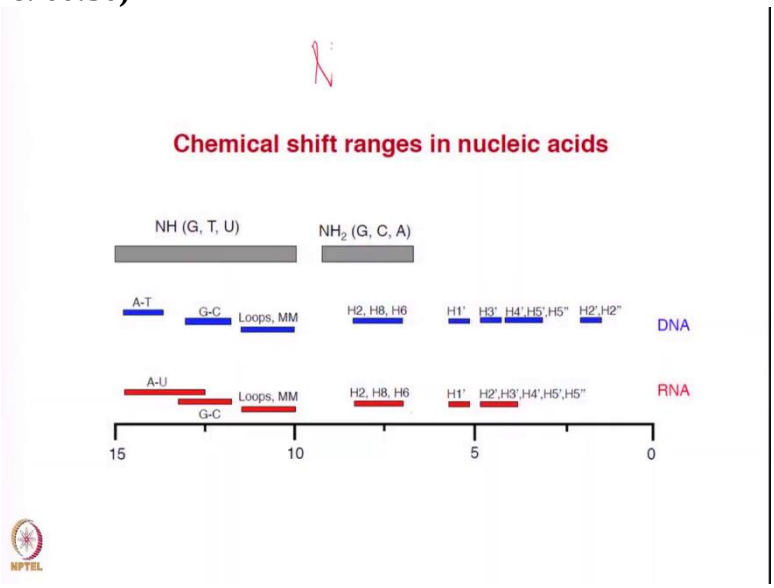
NMR spectroscopy for Structural Biology
Prof. Ashutosh Kumar and Prof. Ramkrishna Hosur
Department of Chemistry
Indian Institute of Technology - Bombay

Lecture: 27
Application of NMR in the Area of Structural Biology: Structure of DNA and RNA 4

(Refer Slide Time: 00:30)



So, far we discussed about the various aspects of the DNA structures. Now we are going to talk How NMR can be used in this. So, what is our first observation in NMR we have to talk about the chemical shifts proton chemical shifts we talk about the proton chemical shifts and what are the protons which are possible. We already mentioned about the various protons in the in the DNA and in the RNA in various places and I will show you here I have just listed them here.

And the proton chemical shifts as we have already seen can go from 0 to 15 ppm with respect to a particular reference typically when we all we record in water typically we use a TSP as a reference then you have this ppm 0 to 15 ppm which protons appear where. Now what are the protons which we have we have the sugar ring protons. Sugar ring protons are here 1', 2', 2", for the DNA 3', 4', 5', 5'' these are all part of the sugar ring 5" lie outside in the backbone.

And then these are the bases H2 H8 and H6 which H2 and H8 are present in both G and A and H6 is present only in C H2 H8 are present in both G and a and H6 is present in C these nomenclatures are important one has to remember this. Now with regard to the loops and these are the imino protons NH protons are the iminoprotons and these all on the in this area we also have the NH_2s GCA we have the NH_2 in all these 3 bases we have seen that.

When they when we are talked about the various base pairing schemes we saw that there are there are NH2's in each of one of these and they also appear in this region and this is what is this ppm? This is about 7.5. So, 5, 7.5, 10, 12.5, 15. So, around 7.5 area you have this 1.5 to 8 8.5 you have these base bases and then you also have the aminos. These are the GCA and you come in this area from 10 ppm to 15 ppm you have the iminos.

Iminos of G, T and U, G, T and U these iminos appear in this area and these are the ones which are present in the loops these are not hydrogen bonded these amino protons are not hydrogen bonded. Therefore these are mismatches these called as mismatches that are the loops. This hydrogen bonded one this is the GC pair and AT pair GC pair appears earlier. So, this is around 12, 12.5 and the AT pair appears later of course this 15 ppm is a little bit this is a wider range here typically you will find them around 14 things like that 14, 14.5 that is where you will find them.

