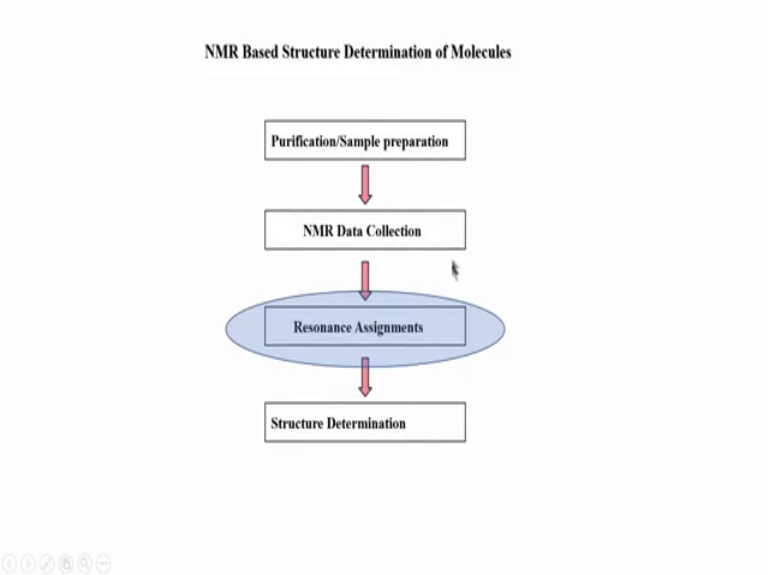
Multidimensional NMR Spectroscopy for Structural Studies of Biomolecules

Prof. Hanudatta S. Atreya
Department of Chemistry
Indian Institute of Science, Bangalore

Lecture – 28
Resonance assignments of Proteins – Part I

We will now begin with how protein structures can be determined by NMR. Before we start looking at actual structure determination, the first one step you have to do before that is called Resonance assignments.

(Refer Slide Time: 00:42)



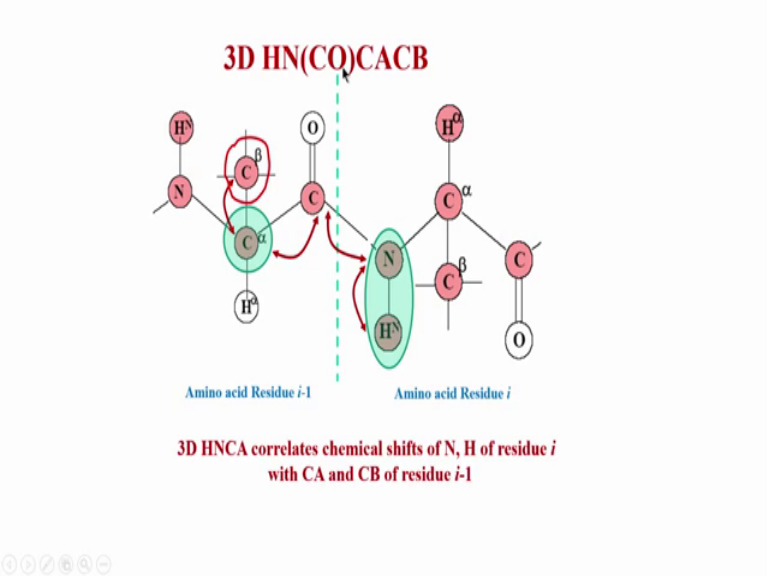
So, this is the flowchart shown here for protein structure determination in general. So, we have gone through this step in the last few classes, we looked at isotope labeling and how sample is prepared. Then we looked in the first part of this course how to collect data what are the different 1 D, 2D, 3D NMR data that we collect how they are collected.

Now, we will see that once you get those data, how do you actually assign a protein. So, this is very important step here assigning means, that the chemical shift of every hydrogen carbon and nitrogen in the protein should be known means what? It belongs to which amino acid it belongs to what is its chemical shift and so, on. So, that is very important to know and that is not a trivial step, it takes lot of time energy and effort to do this. Sometimes you may not be able to achieve success here at 100 percent, typically if it maybe 160 70 percent. So, that is a way 100 percent success is very difficult especially

if the protein is very large in size and therefore, because we have lot of peaks which are missing sensitivity of the signals are low and so, on

So, its very difficult step here, and once we achieve this then we go to the next level or next step where we actually solve the structure. So, we will come to this once we have understood how we can assign a protein.

(Refer Slide Time: 02:09)



So, let us see again how what is assignment procedure, what is to be done in assignment. So, first thing you have to know is the exact sequence of the protein. So, as I said this has to be given to us we should know this correctly. So, what is a protein primary sequence? A protein primary sequence is a linear sequence of amino acids, join to each other by the peptide bond. So, these bonds are peptide bonds. So, we can see here a schematic drawing of some protein. So, these are different amino acids. So, one thing of course, you notice in this protein is at many in some of the amino acids are repeating.

