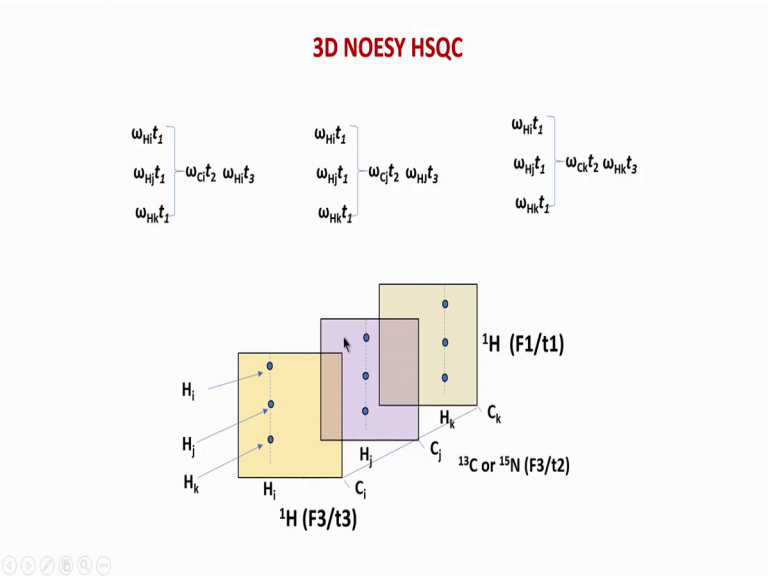
**Multidimensional NMR Spectroscopy for Structural Studies of Biomolecules**
**Prof. Hanudatta S. Atreya**
**Department of Chemistry**
**Indian Institute of Science, Bangalore**

**Lecture – 35**
**Determination of Protein Tertiary Structure from NMR Data-Part II**

So, we are now in this part of the course we are looking at 3D NOESY HSQC and we started looking at how the correlations are obtained in the last class. So, you can see here this is typically how the spectrum is shown here we can see these are the different correlations that we obtained in the NOESY.

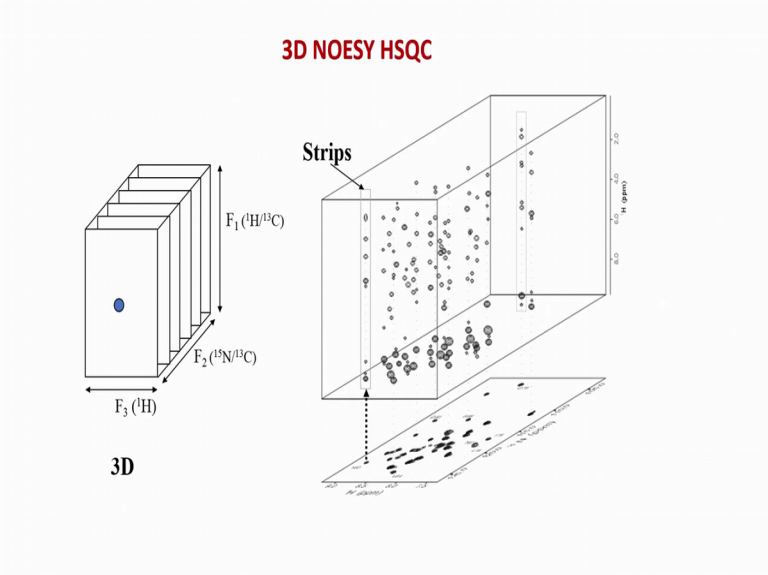
(Refer Slide Time: 00:39)




So, this is along the t 1 dimension which is basically the vertical dimension shown here t 1 then you have carbon along t 2 that is this horizontal going into the plane of the board and t 3 is the proton. So that means, for every given carbon proton pair you will see a set of correlations along the F1 or t 1 axis ok. So, in this particular example was what we saw it is basically in the same molecule. So, all atoms are correlated to all the atoms. So i, j, k here is correlated to i. Similarly i, j, k is again correlated to j and i, j, k are also correlated to Ck and Hk. So, you can see if you look at this 3 correlations all the 3 so, you see all the 3 peaks here for H i, C i, pair you see all the 3 correlations for H j, C j pair and all the correlation or Hk, Ck.

So, now if we look at this F3, F1, F3, F2 plain sorry this should be a correction here this is F2 so, if you look at the F2 F3 combination that is H i C i, H j C j and H k C k they represent nothing but the HSQC. So, HSQC is the F2 F3 plane and that forms the basal plane that is the basal projection and on the third dimension we see the correlation.

