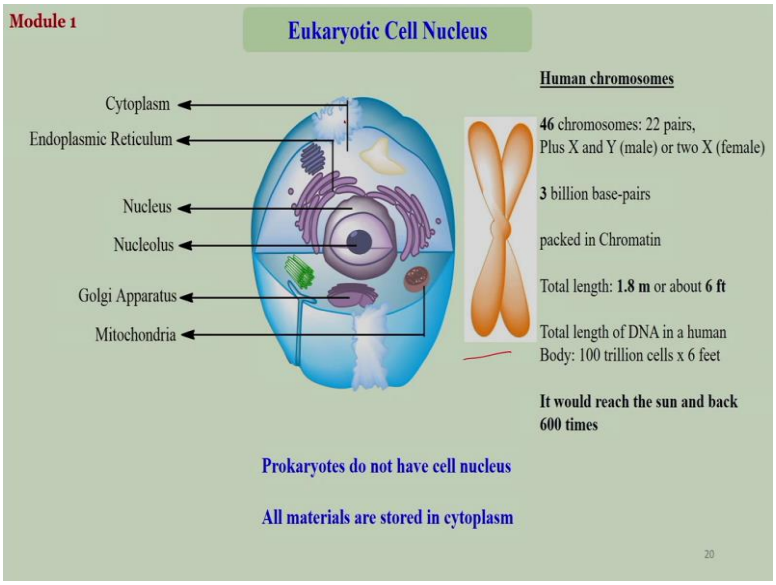
Or exact is actually 10.5 bases nucleus bases are there for helix pitch. If you look here one nucleus and the second nucleus base down on top of each other this distance between one nucleus base and the second nucleus base is; so distance between two nucleus bases to nucleus bases means the neighbouring nucleus bases is approximately 3.4 angstrom and another important thing is the distance between the two helical moieties the distance between two strands one strand is represented by this pink other one is by this purple one this distance this distance between the two DNA strands is known as the helix diameter.

This is approximately 20 angstrom so that is how they are closed in and because of this structure now there are two important pockets that are formed within the DNA this is the B form of the DNA that I am talking about which is the most natural one. One is called the major groove other one is called the minor groove. So, if you see it forms here it is a large pocket little bit larger pocket in the helix. So, this is known as the major group the relatively larger pocket.

So and these are very important as they are the positions of where the other molecules bind and do chemistry or do biological transformations. Major group is for the binding of the larger molecules for example proteins or other nucleic acids like RNA and so on. So, larger molecules would bind to the major group because it has a relatively larger pocket. And another type is the minor groove here.

Here this part this is a small pocket low space that is why this is called a minor group. So, the minor group is for binding with smaller molecules because the size is small and the space is small. So, minor group basically binds with ions or smaller organic molecules for various kinds of biological transformations. So, these are very important aspects of a DNA double helix.

(Refer Slide Time: 14:14)



So, now how is DNA present inside the cell. So, this is a drawing of a schematic presentation of a cellular nucleus the inside nucleus and other parts of the cell and this is for the eukaryotic cells and that means the higher organisms. Lower organisms are called prokaryotes and they do not have cell name class most of them. So, this is a nucleolus and this part is basically the nucleus where DNA stays.

All DNA the genetic information is stored inside the nucleus of the cells. And just outside the nucleus there is cytoplasm and endoplasmic reticulum, Golgi apparatus Mitochondria. So, inside the cellular nucleus DNA lives and most of the other biological molecules or biomolecules they mostly stay in cytoplasmic area. And DNA is all DNA is packed inside the nucleus as chromosome and they are packed in chromatin the space is called chromatin.

