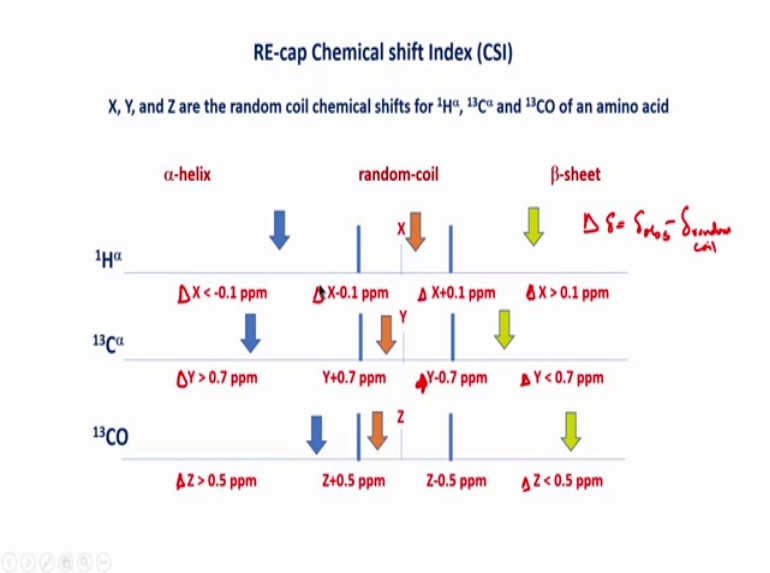
Multidimensional NMR Spectroscopy for Structural Studies of Biomolecules
Prof. Hanudatta S. Atreya
Department of Chemistry
Indian Institute of Science, Bangalore

Lecture – 33
Determination of Protein Tertiary Structure from NMR Data: Part I

In the last class we looked at how to determine secondary structures of proteins using the chemical shift assignment index method and from the measurement of J coupling between the HN amide proton and the H alpha. So, in this class we will now start with 3 D structure determination, but before we go into the 3 D structure determination, let us recap what method was used in CSI, because CSI is a very important step before we go to the 3 D structure. So, let us just recap again what we did yesterday with respect to the CSI method.

(Refer Slide Time: 00:59)



So, this is shown here. So, this is a CSI method that we used. So, here for example, in this slide you can see that X Y Z are basically just denoted as the random coil chemical shapes of H alpha C alpha and CO of an amino acid. So, consider for example, any amino acid, it can be any of the 20 amino acids and consider that the X Y and Z are simply the values for the random coil chemical shifts of these three atoms, so this is shown here. So, now, what do we do in the CSI method is the following.

So, for example, consider proton H alpha. So, for h alpha let us say that this is the X value, this is a random coil value for a given amino acid, then we look at this side that is negative less than point, less than 0.1 or greater than 0.1. So, this you can think of it as a boundary, wherein if the chemical shift of any amino acid whichever we are looking at. For example, X could be lysine or X could be arginine or a aspartic acid or any of those 20 amino acids. So, if you are looking at those 20 amino acids then the chemical shift boundary is defined with respect to the random coil chemical shift of that amino acid. So, let us say lysine has a value of 4.33 ppm then this will be 4.33 ppm. So, what we do is, we look at this boundary which is 0.1, 0.1 on this side, either side.

So, if the chemical shift lies in this zone then it is considered as a random coil. If the chemical shift lies beyond this line on the left side, that is it is less than minus 0.1, this means the difference. So, let me write this here, this is the actually the secondary chemical shift; delta. So, delta, basically the secondary chemical shift here, where delta is essentially given as delta delta is delta observed minus the delta random coil ok. So, if the delta delta; that is a delta observed minus delta random coil is less than minus 0.1 ppm, then it will lie in this zone. And if the delta delta min is greater than 1.1 ppm, like here delta delta then it is in the, this is the right hand side zone.

So, if the chemical shift of this amino acid observed value minus the random coil; that is this X value, if that turns out to be on this side then for that amino acid we consider it is in the alpha helical structure or if it lies on this side; that is this delta delta lies on this side, we consider it as a beta sheet, so same thing applies to the C alpha. In the C alpha case also we can think of the random coil value as Y and there is a boundary down on either side which is plus and minus 0.7 from the center. So, here we do not need delta, this is just the Y.

