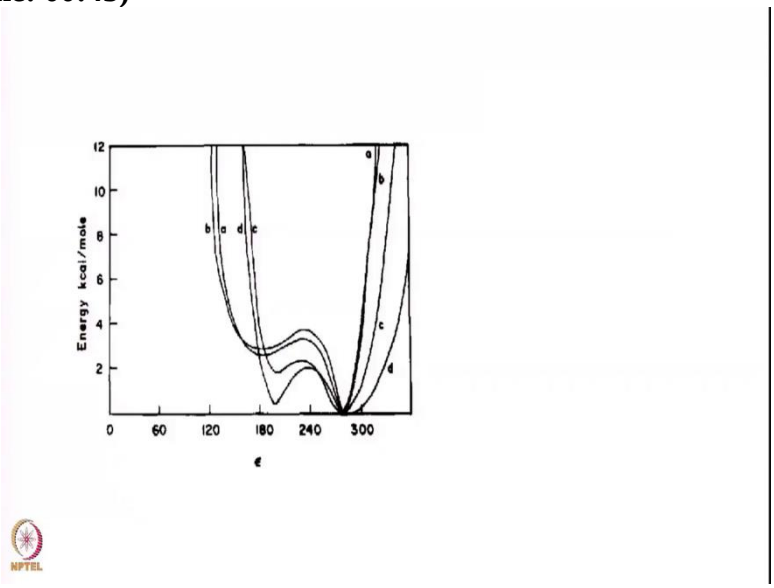
NMR spectroscopy for Structural Biology
Prof. Ashutosh Kumar and Prof. Ramkrishna Hosur
Department of Chemistry
Indian Institute of Technology - Bombay

Lecture: 30
Application of NMR in the Area of Structural Biology: Structure of DNA and RNA 7

So we will continue with the structure calculation of the nucleic acids. So, we had this input parameters coupling constants which translated into the torsion angles of the sugar geometries then we also have some information from the energetic of the ε torsion angle and then next we have the information from the NOEs the NOEs information is in terms of the distances between the protons the various protons inter proton distances.

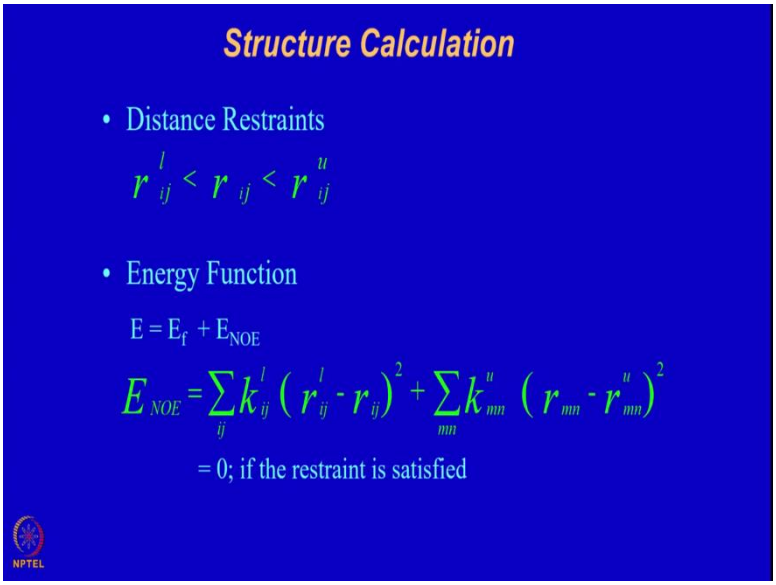
(Refer Slide Time: 00:45)



So, this it comes from the NOESY spectra. So, we have the input parameters let me write here for the structure. The structure we have the sugar geometries then you have the energetics then you have also the NOEs. NOEs will give you the inter proton distances which are the inter proton distances these are the various protein in the in the NOESY spectrum. These are derived from the noisy spectrum and you know the peak intensity is proportional to is proportional to one over sixth power of the distance.

But we get certain distance ranges. So, ranges we do not say we get exactly identical exact distance that this is 3.5, 2.5 and. So, on but we get ranges of distances and these are our input parameters now then what you do after this we use this in a kind of an algorithm and that is indicated here.

(Refer Slide Time: 02:03)



So, you have here the distance constraints you start with the particular model in your particular model the distance between 2 protons i and j is let us say r_{ij} now from the NOESY spectrum and you have a cross peak between i and j from the cross peak intensity you say well this distance should be between this lower limit and this upper limit. So, this is the lower limit of the distance in the upper limit of the distance.

I mean upper limit anyway cannot be more than 5 angstrom this is the highest value is 5. However on the basis of the intensities we can say that this distance should be between 2.5 and 3.5 or it can be between 3 and 4 or between 3.5 and 4.5 that kind of ranges we can give for the different peaks depending upon their intensities. These are called as distance restraints then what we after that you do a calculation to calculate the energy of the whole molecule.

Your energy consists of 2 components one is the so, called in intrinsic energies these are the various bond distances bond angles torsion angle energies and things like that then you have the NOE potential. You define an NOE potential this will tell you this is artificially introduced potential and that is defined in this manner. So, you have 2 components to it this potential this portion comes reflecting the deviation of the distance in your model from the lower bound.

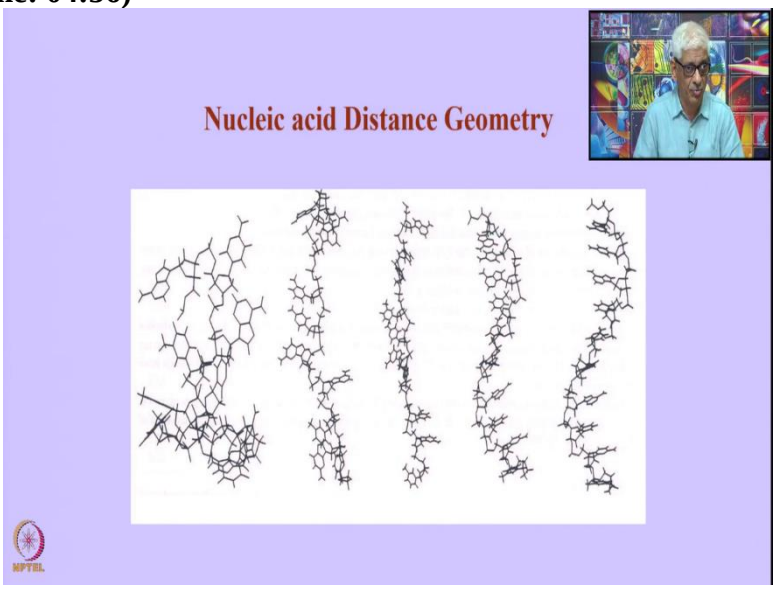
These are harmonic potentials here hormonic oscillator potential functions. So, deviation from the upper lower bound you take the scale you have the force constant here like a spring force con this like this is the force constant there. And then you have a deviation from the upper bound upper bound distance this much and you calculate this for all your distance pairs. Let us say you have thousand distances for each one of those distances you calculate this.

For the thousand distances meaning your thousand distance range restraints there each one of them I have an upper bound and a lower bound. So, for each of those these pairs of the distances I calculate what is the deviation from the expected value. So, you have this particular force constant which is kept there and then you define this potential function for violations of the distances for the lower bound and violation from the upper bound.