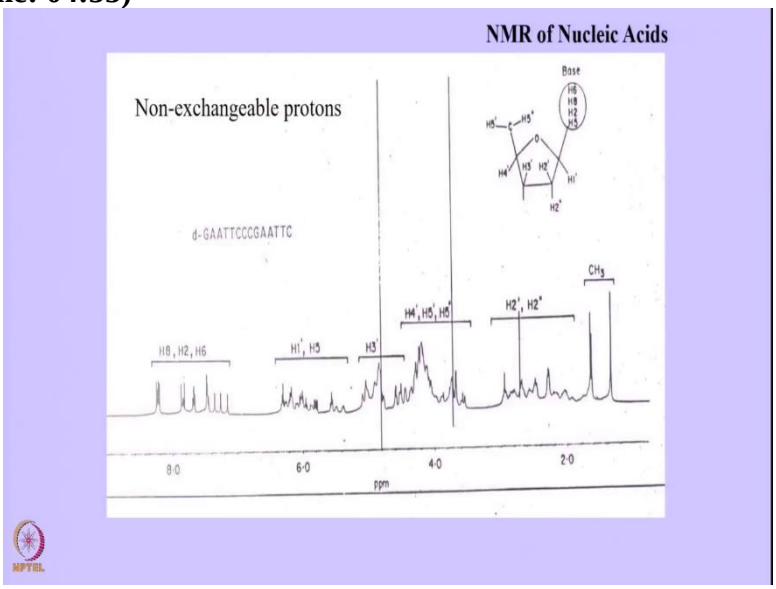
And you have this GC pairs appearing in this area and the ATs appear in this area and these are all at pH7 around pH7 because it is important to mention the pH here because the pH plays an important role in determining what sort of a base pairing can occur because there is a protonation which a protons are exchanging and these are exchangeable protons remember these ones are exchangeable protons.

So, if you want to observe this you will have to record spectra in H_2O and not D_2O these spectra these amino products whereas these ones here these are the non-exchangeable protons. So, these are attached to the carbons these which are attached to the carbons these can be observed in D_2O, D_2O solutions as well. Now if you come to the RNA, RNA does not have H2 prime H2 double by it has only one H2 prime a true double prime does not exist in RNA.

Therefore now you see because of that because there is an OH group at the 2 prime position this H2' gets shifted to this area it comes to this area H2', H3', H4'. So, this is quite a crowded area in the case of RNA the H1' is not affected. So, much this is also not affected. So, much and all the other things remain the same by and large therefore the proton chemical shift ranges differ between DNA and RNA in this area.

This becomes quite distinct in the case of DNA whereas RNA there is quite an overlap in the in this region all right.

(Refer Slide Time: 04:55)



Now here is a typical spectrum of a nucleic acid the segment which is given here. So, you have a deoxy this is the DNA segment. So, it is a 14-mer here G A A T T C C C G A A T T C and this is the sugar ring all the nomenclatures are given once more here. So, you have the one prime and in the basis you have this H6 H8 H2 and H5. H6 and H5 are present in the cytosine H8 and H2 are present in adenines and on the guanine you only have a H8 you do not have the

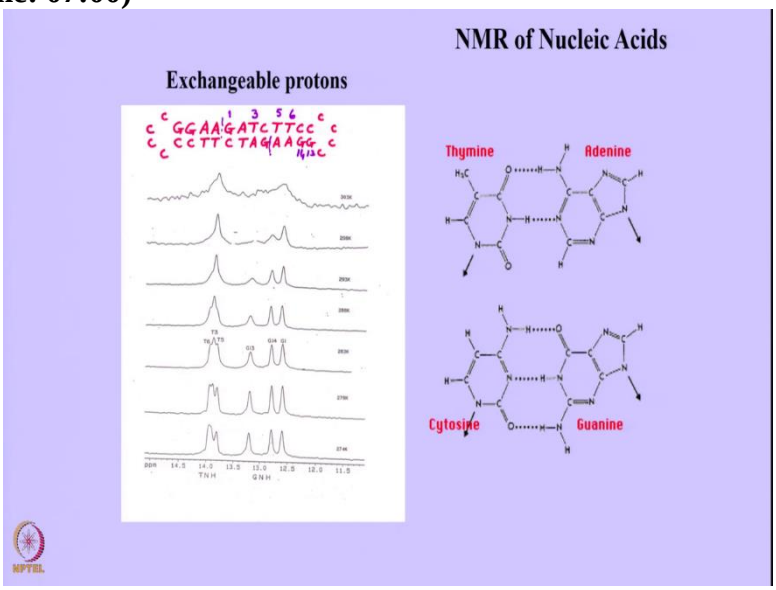
H2. At the H2 position there is amino group in the case of guanine. Now where do these ones appear in the spectrum this is the one dimensional spectrum here.

And all these H8, H2, H6 they are present in this area and the H1 primes and the H5s they appear here and the H3' are here at around 4.5 ppm here and all these H4 prime H5 prime H5" they all get crowded in this particular area the extremely corroded region. You notice for all the 14 nucleotides they are all present in this similarly and all the 14 nucleotide 2', 2" are present in this.

And the methyls here where are the methyls? The methyls are coming from the thymines the thymines have at the H where in the case of cytosine at 5 position that is the H5 in the case of thymine there is a CH 3 group here. So, How many thymines are there here? You have here thymines 1 2 3 and 4 there are 4 thymines and these 4 thymines are present in this area you can actually come because these are quite distinct you have 2 here and 2 more here.