Now, if the chemical shift of any given C alpha of some amino acid; for example, let us say arginine, which we are considering, if the chemical shift of that arginine that turns out to be on this side that is the chemical shift, means that the minus the random coil that is a delta delta, if it will comes out on this side then that particular arginine will be considered as alpha helix.

Similarly if it lies within this zone; that is a chemical shift of that arginine, is within plus minus 0.7 of the random coil value, then we say it is in the random coil confirmation or it

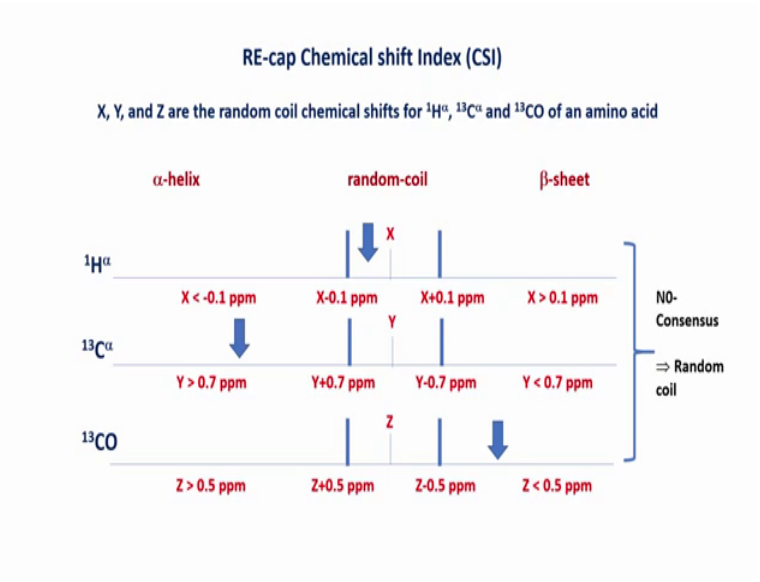
could be on this side then it will be beta sheet. Similar thing we will apply to carbonyl, there also we define a random coil value for every amino acid and with respect to that we say that it is an alpha helix if it is on this side and beta sheet on this side, but remember the chemical shift has to be subtracted to get this delta delta.

So, basically that is what is shown next now. We can see that suppose for any given amino acid, you see that the chemical shift value lies somewhere here; that is a delta delta is somewhere here, for C alpha the delta delta is somewhere here and for CO of the delta delta is here. Then all the three are on this side, which means all the three are telling us that that amino acid which you are considering, is going to be an alpha helical form. It could be that this all these three are actually in this zone; that means, whatever value you are observing for that amino acid.

Suppose it may happen that it is lying in within this boundary for all the three, then we can say that this is a random coil structure for that amino acid or it could be on this side, all the chemical shift three of them have the value such that they lie on this side, right hand side of the random coil then we can see it is in beta sheet confirmation. So, this is the general way of calculating the structure. Not calculating estimating the secondary structure of an amino acid in a protein by looking at this chemical shift values. So, but let us say that all three do not agree, so this is something we looked at yesterday that what happens, if let us say that the h alpha for a given amino acid let us say, we are looking at some aspartic acid number 14 in my sequence, and there it says that the random coil value that proton alpha is within the random coil value. So, it could be a random coil structure, but the C alpha is clearly telling us it is in alpha helical and co is also telling us it is in the alpha helical.