If the violations are all removed then your energy function has to be zero. Therefore what you need to do on your computer you have to keep optimizing your structure on the computer until you reach a stage that all the existence constraints defined here are satisfied once they are

satisfied then you have ENOE is defined as zero and that will be an acceptable structure for you once you have reached that stage.

(Refer Slide Time: 04:56)



And this is done by what is called a particular algorithm called as distant geometry. Now distant geometry algorithm takes a certain set of distances and calculates optimizes the structure to satisfy your distance constraints. For example to illustrate this point you let us say I have a configuration which is like a triangle if it is a triangle like this there are 3 atoms in a triangle how many distances I have here?

I have one distance here another distance there and the third distance there. So, if I define the 3 distances then this triangle is uniquely defined. So, if I have a quadrilateral here let us say then I have 4 distances 1, 2, 3, 4. If I define the 4 distances the quadrilateral is uniquely defined. So, in the same manner to define the structure of the entire molecule I must have a certain number of distances and which will define the structure of the molecule.

Therefore you have to develop an algorithm this algorithm is defined called as a torsion angle distance geometry. This algorithm what I am describing here is called as TANDY this was developed in our lab by a student who called Ajay Kumar and TANDY the torsion angle based distant geometry. So, you start from a particular structure here. So, this is a very random structure these are not necessarily energetically most favourable.

As some kind of a random structure you have for a particular molecule this is the 12-mer DNA segments here and then you put in the constraints and allow the molecule to optimize the structure allow the algorithm to optimize the structure by varying the 6 torsion angles along the backbone and varying the sugar geometry. The sugar geometry has to satisfy the constraints as you mentioned here from the coupling constants.

You can also define the sugar geometry in terms the distances corresponding to those sugar geometries and these are within the sugar ring 1 prime 2 prime 2 double prime etc. You can translate the sugar pucker in terms of the distances as well from what you get from the coupling constants. So, you put in those constraints there and then you optimize the structure in the computer. So, that you have ultimately you have to reach there.

Now you see these are various intermediate steps how the structure is changing you start from here the structure is changing as optimization is going on and eventually you see it come out with a very beautiful helical structure there. Helical structure now it is a single strand here in this case it was a single strand. Now what you do is after that you can add constraints of the hydrogen bond base pairs then you can put base pair constraints.

Once you put the base pair constraints and calculate the structure then you will get the duplex structure for the molecule depending upon the sequence what you have. So, therefore you will have to start with any arbitrary structure start with an arbitrary structure randomized values of the torsion angles randomize sugar geometries. These are not necessarily the most stable structure the energetically there may be quite a bit of steric contacts etc.

All of that may be there now the E_f part in the sugar geometry function which I have here. So, E_f part which is there if there are steric contacts etc this will eliminate those because this also has to be taken care this particular portion involves all the steric contacts and things like that the NOE part is taken care only in the in this part. So, therefore but this steric contacts also had to be satisfied.

So, when you do this put all that together in your algorithm here distant geometry algorithm. So, you reach a structure which is a beautiful helical structure satisfying the all the NMR constraints.

(Refer Slide Time: 08:41)

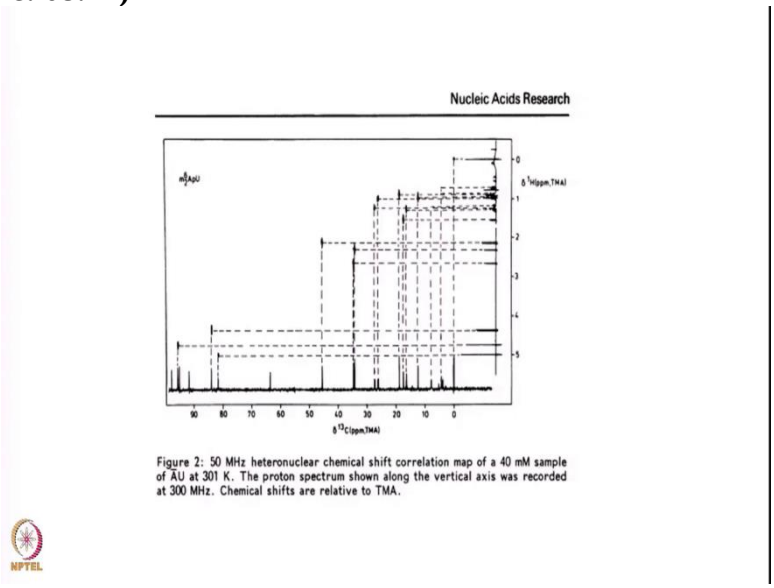
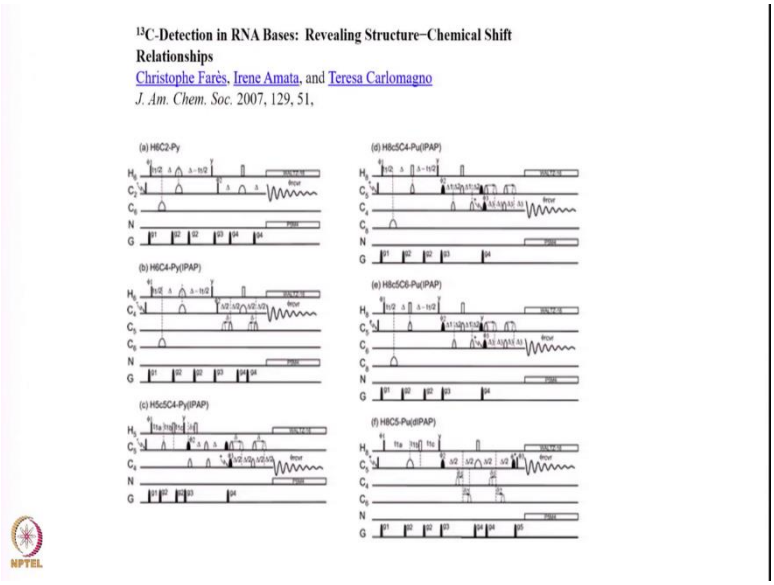


Figure 2: 50 MHz heteronuclear chemical shift correlation map of a 40 mM sample of AU at 301 K. The proton spectrum shown along the vertical axis was recorded at 300 MHz. Chemical shifts are relative to TMA.

So, that is from the structure calculation. Now certain times the hydrogen bond base pairs base pair S information can be obtained from looking at the Heteronuclear NMR. Heteronuclear NMR therefore you will have to use carbon 13 NMR or nitrogen 15 NMR but this will require labeling of your nuclei with carbon 13 and nitrogen 15. And this will have to be obtained by a specific synthesis synthetic procedure and these are certain examples shown here.