So, all the 4 thymine methyls group you can identify in this area of course you do not know sequence specifically which one is which but the thymines are easier to identify because they appear and the methyls appear at a very distinct position in the chemical shift range in the proton. So, this is the typical one-dimensional spectrum.

(Refer Slide Time: 07:00)



What about the amino protons. Now the imino protons as I mentioned to you they appear between 12 to 15 ppm. So, you have these imino protons for these protons these ones have to be observed in water the experiment will have to be performed in water H_2O. Otherwise in D_2O these ones exchange out and you will not be able to see those signals and you see this particular sequence here this is it goes G A T C T T C C C C C G G A A this is the sequence.

So, this sequence has a kind of a two molecules form a stem of a duplex and then it forms a loop here this is and these C's will not be observable to you because they exchange out with water these forces will not be observable to you. Because they exchange with water and this forms a symmetrical duplex here and with the Watson Crick base space. The Watson Crick base pairs are indicated here the thymine adenine and cytosine guanine which are the protons which are observable and those are the iminoprotons.

The iminoprotons appear in this range as I mentioned the amino protons appear around 10 ppm. So, 9.5 to 10 ppm if though and those ones exchange more rapidly with water therefore often you do not see these amino protons you will basically see the iminoprotons. And iminoprotons will produce one iminoproton one signal per base pair because there is only one amino protons whether you take this one whether you take this base pair or this base pair there is only one imino proton.

In this case that is this one here that is that N3 position and in this case there is a guanine which is at the N1 position. So, there now we can see here how many signals are there. So, you can actually count this is the different temperatures this is at 278 degree Kelvin and you slowly small temperature you can go up in the temperature. Why do we do this temperature dependence because you see the duplex here right.

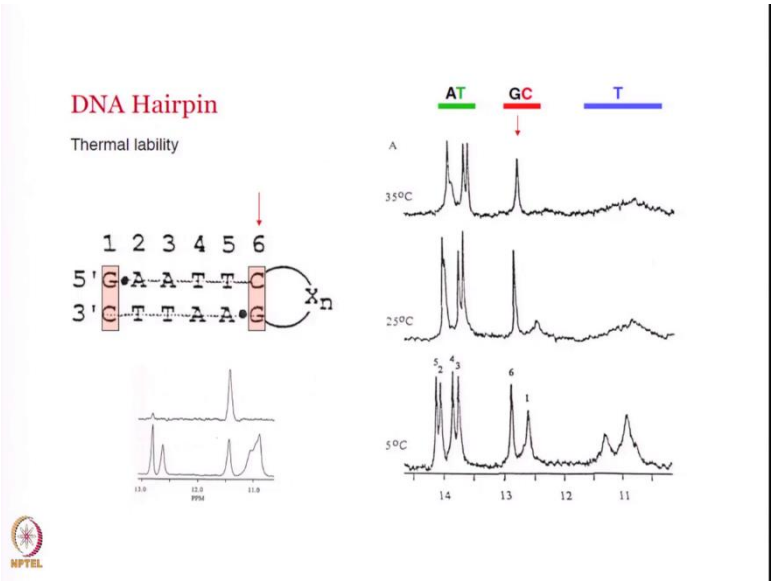
This is a duplex the duplex is the stability of the duplex is the highest in the interior and as you go further towards the end it becomes fragile. So, it is less stable. So, therefore when you in an initial at the lowest temperature you will see all the signals you will see how many CG's you see here you see 3G's these all 3 these belong to the G's there is a symmetrical structure here. So, you see this G this G and one of these G's. So, you see 3 G's because this become at the base of this loop this is not so, stable.

So, you will see that one exchanges out. So, you see these 3 G's which are present this belong to the interior of the duplex and these are the T's which are present that is this one you have 1, 2 and 3 3 T's which are present here you have all of them present and this point they are pretty well seen. Now as you start increasing the temperature you see that the duplex which is present here starts melting.

It starts melting slowly this imino proton is small melting meaning one the base pairs start separating out. Once the base pairs separate out then they become exposed to the solvent to the exposed to the water then they will start disappearing. So, at this temperature this is see that everything is exposed and all of them are exposed they are exchanging with water. So, the completely the lines are broadened out.

If you go still higher there will be nothing seen everything will be the no signal will be present in this. Therefore to find out whether your structure has the duplex or the base pairing has happened or not you have to do the experiments at low temperature and monitor the imino protons.

(Refer Slide Time: 10:46)



And here we show the thermal stability how you can measure determine the thermal stability of a DNA hairpin you have this hairpin here G A A T T C and then you have the X n there are many other base bases here which are in the loop and you will not find signals from there but the ones which are paired here you will find the signals. For this is the one particular molecule which is going around and forming a hairpin structure.