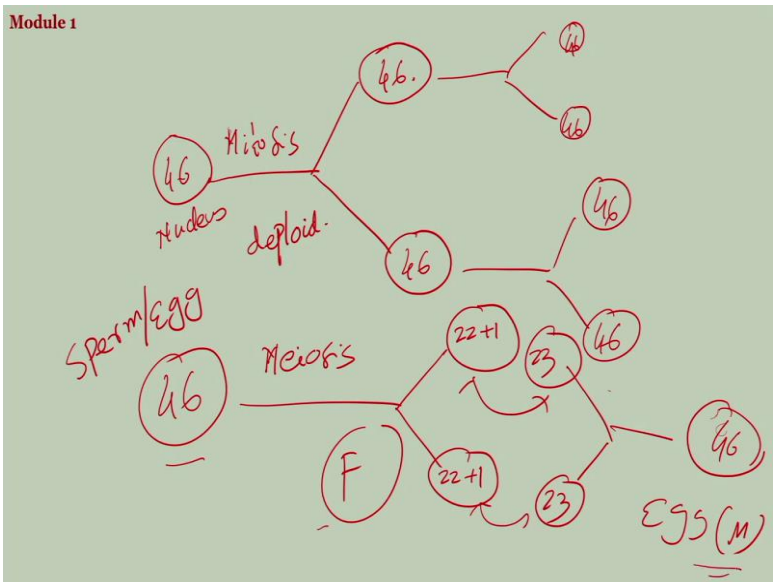
This is how a human chromosome looks like. So, now some statistics if you look so as I have said the distance between two nucleus bases is around 3.4 angstrom. Now if you look at the human DNA total amount of chromosome so in every cell every human cell we have around 3 billion base pairs. So, 3 into 10 to the power 9 base pairs per cell that makes since DNA is double helix that makes here total number of 6 into 10 to the power 9 bases per cell.

Now if you calculate the total distance between the nucleus bases here so that will come out as 6 into 10 to the power 9 multiplied by 3.4 angstrom if you calculate that is roughly equivalent to 1.8 meter or roughly equivalent to 6 feet. So, in our every cell the length of a DNA is around 6 feet long that is a larger area and all of it is stored or packed into the nucleus of the cell into chromatin. Now an adult human being has around 100 trillion cells basically.

So, now if you multiply that with 6 feet a total human body contains the total length of the DNA is equivalent to reaching the Sun and back from there 600 times that is a huge, huge area that our DNA covers within our body well. Our cells are small so around 6 feet of length of DNA is packed inside the cell. And the longest cell longest DNA it is composed of only 6 chromosomes that is Indian one type of Indian Red Deer particular type of Indian Red Deer that has only 3 chromosomes in it we have 46 chromosomes that year has only 6 3 chromosomes in it and around 3.3 billion nucleus bases.

Now I will talk about something how the DNA are there inside the cells and how the multiplication happens.

(Refer Slide Time: 17:57)



So, one interesting fact is that of course that many of you already know if you have that I am writing the cell nucleus and we have at around 46 chromosomes. Now when cell division happens this chromosome or this nucleus so this is cell nucleus 46 chromosomes are staying inside the nucleus. When cell division happens this cell is divided into two new cells and these are nucleus again and all 46 of it goes there all 46 of its goes here that is the DNA replication that is what Watson Crick has found out that had solved the puzzle how the single cell is divided into two cells both of them retaining the exact genetic information's.

So, this is how through a process called replications DNA can unwind which I will talk later. Now all 46 it can make two different copies the same 46 will be here the same 46 chromosome will be here and that is how our cells grows up. All of our cells I am talking about eukaryotic cells now higher organisms all our cells has around 46 chromosomes. And this type of cell division is known as mitosis and it is a diploid cell division. This is called diploid cells because it multiplies into the full 46 to 46 the process called mitosis cell division.

So, the question is now that if all our cells are equal all 46 then how all the human beings do not look like the same if the genetic information is transferred fully to the offsprings then why we all look different and that is true for all other higher organisms animals and all and so on they look very different we have different set of properties. So, the point is usually our cell divisions of course happens through a mitosis state and way.