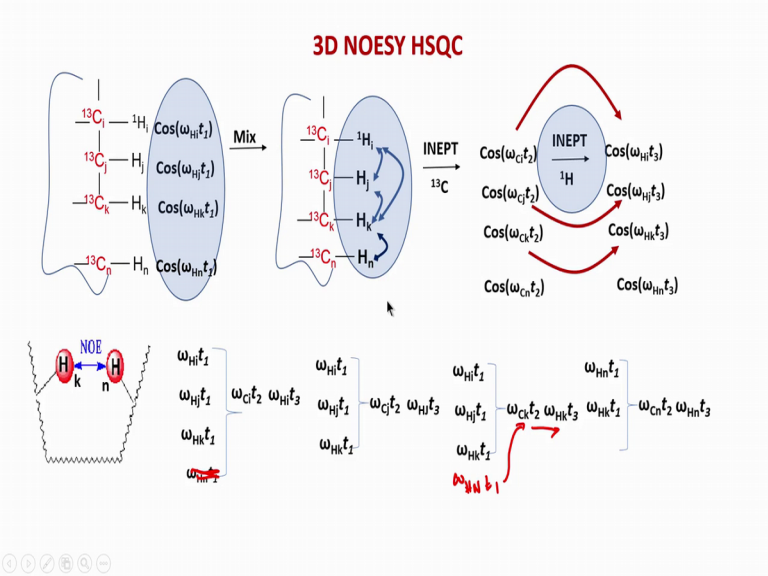
(Refer Slide Time: 02:20)




So, this is what we saw here this is what this is how the NOESY HSQC looks like. So, you have a 3D spectrum the base is HSQC and this HSQC can be either proton nitrogen or it can be carbon proton HSQC and for every peak in the HSQC if you go in the third dimension you will see a set of peaks or correlations and these peaks are the one which could come from the same amino acid or it could come from far away amino acid or it could come from a neighboring amino acid.

So, which is which peak we do not know unless we know the assignment means, unless we know the chemical shift of each and every peak to what atom it belongs in the protein, we will not be able to identify which peak is what type of correlation is it intra amino acid means is it within the amino acid or is it without outside the amino acid, those kinds of information can be obtained only if we have the full complete assignment of every peak with us. That is why I emphasized earlier also now I re repeat that resonance assignment in proteins is very important step without which we cannot do much in NMR. So, let us move further now.

(Refer Slide Time: 03:44)




So, what I showed you earlier was an example of within the amino acid correlations but NOESY is not about that. In NOESY we want to see correlations from a far away proton. So, for example, you consider this case; so the same molecule which we saw earlier, but now imagine that this k is at proton k is coming close in distance to another proton far away n far away meaning in the sequence not in space, in space it is close, but in sequence it is far.

So, this is what we want to capture in a NOESY experiment we want to get the long range interaction, long range meaning the distance is not long, but they are not directly connected they are very far away in terms of the amino acid sequence. So, you can see these two atoms are very close in space. So, let us see how the correlations will come. So, now, I can show you so, what we can imagine that this amino acid is going like this there is a chain and it come back to this n ok.

So, this is just a schematic drawing to indicate that there is a long chain which is also shown here and it comes this atom n comes close to this k. So, now, when I excite the first pulse in the NOESY HSQC if you recall the pulse sequence the first pulse will excite all the protons including this far away does not matter because all are protons.

Now when we do the mixing this is where the important point comes when we do the NOESY mixing, this mixing will now take place; that means, they will transfer magnetization or polarization to each other, but they will transfer only to those which are

close by. So, here we saw that H k and H j and H k are close so they transfer to each other, but now let us say that this far away atom H n is close or near only the H k ok. So, imagine that it is near only to H k and somehow it is not coming close to H j and H i. So, in such a scenario it will transfer polarization only to k ok. So, there will not be any transfer of polarization from n to j and i ok.

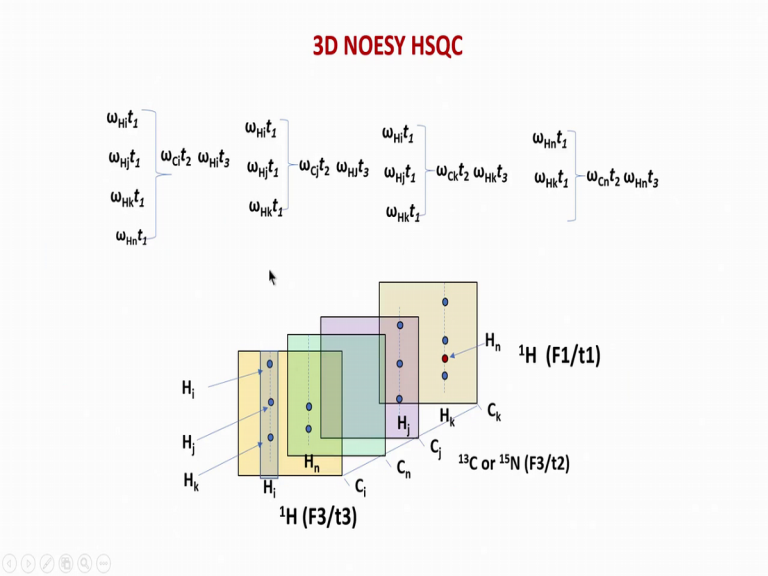
So, they may look near in this picture, but imagine that they are not close so; that means, only n to k transfer or k to n again remember it is both vice each equal and opposite always or vice a mutual so both transfer to each other. So, now, when we do this inept like earlier we draw saw there now will also get the chemical shift of Cn because all the carbons will be excited and then they will correlate to their proton directly attached proton.