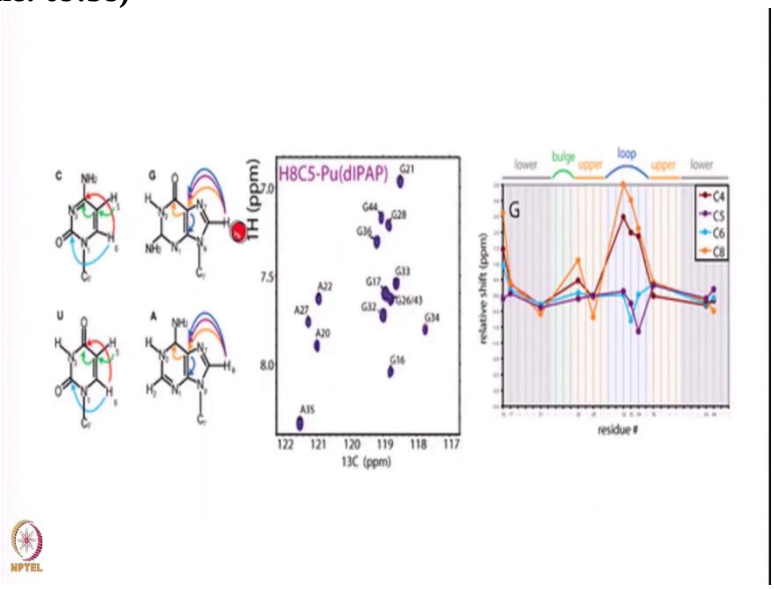
I will not go into the details of these ones a typical example of the spectrum of a particular molecule here and you have the carbon chemical shifts in this area and the proton chemical shifts in this area and all these cross peaks which will identify this helps you to identify those cross peaks this is the heteronuclear correlation spectrum and this actually is extremely helpful to figure out which ones are close by.

(Refer Slide Time: 09:34)



And that is various experimental sequences are present here and in this case you see these carbon chemical shifts are quite distinct carbon chemical shifts are C2 C6 C4 C5 C6 these are very distinct they are quite wide range of chemical shifts are present. Therefore you can apply very selective pulses we are not going to the details of these pulse sequences but these establish correlations of the type here.

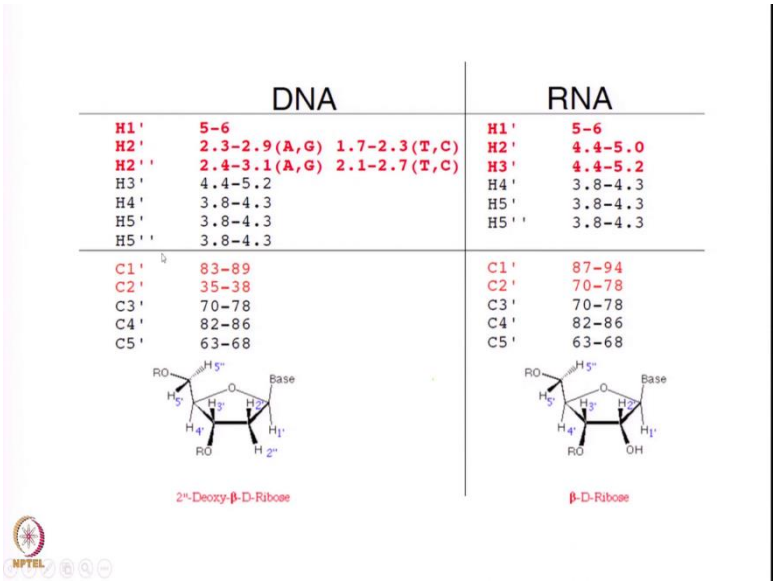
(Refer Slide Time: 09:58)



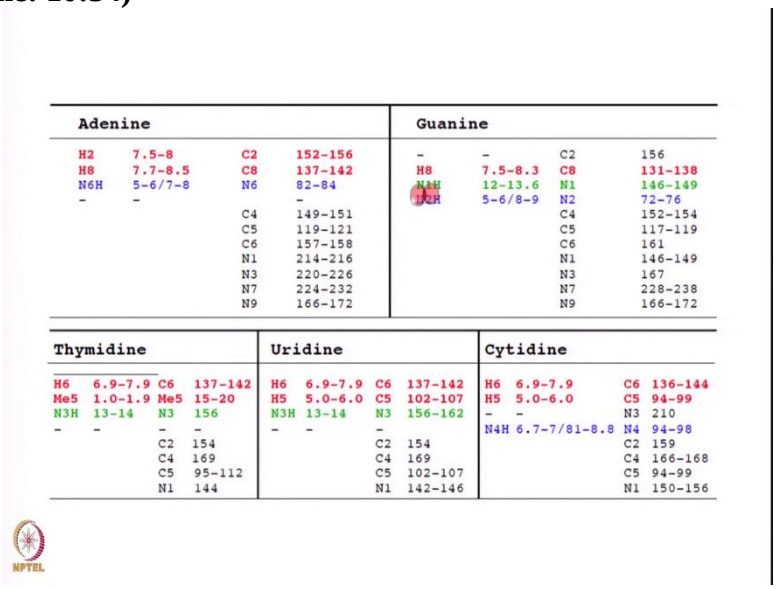
So, here for example for the you will establish correlation from this particular proton to this carbon from this carbon to this carbon and this carbon to this carbon you will establish correlations there. And these ones appear at very distinct chemical shifts and similarly from here you can establish correlation from this proton to this carbon this carbon you can go to this carbon and so on so forth.

You will also establish correlation from this proton to this carbon. So, therefore such kind of correlation experiments are present.

(Refer Slide Time: 10:30)



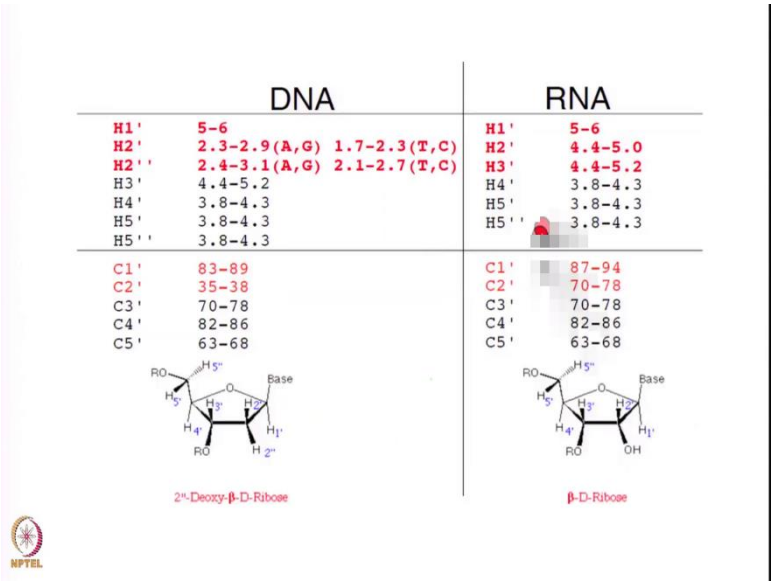
(Refer Slide Time: 10:34)



And here we show you the chemical shift ranges the chemical shift ranges for the various bases where all do they occur this is for adenine, guanine, thymidine, uridine and cytidine. For all of these you have the proton chemical shifts anyway which is all known from the proton spectra this we have already seen and these are the carbon chemical shifts the carbon chemical shifts appear very distinctly in certain region here.

And the C4 C5 C6 you see they are very distinct chemical shift ranges therefore this will allow you to apply selectively pulses to these different carbon. And therefore you treat them as individual channels and the magnetization transfer can be affected from one carbon to another carbon to another carbon and so on so forth. So, this similarly you have this information for the guanines and the cytidine.