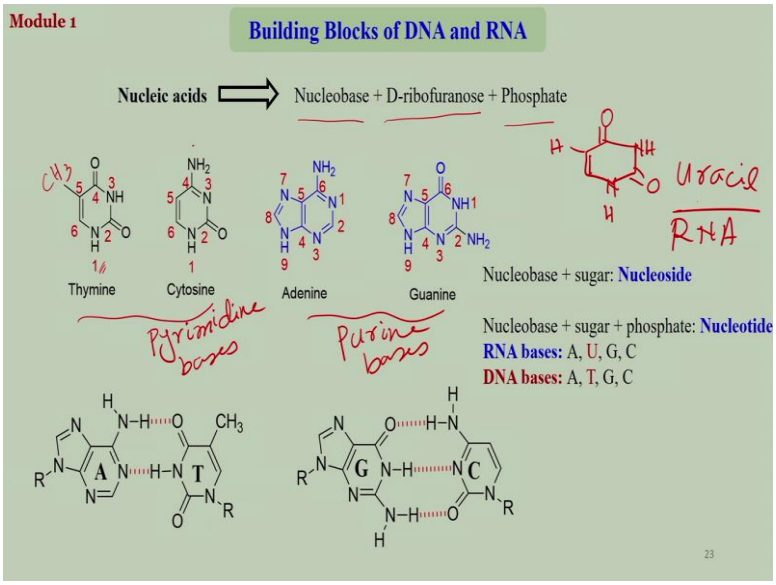
And from here for the cell divisions happens all mitosis all 46, 46 there is one exception that is start with the 46 and then it undergoes not mitosis but meiosis cell divisions where in the new cells it is not 46 it is 23 half of it will go haploid cell divisions 23 means 22 + 1 and here also 22 +s 1 haploid cell divisions half cell divisions and that happens in sperm cells or egg cells father's sperm cells or the mother's egg cells.

These are the only exceptions where cell division is not full is not 46 to 46 say cell division is half Floyd 46 to 23 and 23. Now so if this comes from a father so same thing happens for mother if this is the sperm cells this is the egg cells from mother and here also is 23 and 23 haploid cell divisions and then the cell's fertilized and would produce a mature cell with a single nucleus they will be fused together single nucleus. Now the new cells will have together 46 chromosomes in the single cells and that is how our completely new cell is developed.

Now which 23 will come out of from the mother from the father and which 23 will come from the mother is unpredictable almost unpredictable. So, the new cell that has been generated is completely new completely different than anything that exists. So, that is how the variation is created that is how we all look different that is how we all have different level of immunity different strength and all these things that is a very good protection also that higher organisms have from the environment.

The lower will organisms such as virus and bacteria most of them they have the same nucleic acid structures they have the same genome structures. So, that is bad for them in the sense that if you can kill one cell through a chemical agent you can kill the entire population which is not true for higher organisms because we have different immunity level different strengths.

(Refer Slide Time: 23:09)



Now the constituents of the nucleic acids on DNA nucleic acids contents as I have mentioned 3 different segments one is the nucleus bases the second one is D deoxyribose sugar and third one is it phosphate. There are enormous amount of chemistry involved in the whole structure of the DNA as we will move on we will see. Each of these segments has their particular role their confirmations play an a very important role and so on.

I will start with the nucleus bases first these are the four nucleus bases that are present in DNA thymine, cytosine, adenine and guanine. Thymine and cytosine these are present in DNA. RNA has a different one I will just mention briefly later. Thymine and cytosine if you look at their structures they have both of them has a 6 membered rings and both of them have this kind of

pattern nitrogen two nitrogen atoms present in 1-3 positions. These are called the pyrimidine rings.