So, this top part this part was same as earlier now I have added an extra because now we have an atom which is far away. So, like this it could be many far away atoms I am just giving an example. So, therefore, in our correlations which we saw now we can add one more set of correlations you see here. So, here its coming as H n; H n means not the amide H n this H n which is along t 1 excited because this is coming from here then it gets correlated to C n and H n, this C n and H n.

So, this combination is called the diagonal peak we saw this in the last class and now we are getting an extra correlation which is between k there is H k and H n. Why is that coming because H k transferred the magnetization to H n and H n was labeled with Cn and H n. So, H k was in the t 1 axis here this H k got transferred during mixing to this and during t 2 and t 3 got frequency labeled with the carbon and proton. So, this is extra correlation and this is called a cross peak.

Now this H k will also H n will also come here because as I told you this is a mutual equal and opposite this is if H n, H n transfer to H k H k n transfers to H n and so, in this case H n correlation is also noticed in this part of the spectrum where we are detecting H i sorry it should be coming here. So, let me correct this here it will be here because it gets correlated to this H n is correlated to H k C k and H n H k is correlated to H n C n mutual, equal and opposite. So, these are the this is a new correlation which you are going to see now because of this far away atom coming close to H k. So, let us see in terms of a spectrum what happens.

(Refer Slide Time: 08:24)




So, these are the correlations again I have to I will make this correction this goes from here, it comes here ok. So, now, if you look at the 3D planes, now we have added one more plane if you notice here blue color and that is C n because now we have an extra C n H n combination. So, this C n H n combination is shown here ok.

Now, in that plane along the H n that is along H n along t 3 you will see the long t 1 H n and H k along t 1 or F 1. So, along t 1 or F1 along here we see these two peaks. So, one of them is H k and one of them is H n. So, one which is the H n here will be called as a diagonal peak because, it will be repeating a H n here plus H n along this axis, H n along this axis so it will become a double diagonal peak I mean it is repeated two times it is a diagonal peak.

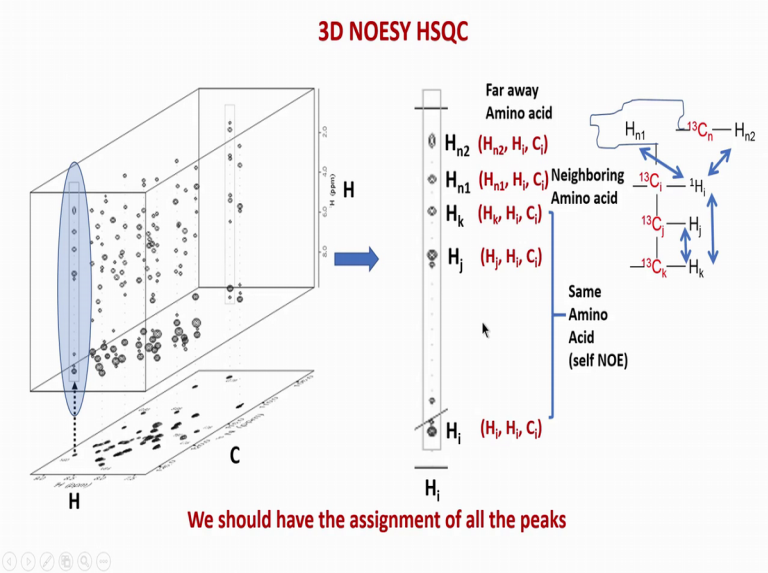
Now if you look at this plane here that is H k C k combinations am looking at here H k C k H k plane in the C k H k plane along the H k that is along t 1 you will see now am observing an extra peak H n coming from this combination H n this chemical shift combination. So, this is a new chemical shift correlation which am now getting because of the NOE effect which was earlier not present because we were only considering within the amino acid. So, this is very important correlation for us because it is giving us a distance information remember every peak in a NOESY corresponds to a distance.

So, every peak uses information about the proximity. So, therefore, this peak shown here in red color although the color wise in a real spectrum they will all be the same, but just

to highlight I have shown it in a different color. So, please do not think that the color of a peak changes for a NOESY peak it is all the same.

So, the H n correlation now to H k is giving us a distance information between these two atoms H n and H k. So, that is basically the essential or unique or a very beautiful idea about NOESY that it gives me correlation chemical shift correlation between far away protons I mean far away in the sense far away in the sequence of amino acids, but near in this space.