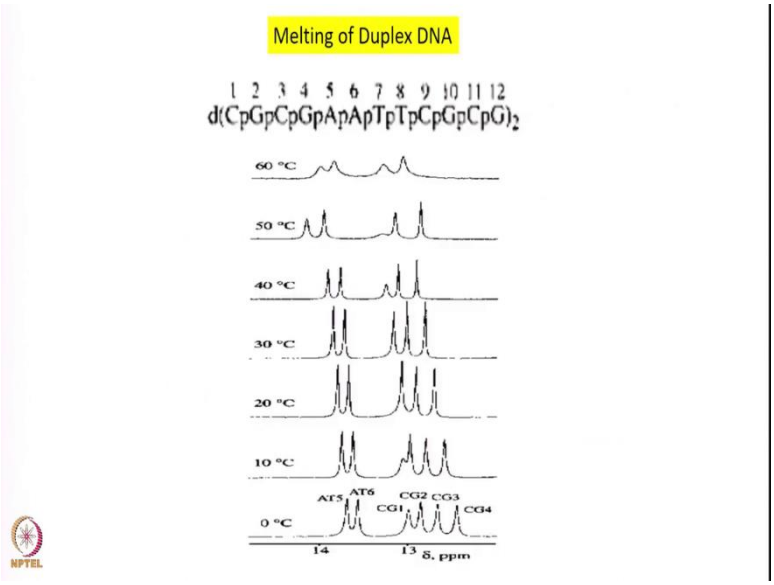
You will see peaks for these ones you will see peak for this G and you will see peaks for the T's and those ones are the ones which are present here and you can see as the temperature is increased this is 5 degree centigrade 25 degree centigrade and you have this 35 degree centigrade and so, this the positions are indicated 1 2 3 4 5 6. So, these are the 1 2 3 4 5 6 these are the ones which are present here.

So, you have the G's which are present here at position this is at position 1 and so, this see the sequence is going like this, this is the single sequence and this G and this G are not the same they are different this comes through the loop here. And this G and this is the position 6 G that is this one there and then you have the other 2 3 4 5 these are the AT base pairs a 3 these are 2 3 4 these are the AT base pairs.

So, therefore these belong to the thymine and 3 iminos. So, those ones are distinctly present here and these ones are the aminos. So, these are belong to the aminos and various places and these ones you do not and you do not see them very clearly even at 5 degree centigrade you see they are all quite broad as increase the temperature these ones will disappear and you have only the iminos which are present in this AT GC pairs.

And these are possibly the T's X probably is a T here. So, you have this T's which are appearing in this which are not clearly observed this is the DNA hairpin.

(Refer Slide Time: 12:47)



And this is for the duplex melting of the duplex DNA indicated another example of the duplex DNA here you go from this particular sequence you have the C G C G A A T T C G C G this is a self complementary sequence you write the other way around here. So, it forms a symmetrical duplex with the twelve base pairs. So, the twelve base pairs you have. So, you have How many you can see 6 of the 4 G's C G C G and C G C G there are 4 G's.

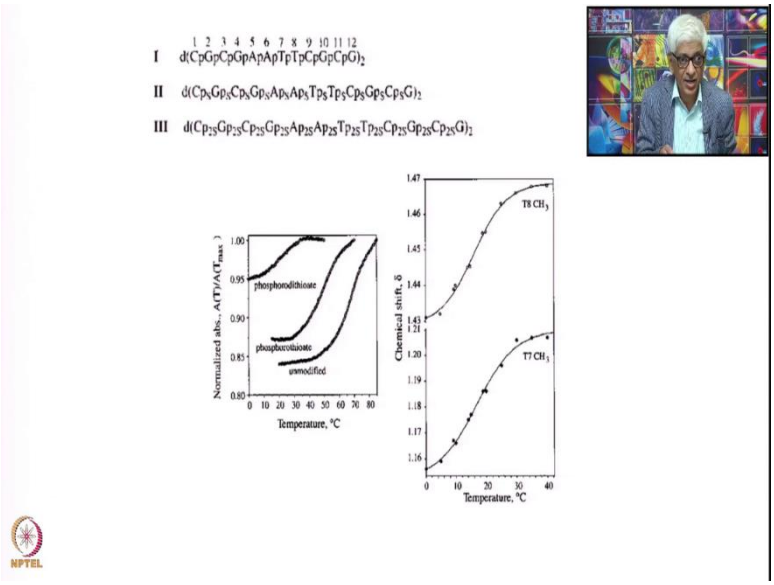
And How many T's? There are 2 T's AA TT there are 2 T's because we are going to see signals in the amino proton spectral region you will see either from the G or from the T. So, therefore you have 2 T's at zero degrees you can see both of them and you can see all the 4 G's these are the 4 G's which are present here at zero degree centigrade and as you start increasing the temperature.