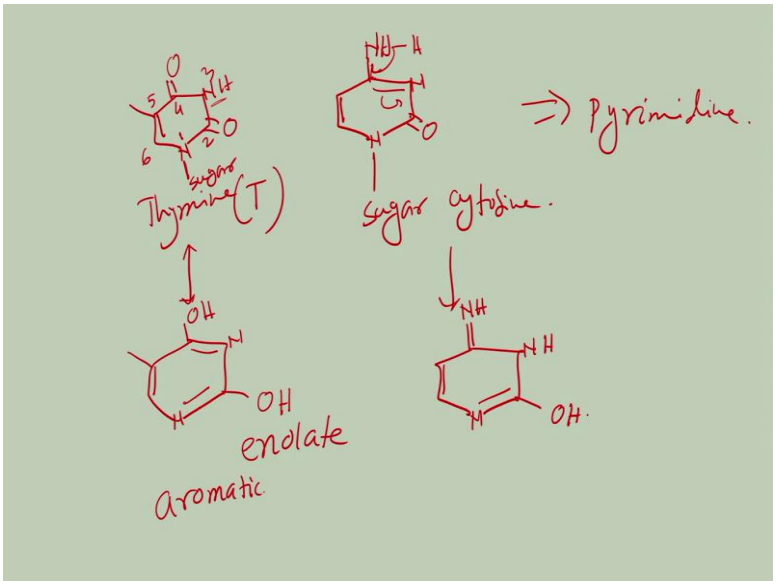
So, thymine and cytosine are of the nature of are known as pyrimidine basis because they have the basic skeleton is pyrimidine structure which has 6 membered ring which has two nitrogen atoms in one and three positions. And on the other hand adenine and guanine have one 6 membered ring fused with a 5 membered ring. Six-membered fused with a five-membered and they have of course this skeleton again 1-3., 1-3.

So, these two are known as purine they have the basic skeleton that help as purine kind of structure. So, these two are known as purine bases adenine and guanine are purines bases thymine and cytosine are pyrimidine bases. This is how the numbering goes if you look at the time in so this NH is 1 2 3 4 5 6 this NH CO NH CO this is the C carbon-carbon double bond and a methyl group.

So, the numbering has to do with of course your standard system that the nitrogen atoms the hetero atoms should have the numberings which is lowest in this case is 1 2 3. So, you have to start with either you can start here 1 or here 1. So, 1 2 3 and the second factor is of course the functional groups has to appear the closest. So, that is why this starts here so that the carbonyl will come at the second position so 1 2 3 4 5 6.

So thymine has basically a methyl group this is methyl CH3 group at the 5 position that is important. In RNA thymine does not exist what exists is called uracil has the same structure as thymine NH CO NH CO double bond and nothing here H hydrogen that is uracil and that is present in RNA. So, in RNA instead of thymine or uracil exists. So, this is what thymine looks like when it comes to cytosine here you have a amine group. And since there is amine group there is a double mode and in this is a difference.

(Refer Slide Time: 26:49)