(Refer Slide Time: 10:57)




So, this is what is shown here. So, now, let us look at more carefully in this strip plot manner. So, this is the 3D NOESY picture which I told you earlier. So, now, suppose I take a strip out of this here so, I take consider this strip which is shown in a magnified zoom manner, now here am seeing here H i C i correlations right. So, this H C combination, but for any given i some proton i and C i plane now am drawing a strip along this axis.

Now here you can see many different correlations or peaks. So, there could these many of them could be the same proton means from same amino acids sorry. So, they could come from the same amino acid i so, H i H j H k are coming from the amino acid same amino acid and then we have some extra peaks here and these extra peaks could be a long range or a NOESY interaction. This is also NOESY all are NOESY peaks here, but these three are coming from this could be from the same amino acid and these could be

from neighboring amino acid or could be a far away amino acids. So, here we can see what is shown in the bracket is the three chemical shift correlation.

So, here H k is correlated with H i which is here and C i because we are at the C i plane you are cutting at the C i plane here so, we are looking at the H i C i plane. So, H i C i is common to all because we are all in H i C i plane, but along this axis it is changing because of different chemical shifts along this. So, now, let us see what could be these two peaks. So, let us say this is some I mean a proton nitrous arbitrary I call it H 1 H n1 the H n1 and H i, C i combination and this could be H n2 some other proton H n2 H i C i here.

So, as I mentioned in a structural in a NOESY we have proteins have three dimensional structure so; that means, suppose this is one amino acid here and it is going like this and then coming back and then there is another amino. So, it is possible that this H i is close to these protons ok. So, if they are close to these two protons H i will show a correlation to these protons in a NOESY spectrum, but in a j couple or TOCSY spectrum we will not see that because they are far away far away in terms of the sequence. So, now, here these three correlation as I said could be from the same amino acid and when the correlations between two protons if they belong to the same amino acid we use the word self NOE self means the same amino acid.

Now this H 1 n1 H n1 let us say imagine suppose this is the next amino acid means neighboring amino acid. So, we can say this is a neighboring amino acid and there is a interaction because they are close by in space. So, this could be this is called as a sequential NOE sequential means the previous or the neighboring, we are also use the word sequential earlier in assignment.

Now, this peak which you are seeing here could be a far away amino acid and it is again showing an interaction with H i because it is close in space, remember the word close makes a big difference it has to be proximal, but sequence wise it is very far so there is no question of any j coupling between the two. So, this correlation what you are seeing is coming because of distance through space interaction.

So, this is how we interpret the NOESY what we do is we cut out strips in a computer like this and for every H i C i and it could be any of these H i C i could be any two carbon proton of course, we should know what they are which amino acid it is. So, that is

why we need the assignments and if you go to any particular amino acid H i C i plane in that we start looking at all the correlations along this axis. And then we try to identify what each peak or each correlation which atom it corresponds to in the protein; that means, again I should know that this peak belongs to H n2 and H n2 is what which amino acid I should know that also and that again comes from assignment.

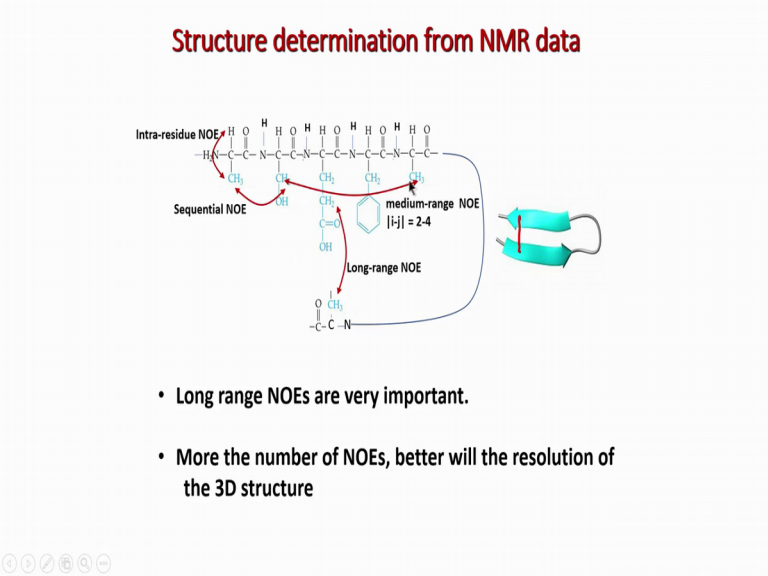
So, basically the bottom line is we should have assignments of all the peaks in the spectrum and that is why the assignment is given so much of importance in fact, there is a journal dedicated to NMR assignment. So, if you have a protein which is about 100 amino acid or more you could actually assign that protein including backbone and side chain and then publish it in a journal.

So, you see that is why the assignment on its own for any protein is publishable and therefore, it because it requires a lot of work and it is very important in the whole process of structure determination. So, now, once we have got these peaks what do we do next we have got the correlations in the NOESY and we have identified that this peak is this, this is this and so on. So, now, how do we convert this into distance because remember each peak as I mentioned encodes the distance between proton.