So, they start melting which goes first you see it is this one goes first right this one is going first what does that mean this will tell you that this is the terminal G okay. So, this terminal G is actually going first. Slowly this is How you identify there as the DNA melts as the looking at How the imino protons are disappearing you will identify them as belonging to which T and which G.

So, at 60° almost all of them are vanished and you can see this continues CG1 CG2 CG3 CG4. These are the base space these are the base parts which are indicated there. There is the symmetry there; there is a symmetry. So, from here to here and from here to this is the duplex because it repeats itself in this manner right. So, various base wave what is indicated here is the base pair which base pair is appearing the first base pair will disappear this is the first base pair this is the extreme end.

So, that is the one which will first disappear then followed by 2 then followed by 3 followed by 4 of course the 4 is quite in the interior and that will be the last one to go. So, this stays here you see all the way up to 50 degree centigrade and this is how melting state these were the standard techniques used earlier to find out what is the melting temperature of the DNA.

(Refer Slide Time: 15:08)



So, here then you plot here the melting temperatures of the DNA. So, you have various kinds of modifications then if you make a modifications in one in your DNA does it increase the stability of the DNA or decrease the stability of the DNA. This you will figure out by looking at the melting temperature. What is the meaning of melting temperature if you plot this normalized chemical shifts of the iminoprotons.

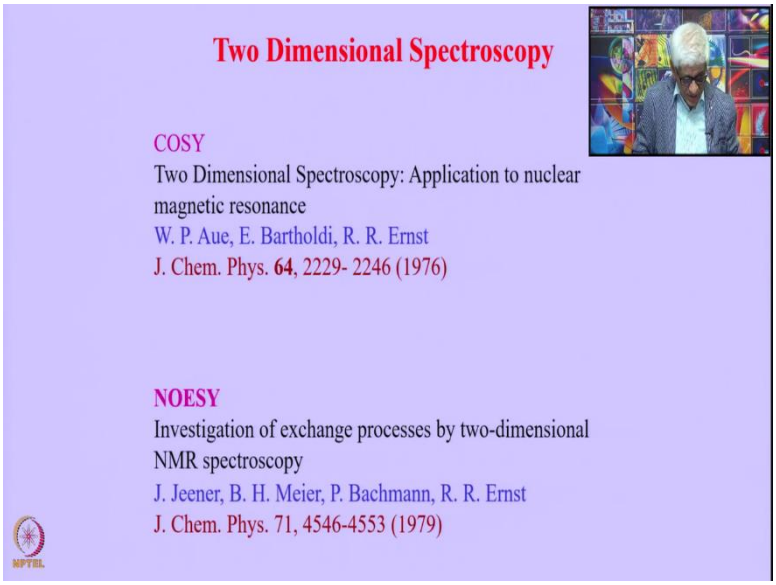
Against the temperature here you will get a sigmoidal curve in this manner and the midpoint of this will give you the melting temperature. So, at that point of course you will have very poor iminoproton signals. And by looking at the melting temperature you can decide whether the DNA is stabilized by a particular modification this is unmodified here and this is modified here at a particular place these are indicated here at place the modification is done.

And then of course you have another one and the thiol means is actually modified by a sulphur group instead of the oxygen you have a sulphur group here you see this is 2s 2s. So, normally you have an oxygen but these are all modified here. So, all the sulphur groups are present here. And here is a normal one and these are the modified modifications at this point you have a sulphur group here and then you have 2 sulphur groups at this point.

So, both of the oxygens are replaced by the sulphur and how does that change the temperature. So, you see when you have oxygen replaced by the sulphur group the temperature has come down quite a lot here. So, it is almost 10 degrees. So, it is very unstable the normal DNA is of course the most stable one because it has the highest temperature the melting temperature is almost about 65 degrees.

So, as your one sulphur one oxygen you replace by the sulphur you have this temperature comes down to 35 degrees or 40 degrees 2 sulphurs it will even goes even further.