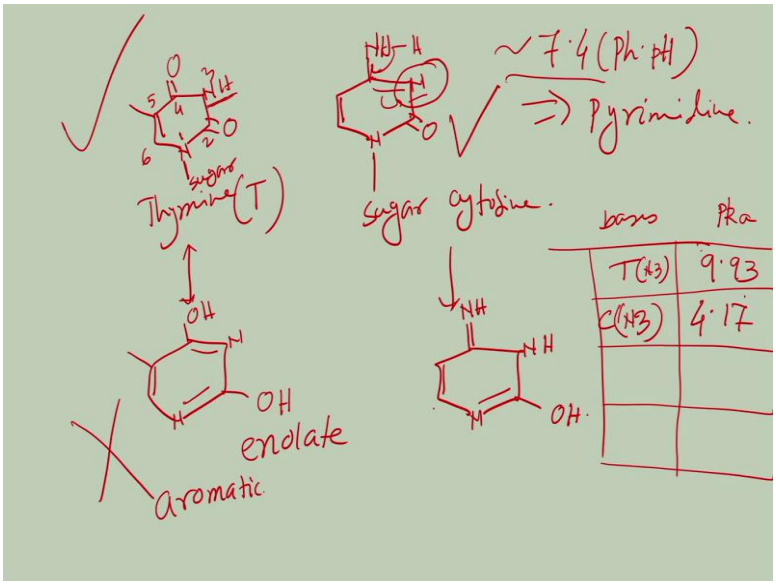


So, thymine and cytosine are pyramid I will draw the structure of the structure of both nitrogen NH double bond here and this is a metal that is thymine usually represented by T. Now when this is connected to the DNA then of course it is connected with the sugar and the sugar comes here in this case here N1 position numbering is 1 2 3 4 5 6 as you have seen. So, sugar is attached to the N1 nitrogen that means in nucleoside the NH is not present only nitrogen is present so you have 1 NH here cytosine N here is your sugar double bond O nitrogen here's an amine so therefore there is a double bond here.

And nothing just another double bond this is cytosine and together they are known as pyramid. Now this is one form these nucleo bases they are actually all aromatic in nature, how? If you look at their resonating structures in this form can exist in enolate form as well. Here you have a double bond O so you can have here N double want H methyl in the free form if you do not have sugar then you have the enolate here as well.

So this is the free nucleus base form this is the enolate structure with 3 double bonds in a 6 membered ring so that makes it aromatic. Similarly the free-form here nitrogen in there will be wage and this can have exist in the other form that is amine NH it is basically the nitrogen NH H that is how the in enolate forms is that. This is another form this is and when it is a re-aromatise it comes back then you can have the aromaticity back. So, these are the enolate forms of it and that makes it aromatic.

(Refer Slide Time: 29:51)



Similarly if you look at the thymine of the purine bases that is adenine and guanine adenine is 6 membered ring fused with a 5 membered ring. Here there is one nitrogen and another one here this is usually attached to it a sugar double bond here N 1-3 another in let us say amine here. So, that has to be a double bond and here nothing so that is another double bond that is adenine represented by A. So, if you look at the numbering again the same way starting with here it will be 1 2 3 4 5 6 7 8 9 number 9, N9 position is attached with the sugar.

In sugar writing s for sugar and then double would not, there is in here so it has to be an H is another in amine here that has to be double bond. So, this is guanine represented by G. Now again as similar to pyrimidine bases they can also exist in all its structures and they are also the aromatic compounds. Now the question is if I go back DNA has a specific structure and therefore the structure or the conformation of the sugar in which form they will be staying in DNA has to be fixed.

It cannot move back and forth to enolate as well as the locked form because in that case the structure of DNA would be flexible. So, question is out of these two which one is the correct structure that DNA exists. Now the nucleobases exists. So, to know that I will give you the experimental PK, value of the nucleobases. So, if you have a thymine T has with sugar in DNA it has a sugar so it has free one nitrogen that is accessible so that is your N3.

So, this is basis and here is the pk value the pk value of NC of thymine is around 9.93. Now that means that this NH this would be deprotonated if the pH of the solution is above 9.93 if the pH is above 9.93 then the NH would be deprotonated and it will exist as N. Below that pH it