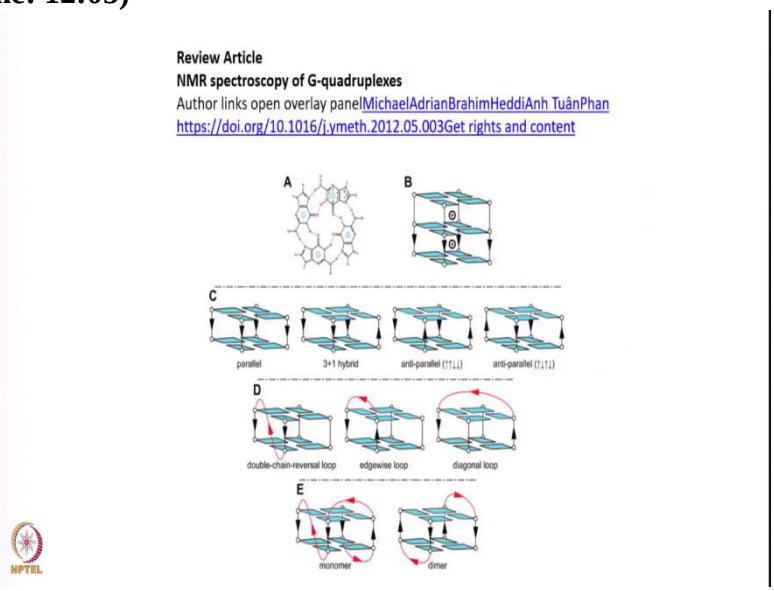
(Refer Slide Time: 11:18)



And the; comparison between the DNA and the RNA chemical shifts that is indicated here. So, you have the all the proton ranges which we already discussed earlier these are the chemical shifts for the protons and then this distant difference between the DNA and the RNA is the H2 prime chemical shift H2 prime is shifted to higher lower field in RNA because of the oxygen present at the 2 prime position.

Whereas in the DNA these ones are at 1.7 to 2.3 and this is a very unique area which will allow you to obtain the coupling constants from the structure and that is what is shown here. So, the 2' Deoxy-β-D-ribose okay this we have already seen before. So, we will not go into that now we will switch to another kind of structure. So, far we are talking about duplex structures.

(Refer Slide Time: 12:05)



Now as I discussed earlier the DNA has to fold itself into multiple structures and there are various ways it can fold and various ways of hydrogen bondings are possible one of the unique structures in this is the. So, called G quadruplex we also talked about this in the hydrogen bonding patterns here this is the G quadruplex. This G quadruplex means the G tetrad. So, G tetrad there are 4 G's with the hydrogen bonding pattern which is like this.

Very symmetrical hydrogen bonding pattern it is present here, and that is indicated in the schematic here. So, you have 4 channels 4 strands which are running in a parallel way then you

have these 4 G's which are hydrogen bonded in this manner very symmetrically they are placed in this manner you have the hydrogen bond this is one particular structure but you can have many other structures the many other ways this base pairing can be obtained the strands need not always run parallel.

So, here the strands are running parallel but here you see there is a 3 + 1 hybrid. So, 3 strands are going in one direction and one strand is going in the other direction. And here you have anti-parallel 2 strands in one direction this and this are one direction this and this are in another direction okay and all of the even so, this for the G tetrad can be formed. But what will then happen is that depending upon the orientation the glycosidic torsion angle will adjust itself.

So, with that sometimes it can go from anti to syn to bring this kind gone in resistor to be able to form this sort of a hydrogen bonding scheme. Now in this you have the again an antiparallel structure but now you see it is looking this way. So, here this like this this one like this and this one is like this and this one is like this. So, all the 4 you can see the way the hydrogen bonding parallel is, is done.

Different ways you can form the antiparallel structures. And in this lower one here now you can have a loop. So, double chain reversal loop chain runs like this there are only 2 G tetrads to chain the lines like this then it actually takes a long loop and comes back and goes in a parallel way here. What is present in this loop area they are not G's they may be some other nucleotides like the T's or the C's and things like that.

When they are present the chain can run like this double chain then there is a loop and these are put it in place there and then it can form a tetrad like this now. Similarly this can form within instead of going like this now if these are going parallel this is the turn it has to take. But if they are anti-parallel if these 2 chains are anti-parallel this can simply go in a loop like this and form a strand like this and form a pair here.

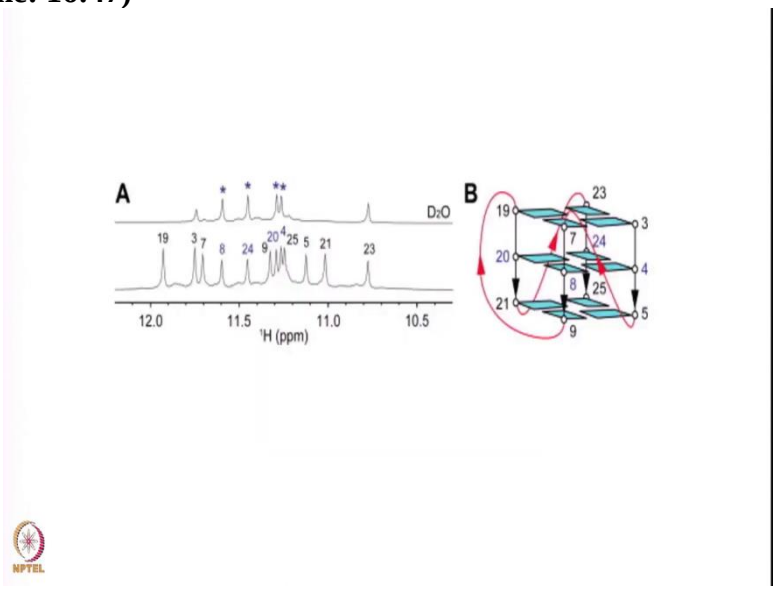
So, similarly you can have another one here this across the diagonal so to say. So, it goes from here to here forms then it folds the diagonal loop. So, this is called the edge wise loop this is the double chain reversal loop and then in this situation you have one molecule which folds in different ways to form a structure. So, in this case you see the chain starts from here it starts from this point it goes down.

Then the loops then comes here then it goes in the opposite direction then loops comes to this position then in goes direction then they get kind of a chain loop and then you have this one for guns come in resistor to form a structure of this type. Now here it is 2 simple loops like this. So, the chain starts here and then the loop here goes in this direction and then there are 2 molecules here there are 2 the connection between these 2 is not shown.

So, then you have another one going here and going like this and these ones are held together in this parallel fashion. So, you see there is enormity there is so, much variation possible in the structures and this is why it is not surprising that the entire DNA can be packed inside a small nucleus forming different kinds of structures. Now it is of course as this at the same time it is also a challenge to figure out what sort of a looping is happened.

This has to be done from the proper NOE based experiments and the hydronuclear experiments to establish the nature of the hydrogen bonds which hydrogen bonds which nitrogen is involved in the hydrogen bond. So, all that has to be seen from this sort of different kinds of experiments.