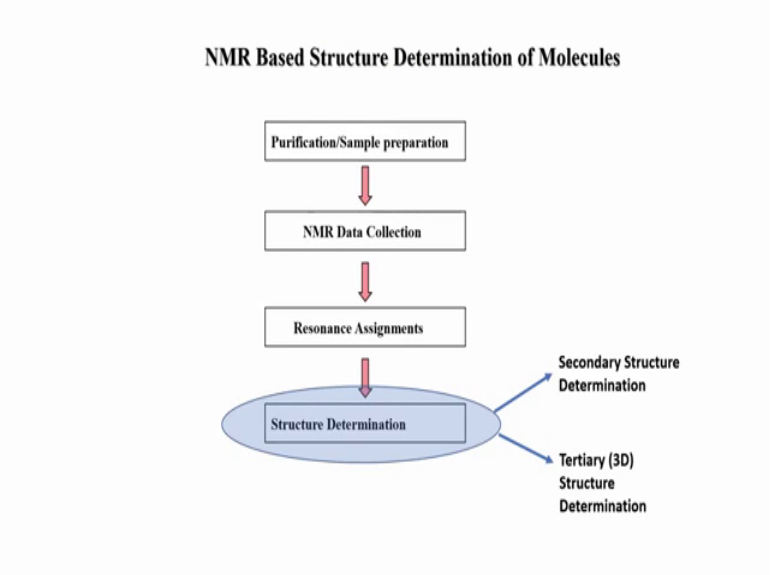
So, we can see that 2 out of 3 are agreeing to alpha helix. So, according to the consensus approach, we will consider that amino acid to be in the alpha helical form, because the majority 2 out of 3 are telling us that it should be alpha helical confirmation. For what happen suppose like this.

(Refer Slide Time: 07:13)



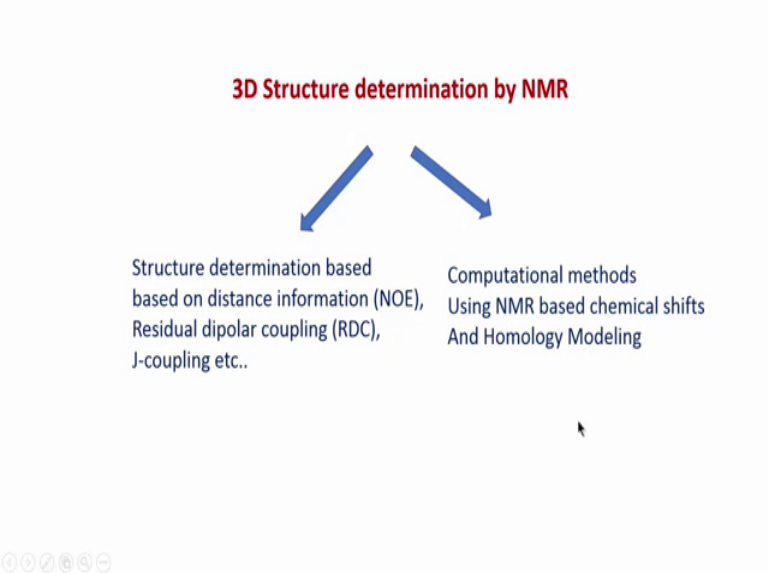
Let us say one says random coil, the second one say is alpha helix and carbonyl says beta sheet, what do we do in such cases? In such cases we can say that there is no consensus, all the three or disagreeing with each other. Therefore, we will consider that amino acid to be in the random coil. So, that amino acid is not considered in alpha helix or beta sheet confirmation, but is assigned a random coil confirmation. So, this is basically the CSI method which we looked at, this was a recap, so that we understand it better.

(Refer Slide Time: 07:49)



So, now let us move to the 3 dimensional structure determination part. So, 3 dimensional structure determination is typically the last step in a protein structure determination process by NMR, because after determining the structure you can then move on to determining the function of the protein, looking at interaction with ligand and all the other parameters. So, this is 3 D structure determination therefore, is almost the last step of the process, but remember to obtain 3 D structure we have to go through these, all these steps. Most importantly resonance assignment; resonance assignment has to be done completed, before you go to 3 D structure determination.