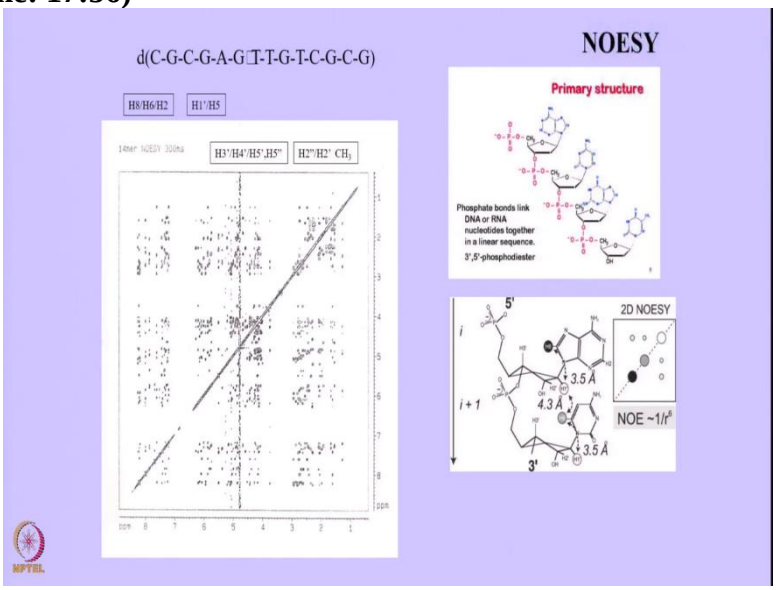
(Refer Slide Time: 16:59)



So, this is this was all that could be done earlier with a single one dimensional NMR spectra. So, melting temperatures see the stabilities of the modifications stabilities of the DNA and what sort of a structure the particular molecule is having. One also studied the interactions with various other small molecules depending upon the from the one dimensional spectra. But then there is the revolution with this 2 dimensional spectroscopy which allowed us to identify the individual protons in the DNA segments.

Non-exchangeable protons we could monitor because in the normal case this non exchangeable protons were not very difficult to monitor because of the extensive overlap of the signals. But in this 2 dimensional spectroscopy this came in the 1970s it became possible to study the non exchangeable protons as well. So, this is the theories of these ones we have already discussed we are not going to go into that.

(Refer Slide Time: 17:56)



So, here I am showing you one particular 2 dimensional spectrum this is a NOESY spectrum. The so, called NOESY spectrum which of a particular DNA segment that is indicated here. So, A C G C G A G T T G T C G C G So, this is a 15-mer you have this 2 dimensional NOESY spectrum. So, you have the diagonal here and see and this area as a as we discussed here this is the one dimensional spectrum this belongs to the 2 prime 2 double prime in the CH 3 groups.

See I also put the primary structure here once more to be able to identify which protons are where and then I have also shown here this H3 prime H4 prime H5 prime H5 double prime those ones appear in this area and next to that is the H1 prime and H5 this belong to the cytosine and here we have a H8 H6 H2 this whole area belongs to the H8 H6 issue these belong to the bases for the various bases. Now what does this tell you.

So, these peaks here this peaks here this tells you about the base-base interactions. So, if there is a base stacking when there is a base stacking you have 2 protons on top of each other and you will find adjacent base pairs you are showing proton-proton interactions through the NOESY short distances and these ones will show you base to this region and what is this region this region is the H1' and H5 base to H1'.

And this is this goes to this area. So, this goes to the H3 prime base proton to the H3 prime correlations sequential correlations distances. All of these depend upon which are the short distances here and these ones goes to the base protons to the H4 prime H5 prime H5 double prime and this goes from the base protons this area is to the 2', 2" and the methyls. So, this is the very well resolved.

Now pretty well resolved of course you also have all from the 1 prime to the 3 prime here and then 1', to the 4', 5", 1', prime to the 2', 2" you will see here then of course from the 3' you will see also the 2', 2" and 4', 5' also to the 2", 2'. So, and then here you have the 2', 2" within that within the same base within the same sugar.

2 prime to double prime these are short distances and therefore you will see NOE cross peaks between these protons. So, 2 prime 2 double prime is actually geminal. So, it is about 1.8 angstrom therefore you will always see this, this is a very short distance. So, therefore you will see all of those ones in this area and you will see from the base protons the methyls are present here from the methyls you will see to the for 5 prime 4 prime 1 prime and then to the other bases H8, H6, H2.

So, therefore the amount of information that is present here is quite substantial quite enormous. Therefore using this you can actually identify the individual proton signals of the individual residues. So, this actually made a big change here. Now here I show you the 2D NOESY what sort of things you will get the NOEs are dependent on the distance between the protons the distance between the protons and that is in the NOE intensity or NOE means I same the cross peak intensities here.

The various cross peak which are present these are proportional to the inverse the 6th power of the distance between the protons therefore the distances becomes quite important which distance is short which distance is long that becomes important in figuring out which peaks are likely to come and it turns out here I will show the distances here this is the nucleotide i this is the sugar ring of the nucleotide i.

And this is the sugar ring of the nucleotide i + 1 these 2 sugar rings are indicated here this base proton this is another base proton these are in the anti conformation i indicated you the anti conformation this base proton comes closer to the oxygen this comes on the other on above the sugar ring. In the syn conformation this goes other way around this goes on that side. So, you will not be able to see that.