So, which distance in this case for example, let us take this peak here it will gives me distance between H j and H i which is am sitting on the Hi; so, between every two protons. So, look at this combination here between H j and H i am getting a distance. Now look at this here this is a H i to H i what is a H i to H i there is no concept of distance there is not 0 distance there is no distant at all.

So, we do not consider diagonal peak at all this is what I have been telling earlier also the diagonal peaks are of no use in a NMR experiment be it NOESY, be it TOCSY be, it cozy and so on because they do not show any new information they are only correlating the same atom to each to the same atom. But now here between H j and H i am getting a peak and the intensity of this that is a volume the area of this peak will give me information of distance between H j and H i. Similarly between for this peak I will get the distance information between k and i, n1 and i here and between n2 and i here. So, therefore, each peak is carrying a lot of valuable information for me in it gives me distance.

(Refer Slide Time: 17:42)




So, basically that is what we do we look at all possible correlations between proton and proton in a NOESY and then all those correlations that is where the peaks are then converted into distances. So, let us see now how do we classify the different correlations? So, number one is suppose I see a correlation in NOESY between the proton like this and a proton in the same amino acid. So, this is an amino acid called alanine, in that alanine am saying between the alpha proton and the beta proton.

So, this is as I said it is called self NOE or intra residue NOE this is not so important for us because the distances do not normally vary they are fixed distances. So, this is not such a important correlation for us, but it is always there in the spectrum. Now this in let us say that is am seeing a correlation or a NOE cross peak between this beta proton of this alanine and this beta proton of this serine. So, this is an amino acid called serine. So, this correlation is now a sequential NOE because it is between neighboring amino acid.

This is also slightly important although not very important for us because, if you take a protein be it alpha helix random coil or in a beta sheet confirmation the sequential means this neighboring distance normally is in a fixed range ok. So, about 3.5, 3 to 3.5 and you would not expect a big difference based on the structure. So, sequential NOE again are not so critical, but it is again present in the spectrum we have to identify them and assign them. A third category is medium range.

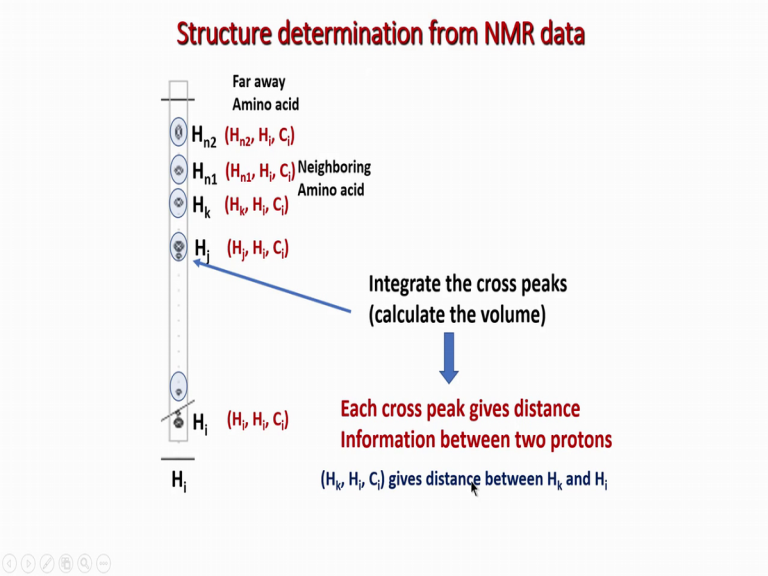
Here the distance between the amino acid from here to here is 2 to 4 amino acid separation. Remember one each NH here is 1 amino acid so I can draw a line here and that become 1 amino acid. So, you can see there is a separation about 2 to 3 means more than 2, 1 if it is 1 it is sequential so, but more than 2 and around up to 4. So, these are interesting and NOEs because they give us information on the secondary structure. For example, I will show you in the next few slides if you want to know it is a helix or a beta sheet it helps us to know that information we need to know the medium range NOE the medium range correlations in the NOESY and from there we can identify the secondary structure not the tertiary.

For tertiary structure what you need is, this you need a long range NOE you need a NOE between some proton like this and some other far away proton like this, this is you can see there is a long chain here. So, this is called a long range NOE for example, in a beta sheet you see this amino acid is going like this and taking a turn and coming back. So, for example, if I have 2 atoms here this distance will be released this will be less than 5 arm strong, but you see they are long range NOE because they have 1 amino acid 2, 3, 4 like this and this will be about 10 amino acid in separation. So, that called long range NOE.

So, the long range NOEs are very very important. They are the one actually which come tell us the distance because finally, we want to fold a protein; protein has to be folded in a structure and for folding you need to bring two far away amino acids to two or more far away amino acids close to each other and that will come information will get from the NOESY. So, the more the number of NOEs better will be the resolution of the structure means if I get more and more like this information correlations long range, medium range and sequential more information I get the better the structure will come out because we will have more information to build a model.