(Refer Slide Time: 08:33)



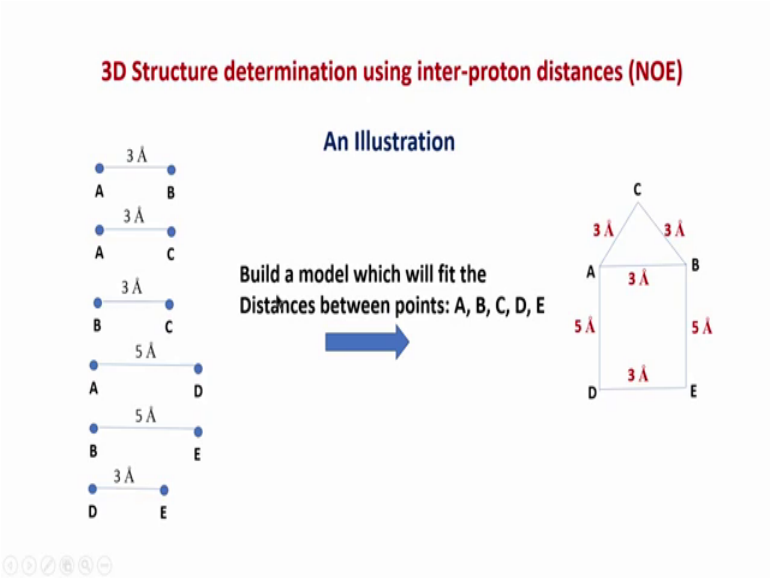
So, let us say, let us see how we can do 3 D structure solving by NMR. So, there are two broad methods today which can be used. So, one is structure determination method or process which is based on distance information. This is coming from this and no nuclear over Hauser effect. We have seen 2 D NOESY and in the previous part of the course. In this part now we will take up 3 dimensional NMR NOESY experiment, but idea is to get the distance information based on this effect and some more parameters are also acquired or needed sometimes too well defined or define it in a precise manner.

We will not go into detail of this particular parameter in this course; that is called residual dipolar coupling, but we will, I just given information that that is some time used for, some time for proteins which are very dynamic. J coupling is also used for structure determination which is already we saw that you can determine the J coupling

from HN HA and that J coupling value is combined with distance information to get the 3 D structure. Now we can also avoid N O E and J coupling based approach and we can directly use simple computational approaches, but which use NMR base chemical shift. That means, a chemical shift what you get from assignments are still utilized, are still used, but we do not go through this process of NOE and J coupling measurements, because these are little cumbersome, especially the NOE part is not so easy, because there are a lot of peaks, lot of information that has to be properly analyzed.

So, it is a time consuming process. So, what typically is done that once you have NMR chemical shift backbone and side chain chemical shift assigned, and if the protein has some homology structure known in the database, then we do not have to go through this step and we can use some computational approaches to solve the structure. So, we will not go into this part in detail in this course. We will just briefly I will show you some software which are used today for calculating on, based on computational approaches.

(Refer Slide Time: 10:51)



Let us focus more in detail on the NOE based approach. So, what is done in that approach? So, now, if you, this is a very important point which you should note carefully that in NMR, the 3 D structure determination process is nothing, but a model building ok. So, we are trying to build a model. We do not say structure actually we use the word model, because real structure of the protein is not known to anybody. We are indirectly obtaining the structure of the protein through some information; such as NO E J coupling

etcetera. So, the actual structure of the protein is actually never known to anybody. So, therefore, we are trying to construct a model and a proposed predicted model of the protein structure rather than the actual structure.