So, in this confirmation you can see in the anti confirmation you will see peaks from here to its own sugar rings to own sugar protons but you will see also from the base proton one base proton to the sugar ring of the next one on the 5 prime end. Notice this is at the 5 prime end this base proton to short distances are there to the sugar protons of then of the nucleotide which is at the 5 prime end.

So, this one to this you will not see from this proton this to this sugar you will not see that that is the long distance you will not be able to see those ones. So, first of all you will see within this nucleotide from the base proton to its own sugar rings okay and then from the base proton to the sugaring protons of the residue adjacent to it on the 5 prime side. Therefore that is the important information present here.

If you look at this at each base if you look pick out this area here this is the base proton to the one prime protons. This is the base to base distance the base to base protons cross peaks. So, this will tell you, you can just simply walk from here one residue to another residue one nucleotide to another nucleotide following this base-base interactions you step wise you go from one base to the next to the next to the next and you can go up like that.

And the same thing you will figure out from this as well or this as well or this as well. So, you have from the base proton to the one prime of the same risk of the same nucleotide and to the 1' of the nucleotide at the 5' end. Remember not to the 5 nucleotide at the 3' end therefore this provides a directionality while this one does not provide directionality this base to one prime.

And the base to the sugar one prime 2' or 3 these ones will provide you directionality of to the sequential walk which you may go through in the case of the DNA. So, you will see 2 peaks from each base proton one to this one 1' and to the one prime on the 5' end.

(Refer Slide Time: 24:29)

	RNA		DNA		
	A^1	A'^1	A^1	B^1	Z^2
C-H6/C-H1'	3.6 Å	3.6 Å	3.7 Å	3.7 Å	3.7 Å
C-H6/C-H2'	4.0 Å	3.8 Å	3.7 Å	1.8 Å	3.1 Å
C-H6/C-H2"	--	--	4.5 Å	3.4 Å	4.3 Å
C-H6/C-H3'	2.9 Å	2.9 Å	2.5 Å	3.8 Å	4.5 Å
G-H8/G-H1'	3.8 Å	3.8 Å	3.9 Å	3.9 Å	2.6 Å
G-H8/G-H2'	4.2 Å	4.0 Å	3.9 Å	2.1 Å	4.1 Å
G-H8/G-H2"	--	--	4.7 Å	3.6 Å	4.7 Å
G-H8/G-H3'	3.1 Å	3.1 Å	2.8 Å	4.2 Å	5.3 Å

(b) Calculated internucleotide proton-proton distances for the base protons of nucleotide (n) to the ribose protons of the 5'-adjacent ribose ring (n-1).

		RNA		DNA		
n	n-1	A^1	A'^1	A^1	B^1	Z^2
C-H6/G-H8		5.4 Å	4.9 Å	4.9 Å	5.2 Å	6.1 Å
G-H8/C-H6		5.1 Å	4.6 Å	4.6 Å	4.8 Å	5.3 Å
C-H6/G-H1'		4.6 Å	4.1 Å	4.0 Å	2.8 Å	3.7 Å
C-H6/G-H2'		1.9 Å	1.6 Å	1.5 Å	4.0 Å	3.1 Å
C-H6/G-H2"		--	--	3.0 Å	2.5 Å	4.3 Å
C-H6/C-H3'		3.4 Å	3.4 Å	3.5 Å	5.1 Å	4.5 Å
G-H8/C-H1'		4.7 Å	4.2 Å	3.9 Å	2.9 Å	6.5 Å
G-H8/C-H2'		2.0 Å	1.7 Å	1.5 Å	3.8 Å	7.1 Å
G-H8/C-H2"		--	--	3.1 Å	2.4 Å	>7.5 Å
G-H8/C-H3'		3.3 Å	3.3 Å	3.3 Å	5.1 Å	6.6 Å

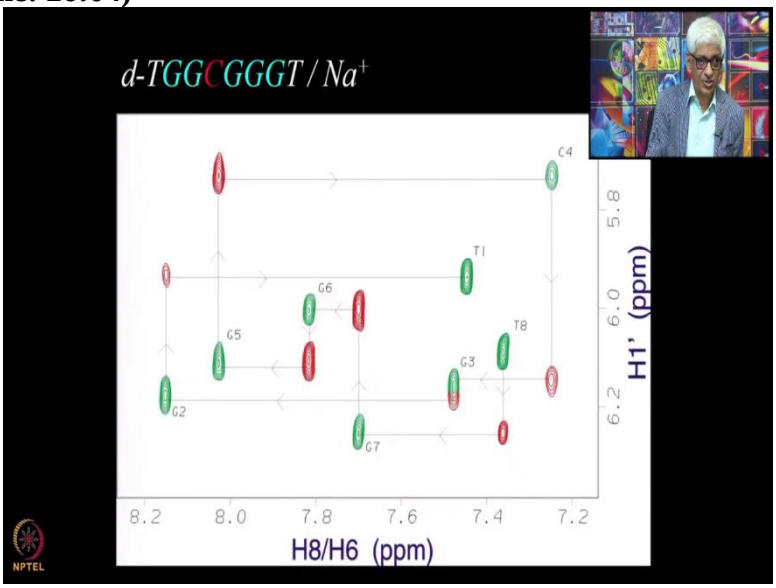
[1]Distances calculated on basis of the torsion angles given in ref. 33. Note that some of the interproton distances are unrealistically small (<2.0 Å) due to the assumption of a rigid, uniform geometry and the omission of taking hydrogen atoms into account in the derivation of the A- and B-models. In reality these distances will certainly increase to values in the range of 2.0 - 2.3 Å.
[2]Distances calculated on basis of the torsion angles given in refs. 6 and 35.