(Refer Slide Time: 16:47)



So, here is one illustration there the illustration of course we will not go into the details of this what is the nucleotide numbering etc here. How these were obtained but this is showing look at this the chemical shift spread this is the record this experiment is recorded in D_2O this is recorded in H_2O and all of these peaks are present here even D_2O some of these are present these are these are the amino protons the iminoproton region.

The quadruplex the iminoprotons appear from the 10.5 ppm to 12 ppm these are all G-iminos you see how many how many are there these are 12 here the 4 4 4 into 3, 12 and all of these are distinct this is the distinct there is no symmetry here because of that you are seeing all the 4 G's are distinct and they are all present here you can count here 1 2 3 there are 12, 3 here 4 here 7 plus 5 12 all the 5 G's all the 12 G's are distinct.

These are all formed within the same molecule you have a particular sequence running like this yes goes like this then he has a loop of this type then it comes down here like this then you make another loop like this and then of course it goes turns here then another loop here and ends up here in this. So, it is amazing, amazing the way that the chain is folding here and because of that you will have different orientations of the glycosidic torsion angle the sugar geometries will change.

And they clearly because of that then on all of the become non-equivalent and you see the distinct chemical shifts for the 12 G's which are present in the G-quadraplex once you have this then you will also have the NOEs between these various imino protons to the sugar rings and then from then you will figure out what are the sugar geometries and the glycosidic torsion angles see what are the individual glycosidic torsion angles.

(Refer Slide Time: 18:38)

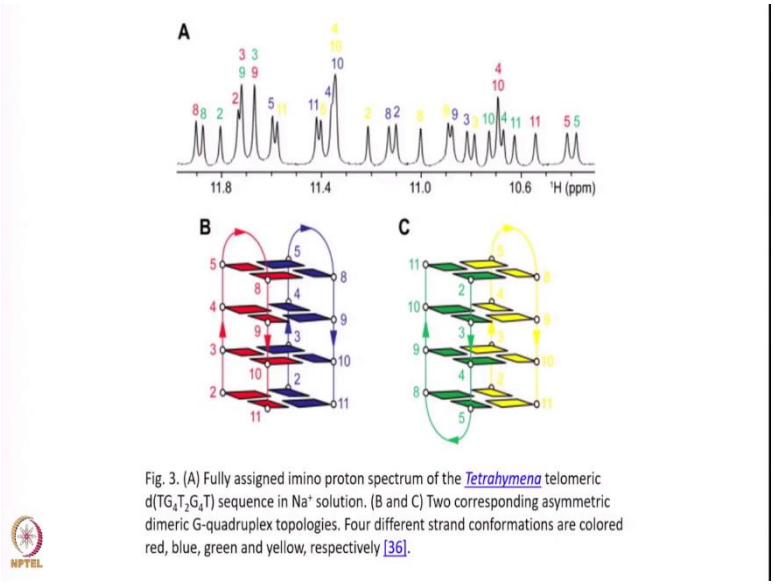


Fig. 3. (A) Fully assigned imino proton spectrum of the *Tetrahymena* telomeric $d(TG_4T_2G_4T)$ sequence in Na^+ solution. (B and C) Two corresponding asymmetric dimeric G-quadruplex topologies. Four different strand conformations are colored red, blue, green and yellow, respectively [36].

Now this is another example there how these various the nucleotides are formed is one particular sequence which goes like this other one which is like this then you form a quadraplex here and in this case this is a different kind of a structure you have this one go this starts here goes like this and comes down and another one which goes like this and comes out these are 2 parallel loops.

These are going in the parallel they are on the same side these 2 loops are on the same side of the quadruplex of the straw of the structure here the 2 loops are on the opposite side of the structure both the structures are possible both the things may coexist. So, when you have this you will have peak coming from both of them there is a certain symmetry here and those peaks this and you will you will establish the connections from the base-base NOEs.

Base-base NOE is here and the sequence that is given here is this TG4T2 G4T these are the 2 4G's which are forming the quadraplexs here. So, these are 2 3 4 5 then you have 2 3 4 5 8 9 10 11 because you see then you have the T2 here the T2 is in the loop and in the result T2 is in the loop G4 again these 4 nucleotides here these are 8 9 10 11. And the same sequence can form a structure of quadruplex in this case this is a symmetric this is symmetric and you have both the structures possible.

And you will see separate peaks for each of these structures and therefore you see so, many peaks beautiful spectrum. You can see all of these are very well resolved here all the imino protons the red and the things you can see all the different strand peaks can be identified they are identified with the different colours as I indicated in this. So, therefore this can actually become as complex as protein spectra.

So, see the small molecules even a small molecule like this is presenting such a wonderful dispersion and so, many peaks and it is of course quite a challenge to establish all of these connectivities.

(Refer Slide Time: 20:38)

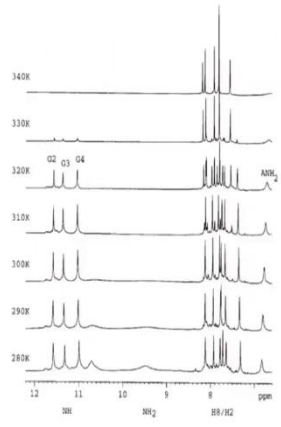


Figure 1. Imino, amino and base proton spectra of d-AG$_3$T in H_2O solutions containing 5 mM K^+ ions at pH 7.3 as a function of temperature. The G-imino and A-amino proton assignments have been marked in one of the spectra.

Now this is the simple example of a small molecule a small molecule of the AG3T. This is only a 5-mer in potassium solution. Incidentally notice that this will not form in the in sodium the quadruplexes are formed in the potassium solution. Because it so, happens that the space between in the G tetrads the potassium can fit in very well and that will stabilize the quadraplex structures the sodium will not be able to do this.

Therefore typically you do not find this in the sodium solutions. We will find this in the potassium solutions there is not quite an interesting feature. And in this case look at the stability of this quadruplex there are G3 there are 2 3 G tetrads there and this is the amino protons of the G's and these are of course the aminos here and the base protons H8 H2 and the iminoprotons are stable all the way up to 330 degree.

These brought out here melting if you see this is extremely stable even your normal duplex DNA of the 12 DNA will not be stable we will not be so much stable typically you will have the melting temperatures of 45 degrees 50 degrees and things like that. So, this is quite stable structure and these and by looking at this of course you can identify which nucleotide is which and incidentally in this case we also discovered something unique.

(Refer Slide Time: 22:02)

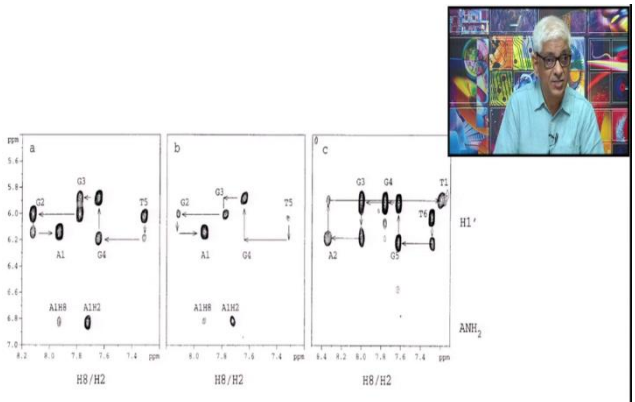


Figure 2. Selected spectral regions from NOESY spectra at 7°C in H_2O, pH 7.3 of (a) d-AG$_3$T (180 ms mixing time), (b) d-AG$_3$T (60 ms mixing time) and (c) d-TAG$_3$T (250 ms mixing time). In (a) and (b), cross peaks from A1NH$_2$ to A1H8 and A1H2 protons are also seen. In d-TAG$_3$T the corresponding region is empty indicating the absence of the A-tetrad type features, in fact, the A-amino resonances were not seen in this molecule.