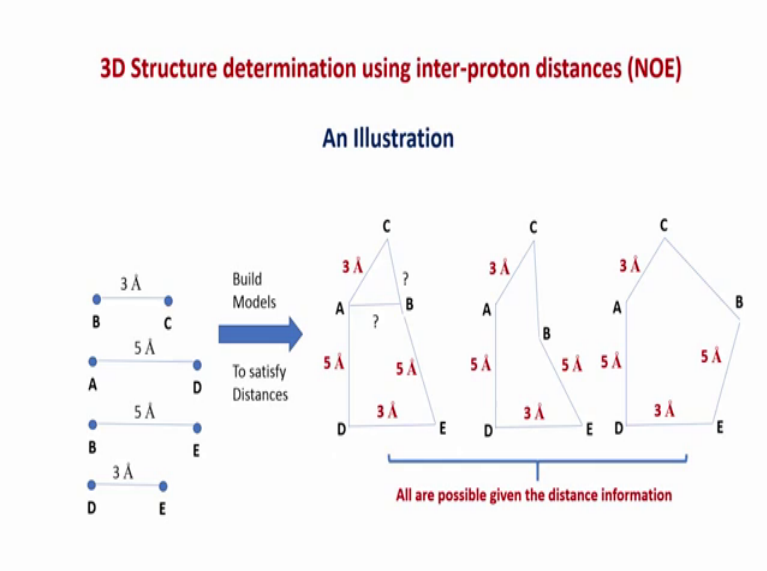
So, although I keep using the word structure everywhere, please remember that the final structural confirmation whatever we obtained from NMR, is just a model, it is just trying to represent the protein structure rather than actually being the structure itself. So, what is this model building, let us see an illustration here. Now let us say that you have some points, some arbitrary points A B C D E and let us say you have given the distance between those two points, any two points. So, between A and B, suppose you are given the distances as 3 Armstrong's and between A and C also you are given this distance of 3 Armstrong's, and between A and C also you are given this distance of 3 Armstrong between B and C is also 3 Armstrong's, and let us say you are told that distance between a And D is 5 Armstrong, and between B and E is 5 Armstrong and finally, between D and E is 3 Armstrong.

So, these are the distances or between two points which are given to you ok. Now the question is, how can you build a model which will fit the distance between the points, distances between these points ok, so this is a point. So, this is like a puzzle, you are given point distances, show me a figure which will fit all the distances. So, one of the possible model or a figure could be this on the right hand side. So, here you see all the distances are satisfied between the atoms or the point. So, you can think of them as atoms or we can think of them as points. So, what have you done here? We have basically taken some set of distances between atoms or points and based on those distances we have tried to build a model which has satisfied all the distances ok.

So, this is precisely, this is exactly what we do in NMR based protein structure determination which uses NOE as shown here. The idea here is NOE gives you at the set of distances between hydrogen atoms. So, these are A B C D E will be hydrogen atoms and between two hydrogen atoms, any two hydrogen atoms I can get distance if it is less than 5 to 6 Armstrong; that is remember that is the major point, NOE is only observed between two protons which are less than 5 to 6 Armstrong's. So, we try to get the distances of the protons from NOE and based on those distances we try to find a structure of the protein which best fits the distances which are observed. So, this is the basic overall theme, and therefore, one should pay careful attention to what is happening

here and this is what we will further take it up now and see how it is done with the nosy ok.

(Refer Slide Time: 14:37)



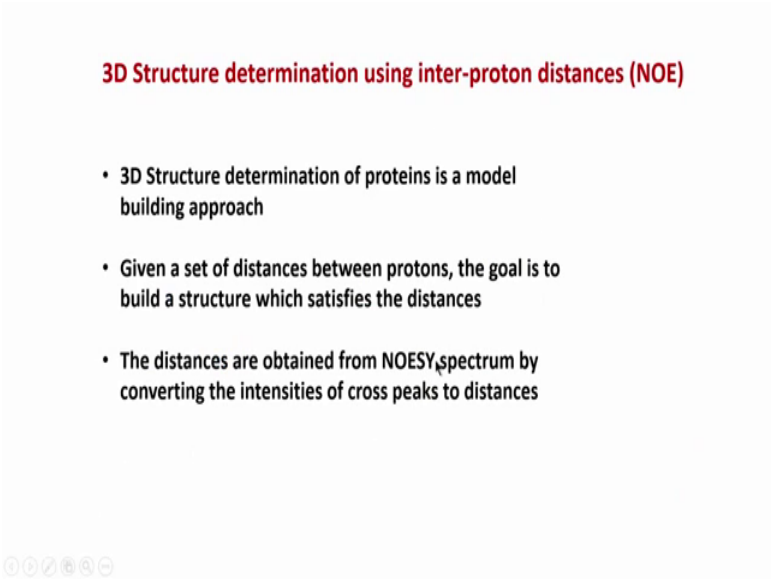
So, now suppose instead of all these distances let us say, I am missing two distances. For example, let us say I am missing these two distances ok. So, these two are not given to me, they are not there. Now can we build again a model which will satisfy the remaining distances, given that remaining are present for given to me, but I am not having the first two. What will happen in such a case let us see? In such a case because I do not know the distance between A B and B C shown here like question mark, I do not know that distance.