Now here is just a listing of the various short distances which are present in the case of RNA and in the DNA. And these are important which are the short distances this is about 3.6 angstrom 3.6 and H6 to H2 prime these are the self within the same within the same nucleotide unit these are the ones which you will actually see. So, these distances you see they are all less than 5 angstroms 3.6, 4, 3.8 and things like that therefore you will see all these peaks.

In the DNA these distances are roughly similar distances which are given is less than two angstroms are not valid this there is a some errors there in those one that is indicated here unrealistically small due to the assumption of a rigid uniform geometry and the omission of taking by hydrogen items into account from the. These are generated from the fibre diffraction models. So, therefore the ones which are less than 2 angstroms are not really meaningful.

So, because there is a when the it is less than the means there is a steric overlap and that is not practical but the other distances are other valid. So, you will see these distances they are in the range between 3 to 5 angstroms. You will see all of these distances these are the sequential distances sequential meaning from C to G, G to C, C to G, C to G see these are all the distances between 2 nucleotide units one nucleotide unit to the nucleotide unit at the phi prime end. So, you will see all of those distances there.

(Refer Slide Time: 26:04)



Here is an illustration here is an illustration of a particular portion of the spectrum you protect this particular segment here T G G C G G G T. In fact this actually forms a quadruplex but so far is the analysis is concerned one strand analysis remains the same in either case. So, you see here what is present here these are the H8 or the H6 proton of the base H8 or H6 proton of the base okay. And H2 proton does not produce NOE's.

H2 proton is far away. When the H8 is close to the sugar ring the H2 proton is on the other side of the purine ring that will be further away. So, you will not see a NOE cross peak from the H2 protons to the sugar rings you will always see from the H8 or likewise you will see from the H6. Now you see here the cross peaks from but every particular cross p every particular base you will have 2 cross peaks one to its own one to the sequential.

For example if this is this the self peaks are indicated by the green and the sequential peaks are indicated by the red. So, for example if I start here so, I get here from C 4 self peak its own then I go to the sequence I get a sequential peak here. Now from here I move horizontal to find a base which produces the cell peak which produces a green peak. So, this is base proton of C4 this is the H6 of C4 from the H6 of C4 I find a H1 prime to its own and then the H1 prime of the residue on the 5 prime end.

The 5 prime end means it will be G3 it will be G3. Suppose I took this, this is a 5 prime end. So, this is G3 I will see this correlation from C4 to G3. So, C4 to G3 I will get here. Now I go

horizontal to find the base proton of G3. So, I get the green signal here this is the one which is the base proton of G3 and it turns out that G3 sequential peak is also close by also on the same one. So, then overlap here because this portion of the signal is in line with this G2.

Where I should go from G3 I should go to G2. So, therefore from G3 I go here sequential peak I go to the G2. So, go to the G2 then I have a sequence this is the self peak then I get the sequential peak here from G2 this is to the T1 from the T1 so, I go horizontal again to find the self peak and then T1. Now how do I know I am correct here. Now you see once you reach the end T1 I do not have any other residue on the other side therefore there will be no signal below this.

There will be only one peak there only the self peak therefore this confirms that what we have done here is correct. Therefore this is the T1 base and there is nothing below here there is no sequential peaks. So, you can actually continue in this manner once you have that. So, you go from C4 you go Horizontal where do I go? I go to the G5. This is the sequential peak from G5 to C4 this is the sequential peak.

Now I go from here to find the red peak here I get the G6. So, from here to here I go I go to the G6. Then I go to the red one here I go to the G7 then I go from G7 I go to this red one here and go to the T8. So, therefore from the T8 I have T8 to G7 G7 to G6 G6 to G5 G5 to C4, C4 to G3 G3 to G2 and G2 to T1. This completes the cycle this is a beautiful demonstration of How the NOESY spectrum can be used to identify the individual protons in the individual nucleotide units. I think we can stop here.