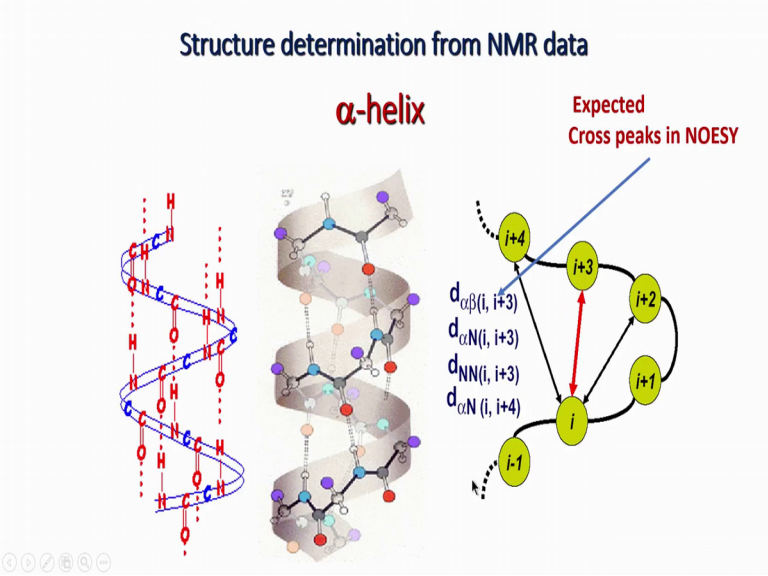
So, remember again in the last class I mentioned the NOE the NMR structure determination in an basically a model building approach we are building a model, a picture we do not have really the real structure with us, we do not know the real structure we just build a model which will fit all the distance information that we give ok. So, that is how it is done.

(Refer Slide Time: 22:01)




So, coming back to this NOESY correlation plot which I showed earlier you can actually convert this into distances. So, this is what is shown here we integrate means we calculate the area of or the volume of the peaks the volumes are calculated by integration; integration can be done in a software is very straightforward to integrate the cross peaks and basically which means calculate the volume. Now volume as I said each peak gives me the distance information. So, H k H i C i correlation peak for example, here this one gives me distance between H k and H i or correlation between ha H k and H i.

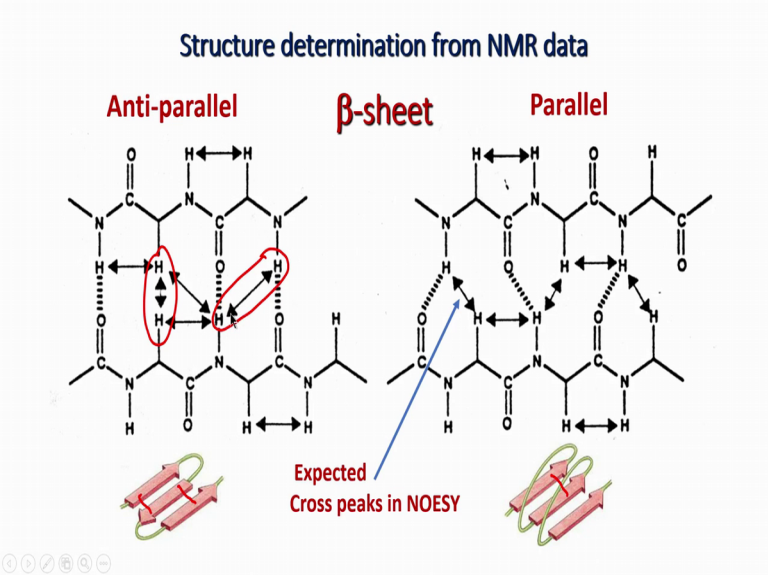
(Refer Slide Time: 22:43)



So, as I said the medium range. So, how do we get the information of a secondary structure from the medium range NOEs so we can see here for example, in alpha helix like this the amino acids will turn and come like this. So, there is expected to these distances become lesser than 5 arm strong. So, we can see from i to i plus 4, they could be the cross peak between the residue i and i place 4 amino acid because the distance between those two will be right. So, these are the different characteristic NOEs which are known to tell us whether it is alpha helix or not. For example, n means H n amide proton.

So, if I get a correlation between amide proton of residue i with amide proton of residue i plus 3 in a NOESY spectrum; that means, that particular set two residues are in alpha helical geometry or confirmation. Similarly if the alpha proton of ith residue gives me correlation to the amide proton of the i plus 3 residue then that correlation very is a signature of alpha helix. So, these are from 4 different correlations which are shown here are very important to identify alpha helical part in a protein remember again we are not looking at the full structure here 3D we are only looking at the 2D I mean the secondary structure.