So, I can put B anywhere I want, because B to E is given to me here and all other distances are given to me here. So, those are satisfied means those are maintained, but A to B and B to C is not maintained, because that is not known, I have missing that, I do not have that information. So, you see this is one possible structure or model of a at, of a protein or a molecule based on those information that we have got. Now I can also build another structure something like this. Here also that other distances 1 2 3 4 which are known here are given to me, are satisfied, but A to B and B to C again is not known. So, I do not know what is a distance so I can put B anywhere I want, as long as it satisfies the other distances. There could be another possibility like this, here again I am satisfying all the other distances, but between B and C and A and B I do not have any information.

So, this is how model building is done by trying to satisfy as much information as you are given ok. So, all these are possible, all these three models shown here are actually possible, because it is satisfying the distances information in in of this, information of distances that are given to us. So, again coming back from to the protein structure point of view this is what is happening, happens in a protein structure determination process. We are given a set of distances between atoms, but what happens is, we do not have all the distances. Some of the distances will be missing, because there will be a missing peak, the intensity may be weak, there may be peak overlap so on so forth.

So, in such cases we will never get a distance information between a few atoms or many atoms actually. So, if the distance informations are missing we have to then rely on whatever is available to us and try to find the best possible structure or a model which fits all those distances or satisfies those distances. So, this is a similar thing which I was been mentioning in the previous slide, that NMR structure determination is a model building approach, wherein we try to satisfy the distances as much as possible ok.

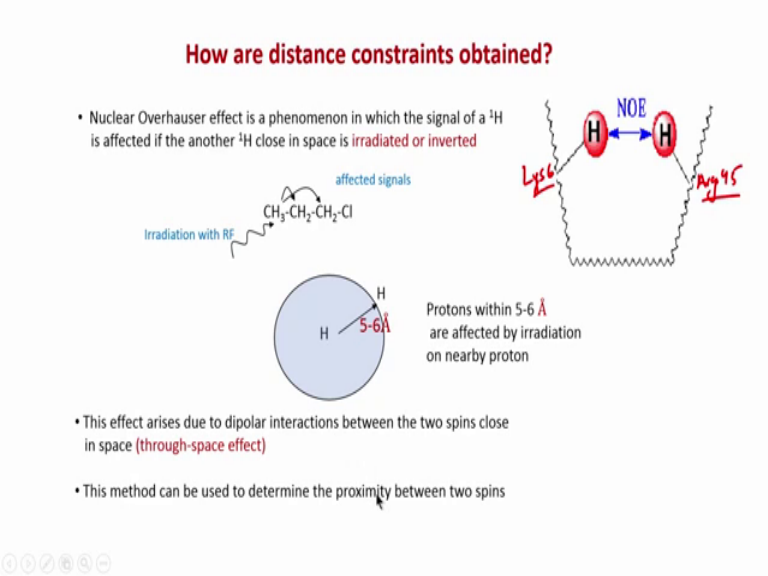
(Refer Slide Time: 17:41)



So, now how do we get this distances that is something we will see now. So, 3 D structure determination now, therefore, as I as I said is a model building approach, given a set of distances between protons, the goal is to build a structure or a model which satisfies this distances. Now the question is how do you get the distances, that distances are obtained from the NOESY spectrum, by converting the intensity of the cross peaks to

distance, and this is something we will see in detail now. We have seen earlier that in a 2 D NOESY spectrum, the distance intensity is proportional to the distance, how inversely proportional, one intensity is proportional to 1 over R to the power 6, we will see that again now. So, this is how we get from the intensity of the cross peak, we can extract the distance value between two atoms or two protons and those distances can now be used to build the structure ok. So, remember those distances what we will get from NOESY, is an approximate distance. Many of, many times people think that these distances are exact values you. No, that is not correct there is no way to get exact distance between two atoms from NOESY, it is an approximation ok. So, we are trying to all get approximation and finally, build a structure.