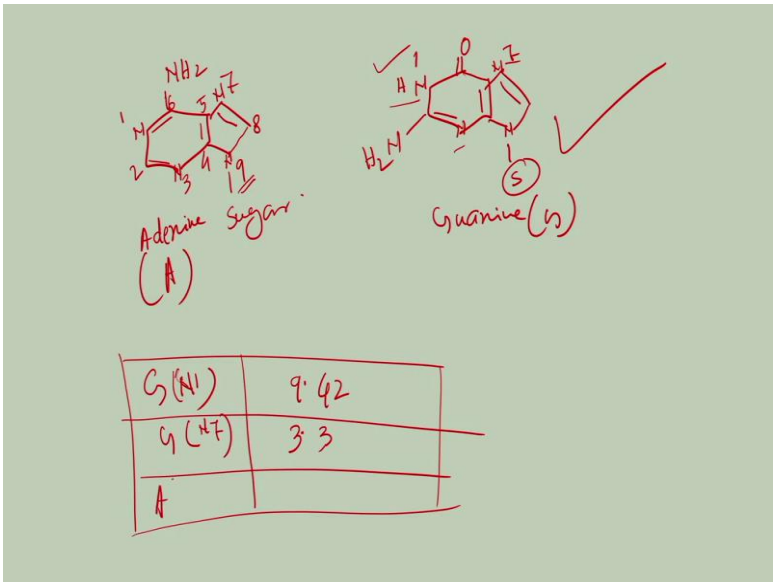
will stay as NH. So, now in our body we call it physiological medium, physiological pH is around 7 7.4 actually. So, that is the physiological pH is around 7.4.

So in our body system we have a pH that is most of the cells there are other cells also which has a low pH. So, 7.4 is that in general the pH of our body so in that solution how the thymine exists in which formula exists. It is obviously lower than 9.9 so NH would stay as this. Therefore this enolate form does not exist, this is the form which exists. And that has very important to value because that is very important for base pairing.

If you consider enolate form the base pairing would be completely different compared to the the original this form non-enrolled form. Similarly if you look at cytosine, cytosine also has N3 and that entry of cytosine has a pk value which is 4.17. Now again this means that above 4.17 it will stay as deprotonated form and physiological pH is 7.4 that means that is already above pH 4.217 so that means that in cytosine the nitrogen would exist as N not NH.

So, again this is the form that will exist in biological system in DNA not this form. So, thymine and cytosine both would exist in this form not the enolate forms.

(Refer Slide Time: 35:03)

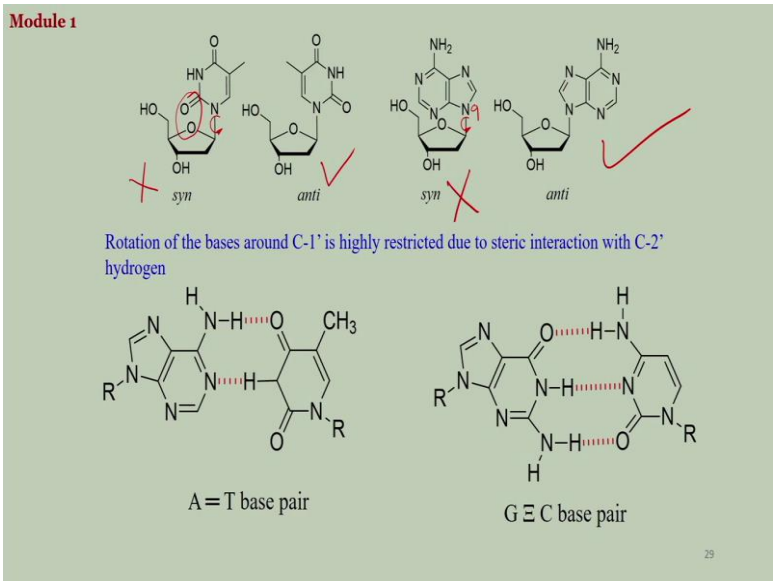


Similarly if you look at the adenine and guanine, let us consider one in first guanine has one entry position here as well as N7 here these two this is attached to the sugar so guanine entry has a pk sorry guanine N1 the pk of 9.42 and guanine N7 which is this, this is 7th position as a pk of 3.3. Now again using the same idea you can find out that guanine would exist in future under physiological condition the N1 would exist at with the hydrogen attached to it.

So, this form on the other hand this is 3.3 and 7 which means above 3.3 it will be deprotonated and this will stay as deprotonated form. So, this is the correct structure of guanine that will exist in DNA. Similarly adenine also you can find out adenine in one I have the value of 3.52. Similarly adenine N1 is this so that will exist in a deprotonated form. So, this would be the proper form that will exist in DNA.