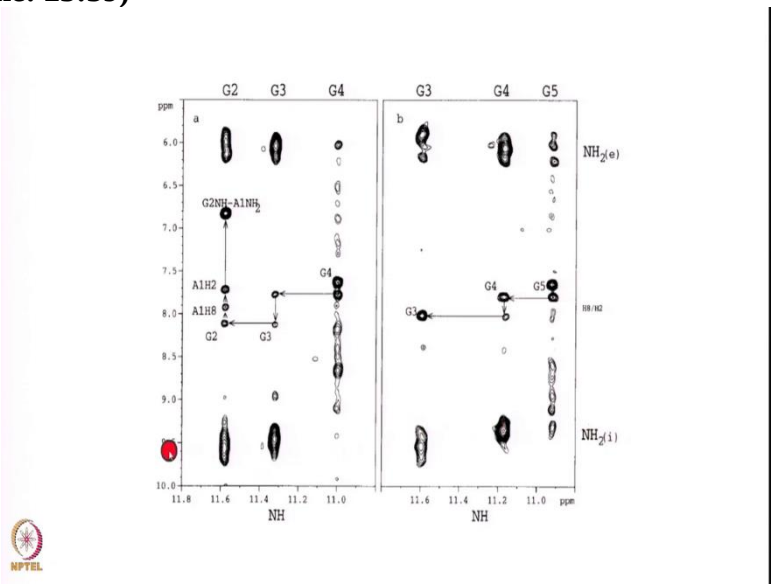
And that is this particular slide we showed you see here a cross peak these are the sequential connectivities as we see in the normal case this is the parallel stranded 4 stranded structure the G qudruplexes you have the identification of the sequential connectivities here T5 to G4 G4 to G3 G3 to G2 G2 to A1. And incidentally here you see a cross peak to the A1 H2 and this cross peak is seen to what this is seen to the amino protons.

These are to the amino protons this is ANH 2 amino protons are normally not seen and they are typically in the in this area chemical shift area amino protons are not seen because they exchange with water unless they get involved in the hydrogen bond. But now you see here we see an interestingly in this particular sequence AG3T you see this amino protons and what are these? These are the 2 different mixing times this is 180 millisecond mixing time of the NOESY.

This is 60 millisecond time NOESY and you see these are present why did we have to do this at 2 mixing times because there should not be spin diffusion. Somebody may argue that this is not a direct interaction but there is a spin diffusion but when you go to small mixing time there is no spin diffusion direct and you see amino proton noise to the a A1 H2 peak normally you do not see this.

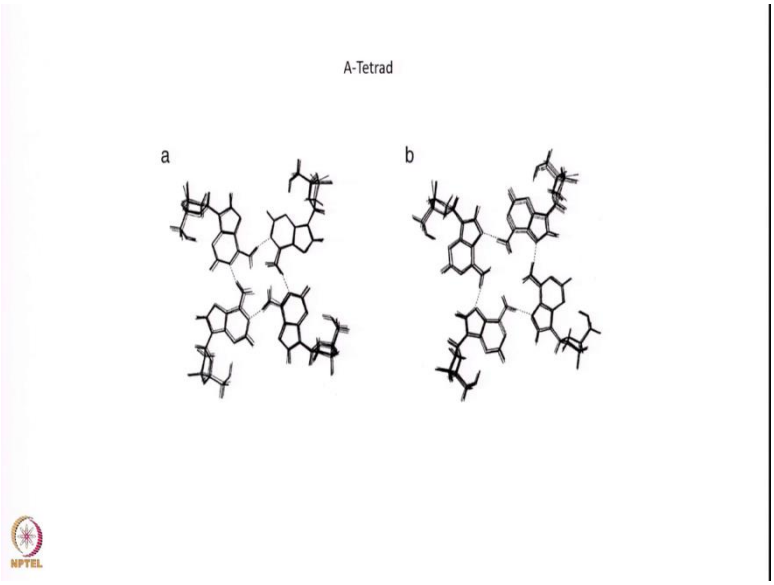
For example if you took this molecule this is TA G3T, TAG3T you do not find that and what does this tell you? This actually is telling us that the A is also forming quadruplex A is also forming A tetrad.

(Refer Slide Time: 23:39)



And that was an interesting observation here and of course this was confirmed by various other amino NOEs here from the amino protons you are seeing NOEs to the various iminoprotons, iminoprotons to amino protons you see NOEs and also to these amino protons. Otherwise you normally do not see the amino protons but here you are seeing the amino proton peaks as well because these ones got stabilized in the formation of a tetrad and that is shown here.

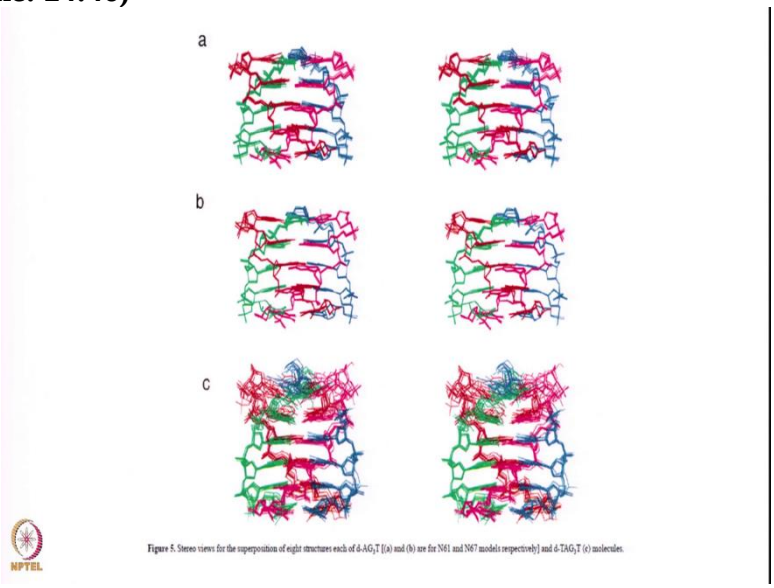
(Refer Slide Time: 24:04)



Therefore this was A tetrad we called this as an A tetrad you can see the hydrogen bonding scheme here found here see, see this is the hydrogen bond. Now what happens is this amino group is close to this proton here and this is the H8 proton that was the peak which we were seeing H8 proton. And this amino is also close to this proton this proton is H2 proton and we were seeing this amino to the H8 amino to the H2 we are seeing this.

And now there are 2 ways one can form this hydrogen bonds and these are the 2 different ways we have shown here and that indicated the structure.