(Refer Slide Time: 19:09)



So, this is basically a recap or overview again of the NOE effect which we have already discussed in the earlier part of this course. So, this slide is, slide is a repetition from there. So, let us have a look at NOE again, because we did it quite long time back. So, let us refresh our understanding of what is NOESY. So, Nuclear Overhauser Effect, is a phenomenon in which the signal of a proton is affected if another proton which is closed in space. So, this is a point here, close in space is irradiated or inverted. So, if you look at this picture on the right side. If you have a two protons which are close to each other, now close meaning between 5 to 6 Armstrong's then if I perturb or do something to 1 hydrogen; the second hydrogen's population alpha to beta state is affected. Similarly, if I

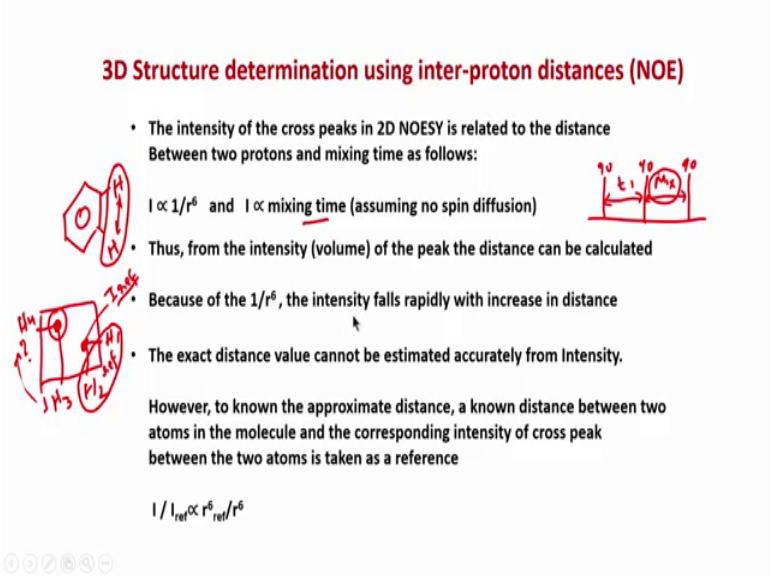
perturb the population of this hydrogen by a pulse that will affect the population of this hydrogen.

So, there is a population effect on each other by these two protons, because of their perturbation to one of them. So, typically in the olden days we used to irradiate the RF pulse. The continuous eradiation was applied on a particular proton and that resulted in affecting the nearby protons, but that is what is called steady state NOE, and that is not done nowadays, what we do is, what is called as transient NOE in which you invert, likes written here. We invert the magnetization of the both the protons and when after inversion when they go; that means, they go to minus Z, when they start coming back towards Z, they start transferring polarization that is a population from one to another and that transfer of polarization is captured as a NOE effect.

So, we will see the pulse short equal, sequence shortly how it works. So, this is what is shown here, that if there are to proton, any two protons for a given proton this radius of 5 to 6 Armstrong, any proton in that radius the zone, within the zone is affected if I do something to in this hydrogen ok. So, this is basically the idea of NOE and this is a through space effect. So, which means if you go back to this picture on here, this two red colour hydrogen atoms are not connected by a bond, they need not be connected.

If they are connected by a bond then definitely there will be a interaction, because they are close to each other, but if they are not connected also like here, you see there is a wavy structure shown here just a schematic of a protein, they may be very far away these two hydrogen in terms of the sequence. So, for example, let me write down here, this could be lysine number 6 or and this could be arginine number 45. So, you see they are quite far away in terms of the sequence, but they are quite close in space. So, NOESY effect the type, is due to the dipolar interaction between two spins or hydrogens, close in space. And as I said this can be therefore, used to determine the proximity or distance between two hydrogen atoms.