So that is how you see that all the four nucleobases that we draw in this way they are the correct structures and that will exist under the physiological conditions.

(Refer Slide Time: 37:12)



Now comes the orientation of the nucleus base when the sugar is on. If you look at this structure this is thymine which is attached with the deoxyribose sugar. Now it can exist in two different conformer one is you see here with the double bond of O and NH the amide bond is in this orientations closer to the oxygen of the sugar. On the other hand in these structures they are away from the sugar moiety. The double bond which is planar system is closer to the sugar moiety here the double bond here is the away from the sugar.

So, these are the two confirmations that are possible because this C and N this is a single bond and therefore rotation around a single bond is of course feasible under room temperature. So, it can rotate and the bond ideally can rotate the question is which is the confirmation that will be stable or that will be more favourable. So, this form where the this oxygen and this oxygen are closer is called known as the Syn form the other one where the away is known as the Anti form.

Now if you look here you have a lone pair a lone pair repulsion which is absent in this case and the since this is a double bond double bond is planar flat. So, that do not exist any other steric effect also. So, here there is steric interactions as well as lone pair lone pair repulsion factor that makes this form syn form less favourable anti form is more favourable. And this can be clearer better when you consider the purine bases.

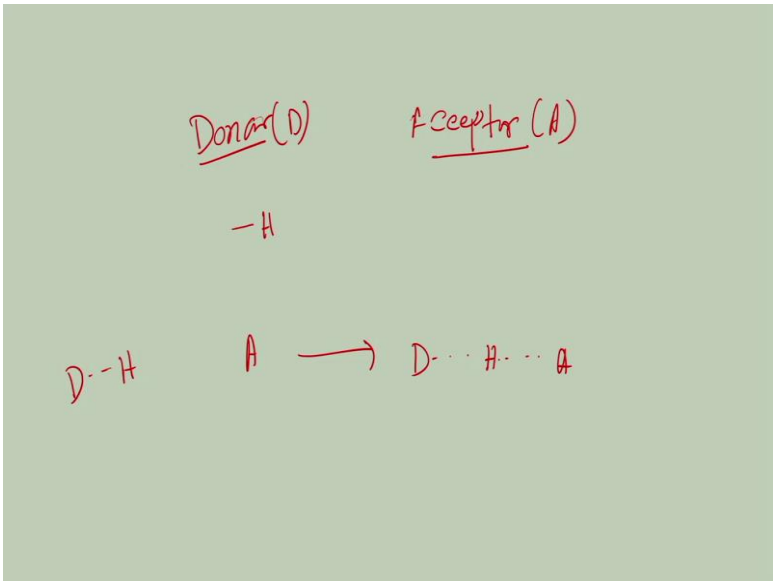
If you consider for example adenine this is the sugar it is the N9 of it N9 of adenine and then this is the adenine structure 5 membered ring fused with the 6 membered ring this is the syn form where the 6 membered ring is just falling on top of the sugar. So, very crowded on the other hand anti form is nice and smooth with the 6 membered ring away from the sugar moiety. This in the syn form you can understand that there are a lot of lone pair alone peripheral sense as well as crowding factor staric parameters are involved.

So this is not the favourable structures at all this would be the favourable structure. So, anti form or the anti confirmations are the favourable confirmations that will be existing in DNA. So all those factors they actually play vital roles in getting the ultimate specific structure of the DNA. So, what we know now we have seen that the structures of the nucleobases are in this form as you usually draw these this like this.

And the orientation of the nucleus base with the sugar is of the type anti and the confirmations. Now these two factors actually they play role in base pairing. So, they are the ones that gave you are the perfect base pairing property. So, this is how the base pairing looks like adenine we are safe with thymine guanine we are safe with cytosine. So, R is the sugar ribose sugar that is why I have written R this is the same sugar.