(Refer Slide Time: 24:46)



One could calculate the structure now on the top here you have this A tetrad you not only have the G tetrads here there are 3 G tetrads there and the top you have the A tetrad as well and that was the previous one which I showed you. In this one this is the A tetrad. A tetrad is formed in this manner the hydrogen bonding scheme is slightly different in these 2 cases and these are called as I will show you what these ones are. These are 3 2 different kinds of structures okay.

(Refer Slide Time: 25:19)

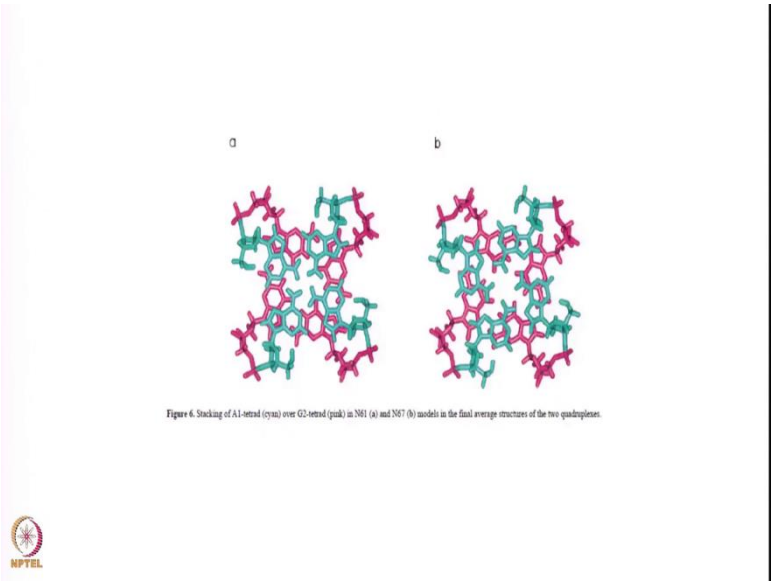
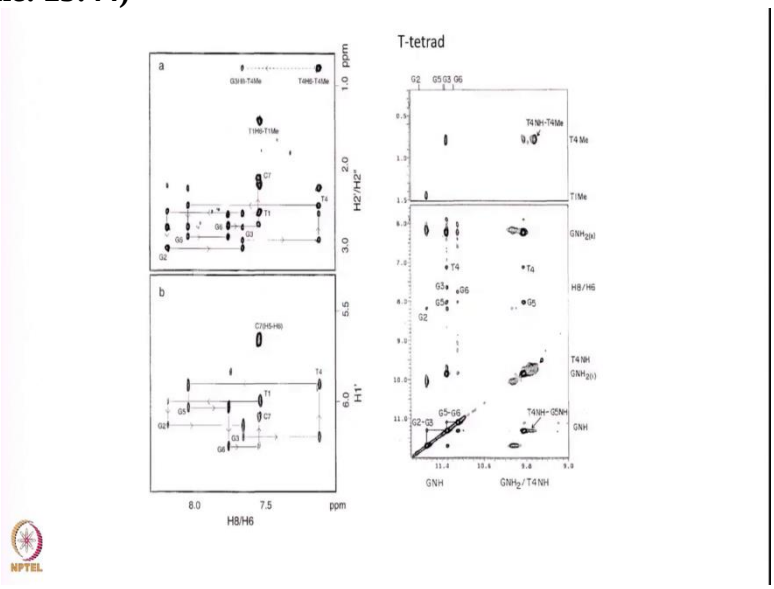


Figure 6. Stacking of A1-tetrad (cyan) over G2-tetrad (pink) in N61 (a) and N67 (b) models in the final average structures of the two quadruplexes.

Stacking of the a1 this shows the stacking of the A tetrad over the G tetrad this provides the stability to the A tetrad. So, the one which is in cyan colour is the a tetrad and what is below is the G tetrad and so look at this how this stacking is happening this stacking of the bases the purine bases provides enormous stability to the structure of the A tetrad otherwise you normally do not see this.

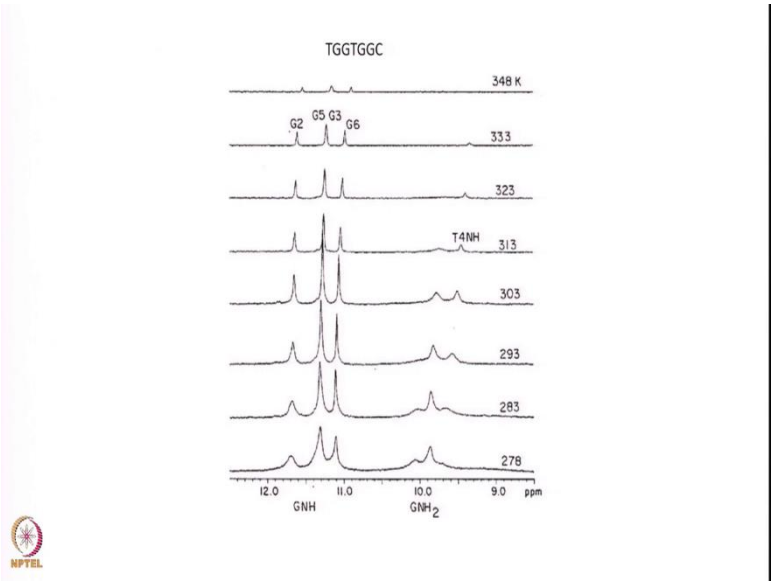
(Refer Slide Time: 25:44)



Now this is another discovery there this is called the T tetrad. So, once you have the G tetrad can you also have a T tetrad indeed it turns out that you can also form a T-tetrad at and that is evident from the T4 NH to the T-methyl. So, this NOE is very characteristic of that one and we will not go into the details of this discussion there with regard to the analysis this is typically the standard analysis of the one prime region of the base to the one prime area.

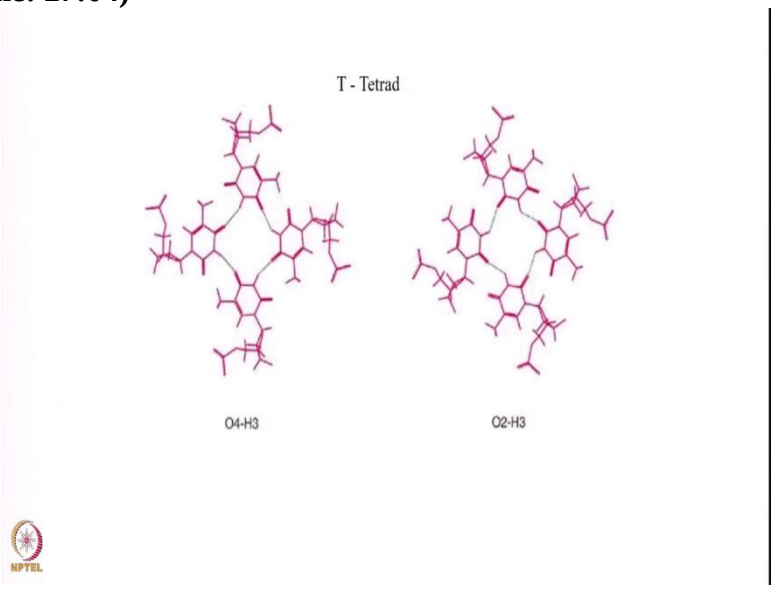
And then to the base to the 2 prime area to double prime area base to the methyl area and this one is the amino protons the amino protons to iminoprotons to the various NOEs iminoprotons to the various G is there and then iminoprotons to the T4 NH. Now this actually clinched the issue that there can be T tetrad as well.