(Refer Slide Time: 24:07)



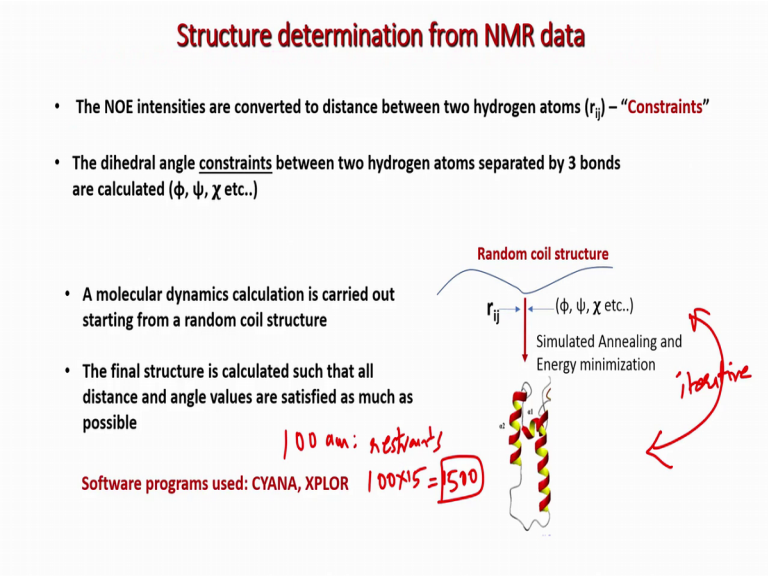

Now in beta sheet, it is little more long range, but because these two strands can come together so let me show you this here for example, in a parallel beta sheet, parallel means the sheets are parallel to each other they are pointing in the same direction.

So, when they come and fold like this these two are actually very far away these two I mean amino acid here and an amino acid here are actually very far away, but you can see that they are close in space because of this close hydrogen bond interaction. So, we can see the hydrogen bonds are shown by dotted line. So, these NOEs if you observe this NOEs in your spectrum immediately you can conclude that this is a parallel beta sheets structure in that protein, again this is a local it is a secondary structure remember we are not still quantum to the tertiary that is 3D structure.

Similarly this side if you look at it this is an anti parallel beta sheet and the sheets are anti parallel the strands are anti parallel. So, any 2 amino acid here for example, like this am drawing with a red line they could be far in the sequence, but they could come close in the space why because of this hydrogen bond interactions. So, you can see this proton to this proton for example, is a very characteristic of the beta sheet this is also very characteristic of beta sheet means if you see the H alpha proton of 1 amino acid giving a cross peak or a correlation to H alpha of another amino acid immediately you can conclude that, there is a beta strand in that region ok.

So, these are the type of signature which we get for beta sheet. So, now, coming to how do we get solve the structure final three dimensional structure of a protein? So, what we do is we take all this distance information remember we are getting the distance by from the intensity of the cross peak between the two atoms.

(Refer Slide Time: 26:07)



So, these distances all of these NOE this intensities are converted to distances between the 2 corresponding hydrogen atom we call it as r, r means distance ij; ij means between two atoms i and j. So, we use we call this as a constraint why because as I said we are doing a model building here, model building means we are trying to build a 3 dimensional model which is constrained by the distances which we are giving. So, our distances what we are giving is restricting the geometry therefore, we call it as a constraint.

Similarly we also saw that we can get j values j coupling and from j coupling we can get the phi value. So, those constraint mean those are values are also given because we are constraining restricting the geometry of the molecule. Then what is done next the next what is done is you start from a random coil structure so take be you can make this a computer you take a simple linear chain of an amino acid structure because you know the sequence of your protein.

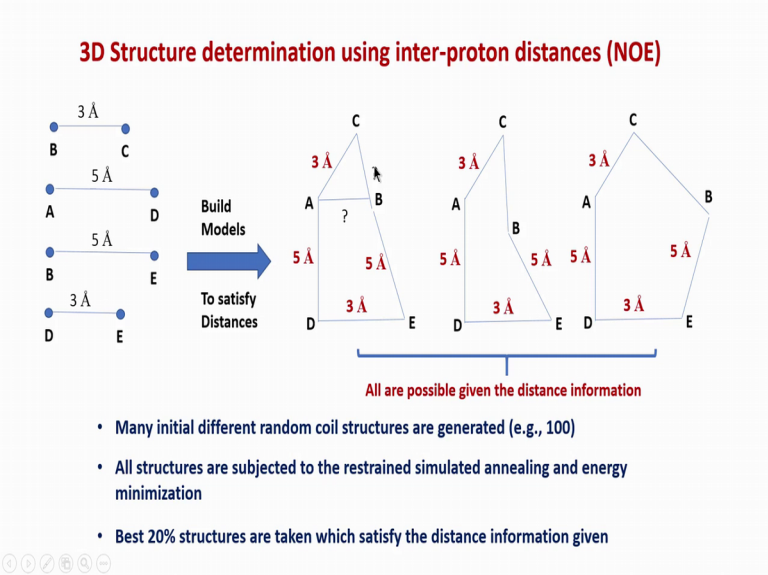
So, you build in a computer random coil structure which is done by software then, you start at small declare dynamics calculation. So, this is something we cannot go into detail, but this is a multi computational technique where in you start from a random coil and start giving the constraint; constraint meaning those numbers that mean distances between the each amino acid to each other amino acid.