(Refer Slide Time: 22:25)



So, let us see how this is done. So, this is typically done from 2 D NOESY in the case of peptide. We will use 3 D NOESY now in the proteins, I will show you how it works, but basically the idea is you take the intensity of a cross peak, whether it is 2 D or 3 D NMR it does not matter, any cross peak between two hydrogen atoms is considered and converted into distance like this ok. So, now, this intensity what we use or what we actually measure, is proportional to the mixing time. We will talk about spin diffusion shortly, but what is this mixing time, if you recollect, if you recall in the pulse program which will come again now, but if you recall for the time being, let me write down here, we had a 3 pulse pro experiment in NOESY, this was a 90 degree pulse, another 90 degree pulse, another 90 degree pulse and we call this period in a NOESY as T 1 and we call this period as mixing time.

So, this mixing time is what is mentioned here ok. So, this typically is of the order of 60 to 80 milliseconds in case of proteins, and for peptides we use a little longer mixing time. For peptides we go up to 200 millisecond to 50 milliseconds. So, that has to be optimized so that we can avoid this concept of spin diffusion which I will show you shortly. And then from the intensity whatever we get, we can get the distance information by using this formula. Of course, directly we cannot use this formula, there is some calibration has to be done and if typically we use a reference distance if available, and from the reference distance we calculate the actual distance for some other unknown to distances, unknown distances.

So, we will see that we will not go into detail of those calibration approaches, there are a lot of technical details involved, but roughly, qualitatively, a broadly you can think of a distance getting from, the distance is therefore, related to intensity in this manner. Now because of this power 6 here, the intensity falls rapidly with increase in distance; that is why the limit of 5 to 6 Armstrong comes, because as you go to 7 8 or 9 Armstrong, the 7 to the power 6 or 8 to the power 6 which will come here is very very small number 1 over 8 to the power 6 or 1 over 7 to the power 6 or 1 over 9 to the power 6 compared to 1 over 3 to the power 6. So, typically when you go beyond 3 4, or even 5 Armstrong, the intensity will fall very rapidly because of this R to the power 6 dependence and therefore, we restrict, we say that the distances are restricted to not beyond 6 Armstrong's.

Typically it is about 5.5 and 5.5 is typically seen as a maximum distance beyond which the two atoms will not show any cross peak call the cross peak, we will be very week ok. So, as I mentioned earlier above that exact distance value can never be estimated from intensity, this is an approximation. So, it does not matter because we really do not need exact distances in protein. We are trying to build a model, a dynamic model of a protein and therefore, exact distances can never be obtained and does not really is not required. So, as a I was also sell, telling that this formula I is proportional to 1 to divided bar to the power 6, we use in a with respect to a reference. And with respect to a reference if you do, so you can see here that if I know the resistance of a reference intensity of a cross peak between two reference atoms, and if I know their distance between those two atoms, then for my unknown distance are can be obtained by simply using the intensity from the spectrum ok.

So, by using this approach with respect to reference, I will be able to very nicely simply get the intensity. So, let me illustrate this here with a picture. So, imagine this is your NOESY spectrum, there is this cross peak between two atoms H 1 and H 2, let us call it these two. And suppose there is another peak here, between some other peak H 3 and H 4 ok. So, suppose we know the distance between H 1 and H 2, that we will call it as a reference. If you already know the distance how do we can, how can we know. Sometimes these two can be a part of a rigid structure, where the distances are well known for example, if I have an aromatic system this proton and these proton, these are fixed, you cannot change much the you, whatever you do their distances are quite fixed

in the protein. So, if I take such two protons where cross peak is there. So, I can take this intensity as a reference and this intensity and distance also I know as a reference. So, I can use that formula here and I know the intensity of this now, but I do not know this distance, this to this distance.

And for that I will then use this formula, because I have three known parameters, I know the reference intensity, I know the reference distance, I know the intensity of between two atoms, but I do not know that distance. So, if I use this formula I can get the unknown distance ok. So, this is how basically the structure determination process works. We will now we have only looked at how the distance information can be obtained from 2 D and 3 D NOESY.

So, in the next class we will take up the actual 3 D NOESY and how 3 D NOESY experiments are built, because 3 D NOESY is what we very regularly do in proteins, 2 D NOESY is done only for peptides. So, we will look at the pulse sequence and other things in 3 D NOESY and from there we will take up how to determine the 3 D structure.