(Refer Slide Time: 26:37)



The T tetrad is also formed and that is shown in this particular way the T4 NH to this one you see is seen even at a high temperature like 313. These are the typical from the G tetrad these 4 G's which are present here this form are shown here. And one of these T is the central T is also involved in the tetrad formation and that is what you are seeing in this area there. This physical peak is shown.

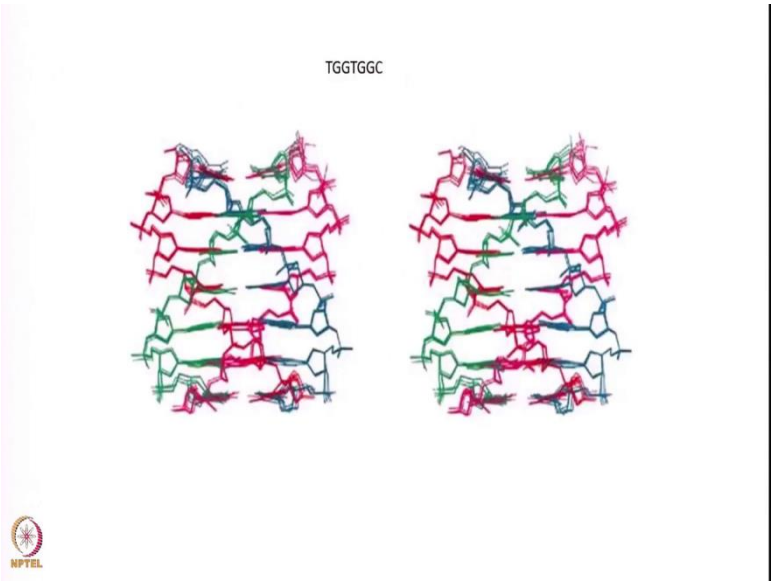
(Refer Slide Time: 27:04)



Now T tetrad there are 2 different ways the hydrogen bondings can be formed which oxygen is involved in the hydrogen bond. Now this is the oxygen and this is the NH this is the TNH and this is hydrogen bonding like this to the oxygen this uses the O4 this is the O4 in this case this is the O2 because 2 carbonyl groups in the thymine right at position 2 and at position 4. So, you can use either one of those to pair with the T3 NH.

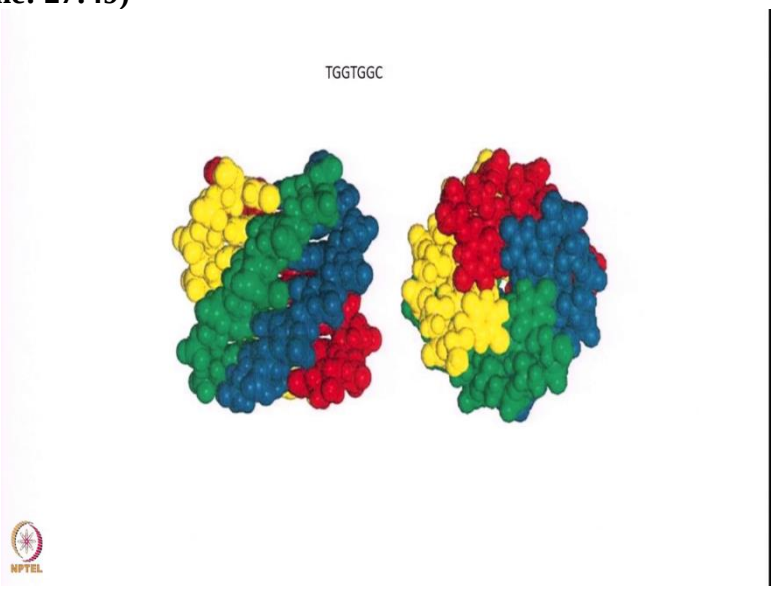
So, once we have that then you have this kind of a 2 types of hydrogen bonding schemes and you can have different structures depending upon that.

(Refer Slide Time: 27:42)



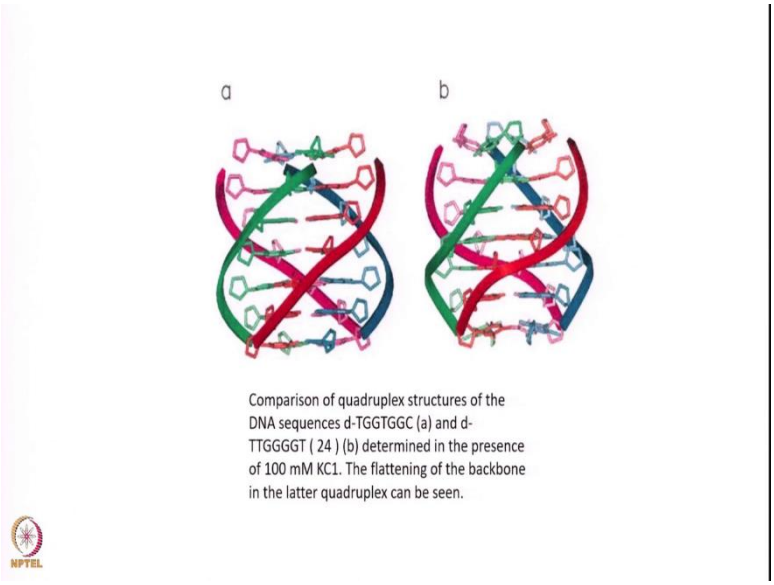
You see you can see different kinds of structures that can be formed depending on this is the stereo picture of the one particular structure.

(Refer Slide Time: 27:49)



And you can have this is the space filling model of the same structure.

(Refer Slide Time: 27:54)



Comparison of quadruplex structures of the DNA sequences d-TGGTGGC (a) and d-TTGGGGT (24) (b) determined in the presence of 100 mM KC1. The flattening of the backbone in the latter quadruplex can be seen.

And here you have the comparison of a this kind of a tetrad G there are 4 G is continuously opening here and in this case the 4 G's are interspersed by a T in this case. Now what is the result of that you can see there is a kind of a flattening of the backbone here in this structure and that is this one and in this case this is more straight the backbone is going in the straight and the tetrad is formed in the middle it is the G tetrad T G G T G G C you also have the T tetrad in the middle.

So, this sort of structures are formed and then I think we have to investigate this various kinds of DNA structures depending upon the sequence what we have the variety of structures is quite large. You have we have talked about the quadraplex there can also be triplexes. The triplex is we talked about the triplexes earlier in the previous lecture when you talk about the structures and there are also characteristic energies for the triplex structures as well.

Depending upon the base pairing scheme whether it is TAT or AAT triples you will have different kinds of NOEs. So, with that you will be able to calculate these triplex structures as well. So, therefore this establishes the enormity of the DNA structures which certainly enables the packing of the DNA inside the nucleus. So, I think we can stop there and that completes the nucleic acid structure calculations and the varieties of nucleic acid structures.