So now these are the two structures if you just compare how do you get the base pairings and what is the best possible way spreading to obtain is this. So, before that I will mention one more thing. So we have seen the thymine exist in this form the keto form here there is an N.

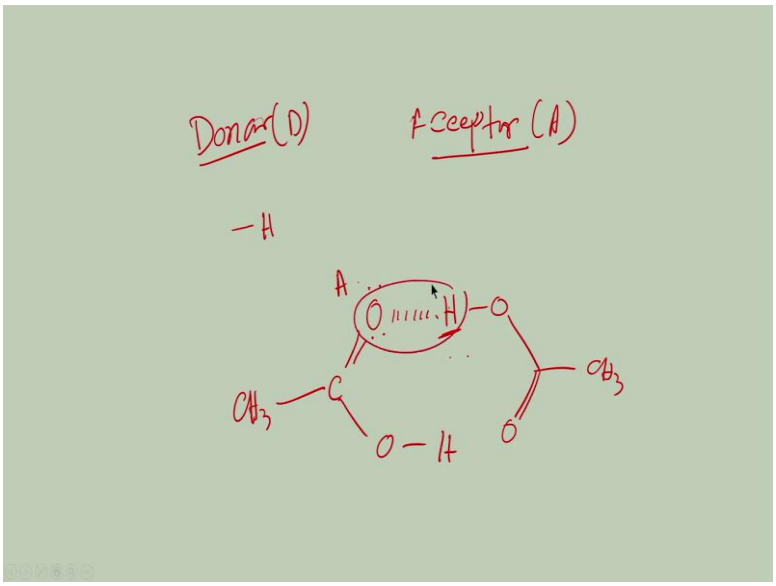
(Refer Slide Time: 41:17)



So, now when you talk about the hydrogen bonding then there is an two factors one is called the donor other is the acceptor. Donar represents D acceptor represents A of course you know all this I will just give a brief recap. So, if you want to do a hydrogen bond you need to have a donor which has a donatable hydrogen H that is donatable. Similarly the acceptor should have the capacity to accept the hydrogen when it basically lives kind of like a proton. So, the acceptor should have sufficient electrons to attract the H atom.

So, that is what the acceptor is a typical hydrogen bonding if you see usually we represent in D - H and then there will be acceptor then the hydrogen bonding that you get is like this H is leaving D and coming closer to A that is what the typical hydrogen bonding is I will give you one example.

(Refer Slide Time: 42:32)



So, the like acetic acid you know OH this is where the hydrogen bonding will take place because this oxygen has lone pair of electrons which can accept the protons. So, this is our acceptor on the other hand the acetic acid is an acid so this process proton is level it is the acidic proton it wants to move away so that is why this hydrogen bonding will take place. It is very strong hydrogen bonding. So, donor should have a donor table hydrogen and an acceptor should have electrons to form the hydrogen bond or to accommodate the proton.

So, here are the structures if you stick to that and the confirmations in the entire formations then you can see you have a donor here you have acceptor here do not here acceptor here. And for adenine you have a acceptor here and then I do not here I mean is the donor here N is the a nitrogen is free it has also lone pair of electrons so it can accommodate the proton. So, acceptor here a donor here acceptor donor this is an acceptor. But if you look at the geometry of this sugar then you can see this CO bond is actually away it is somewhat in the interior part closer to the sugar.

So that is why it does not participate in the hydrogen bonding. Hydrogen bonding comes from these two factors. Similarly same thing happens between G and C cytosine has 3 these this and this acceptor this is acceptor this is donor the this is acceptor this is donor this is donor, donor-acceptor donor-acceptor acceptor-donor so three hydrogen bonds are possible for GC if you try to do any other combinations you will find that this is the best combination possible. So, that is how the Watson-crick base pairings actually happen, thank you.