So, all the distances which we get from NOESY correlations are given as a input to this program and also the different angles then it carries out what is called as a simulated annealing and minimizes the energy of the structure. And finally, it will give us a structure like this such that all distance and angle constraint or values that we have given are satisfied as much as possible. So, remember they were giving going to give thousands of constraints of typically for an 100 amino acid protein the number of restraints which we give is rough calculation is 100 into 15.

So, about 1500 constraints ok. So, we are not talking about a few distances here we are talking about 1500 distances and angles so obviously, it cannot satisfy all of them to make a structure there will be some distances which we give maybe wrong in value. So, such cases it cannot do anything and it has to not satisfy. So, those are the one which we have to remove and we start to we have to refine this structure we cannot keep on it is not one step process we get this structure then we find some distances are violated.

Violated means not satisfied then we go remove those distance or we adjust the cross peak maybe the assignment was wrong so on so forth and restart again and build a new model. So, like this we iteration; so, I would say this process is an iteration iterative it keeps happening till we reach a level where we say we are satisfied with the distances what and constraints that it has satisfied.

(Refer Slide Time: 29:32)

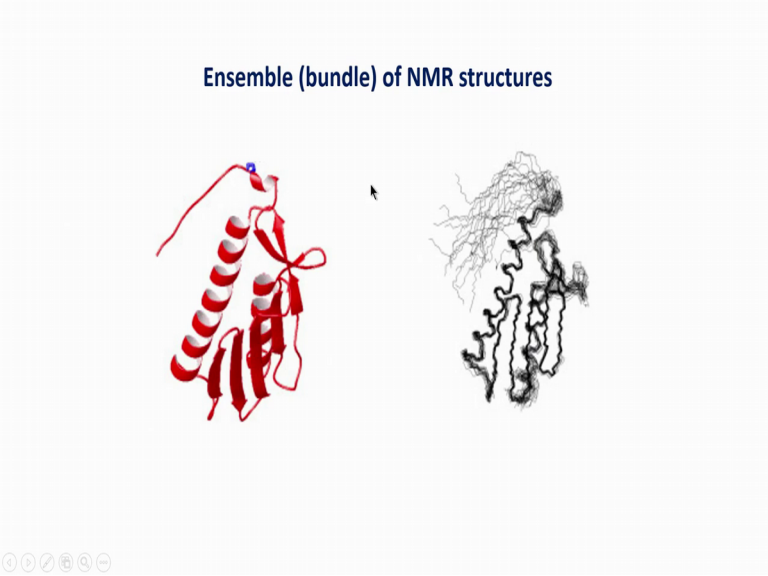



So, that satisfaction once you reach then we can say that our job is complete so, but you recollect if you recall in one of the slides I showed that you know if the distances are not sufficiently in number then, there are a lot of uncertainties means I do not know for example, in this case where should I keep the B; B could be here it could be here atom or this B atom because I do not have C B and A B distance in my list input list this is the input list.

So, therefore, when am building a model to satisfy distances and if I am missing distances I cannot build a correct model not correct model I cannot build a unique model. So, all these three are possible right because I do not have any distance between these two. So, I have to consider all the three as equally correct there is no distinction between these three therefore, in NMR what we are doing is we start from many different initial random coil structures for example, 100 of them then all of them are subjected to the same energy minimization and each time it will give me a different geometry or different structure because each time it does not know to fit this B.

So, sometime one structure may give me this result, another structure model may give me this result and other models. So, like that out of 100 I may get all different models. So, what I do is I will take the best 20 of the 100 which satisfied the resistance. So, I will take your best what do you mean by best? Best means where there is a minimum violation of distance. For example, if this is supposed to be 3 in some structure I may this may look come out to be 3.6 then that is a big violation if it is 3.1 its or 2.9, but if it is 3.5 or 4 and that distance was not correctly satisfied so, I have to reject that particular structure ok.

(Refer Slide Time: 31:22)




So, this is how structure building in NMR is done. So, this is how we as we show in a publication or in a database we show a bundle of NMR structures we do not show a single structure like this we show a ensemble we use the word ensemble why? Because all of these structures which we have got all there are about 20 here in this drawing in this picture all of them have satisfied the distance and angle constraints that we have we want to the software.

So, all of them are equally correct or good. So, we put all together and represent this. So, the average of this could look like this ok. So, this is you can see already they all look same as such there is no big difference, but I could take an average structure and show it as a representative model again we do not know the correct structure.

So, basically this brings us to the end of the NMR structure determination the last point here is we will come back to this in the next class and that you have to deposit this pair protein structures in the database when we publish and we look at this slightly the next class before we move on to the next part of the course that is protein ligand interaction.