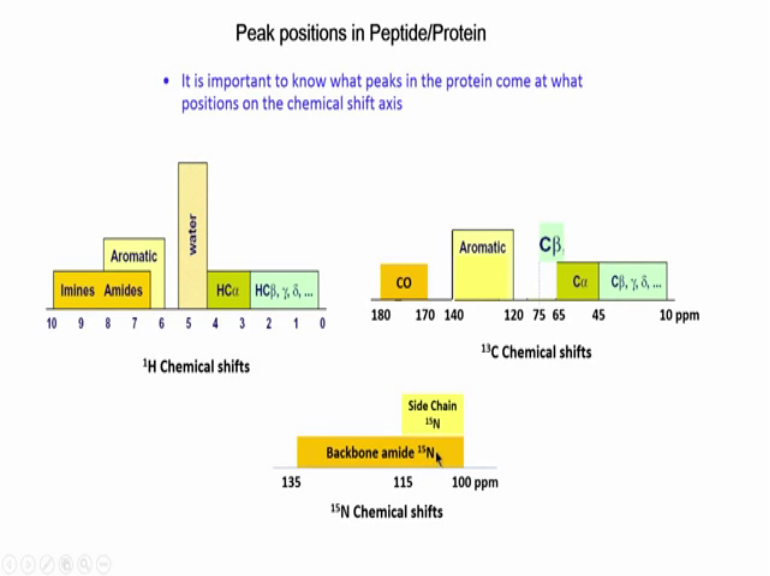
For example leucine is coming twice in this chain. Similarly you can see lysine is coming twice ok. So, like this many amino acids are not just twice, it can come multiple times. So, therefore, our goal in a chemical in assignment, resonance assignment is to assign the chemical shift of each peak can be over spectrum to the respective amino acid type, but not only that we need to also tell that this chemical shift is lysine num[ber]- for lysine number 17 this is for lysine number 24 like that. Similarly this is the chemical

shift of leucine number 16 is so, and so, chemical shift of leucine number 12 is so, and so.

So, we are not only just telling leucine, here also identifying which leucine in this sequence in the order that we are chemical shift belong to. So, therefore, we should have the information of the order means the sequence, then only we can start working on the project and then try to figure out which amino acid and which at lysine leucine etcetera have what chemical shifts. This is therefore, one confusion which everybody many of the beginners in NMR have, they ask the question why do you need the sequence, but sequence is needed as we see it going we as we go long you will see that we need this to assign the protein.

So, this how do you obtain the sequence? Sequence typically is obtained by various biological and chemical methods, which we cannot go into in this course, but I would refer you to some biochemistry book, where you can see how the sequence of proteins are analyzed. So, typically if you know the genome the DNA of if of a cell, you know that if you know the open reading frame or you know which gene corresponds to which protein, from the gene sequence you can then extract the sequence of the protein that is one of the methods.

(Refer Slide Time: 04:44)



The next thing which we should know for assigning a protein is that, where do the chemical shift come generally. So, this idea one should have very well in the mind ok.

So, for example, the in the proton this is your looking at a 1 D proton spectrum. So, if you if you look at a 1 D proton spectrum, typically the aliphatic groups will come somewhere here from 0 to 2 or 3 ppm. So, these are all the beta protons of methyl of the amino acids, gamma protons of the amino acids, similarly methyl groups will all come somewhere between 0 to 1 ppm and they are very easily recognizable because they are very strong. Then you have this region between around 3 to 4 ppm or 3 to 4.5 where the H alpha H alpha of the amino acids will come here.

So, remember is the H alpha is the backbone proton alpha of every amino acid very amino acid has H alpha. Now comes next after that is a huge water signal because remember we make our protein samples in water H 2 O. So, the H 2 O is a huge signal compare to all other signals. So, that will come here then comes aromatic region where all the aromatic amino acids like phenylalanine, tyrosine, tryptophan, hysterine etcetera will come there. Now there is the amids also start coming in this imines of course, we do not worry about in protein, but the amides will start also overlapping, but slightly shifted they come between amides come between 7 to 10 ppm.

So, they are slightly away from aromatic although it will here it looks overlap, but they can be separated, but sometimes amides and aromatic will also be clashing. So, this is a typical range of chemical shifts which are observed in proteins for as for as protons are concerned. Now let us look at carbon where do the carbon chemical shifts come? And carbon chemical shifts comes in this range. So, typically between 10 to 45 ppm you will get all the aliphatic C beta C gamma C delta chemical shift methyl's etcetera will come somewhere close to 10 to 20 ppm 10 to 25 and the remaining all come in 25 to 45 ok.

The C alpha starts from 45 for glycines and goes up to 65. This is something which we saw in the experiment called HNCA and HNCO CA where they come. So, this is the range typically for C alpha of all amino acids. Then the some of the C betas all the C beta is comes in this zone mainly, but two amino acids serine and threonine there C beta comes between 65 to 75 and why is that? Because a C beta is attached to a OH group. So, because of that there is a deshielding effect and the C beta comes in this range then between around 75 80 to 75 to 120 there is a blank region there are no peaks in this region in proteins and then you start getting the aromatic carbons SP 2.

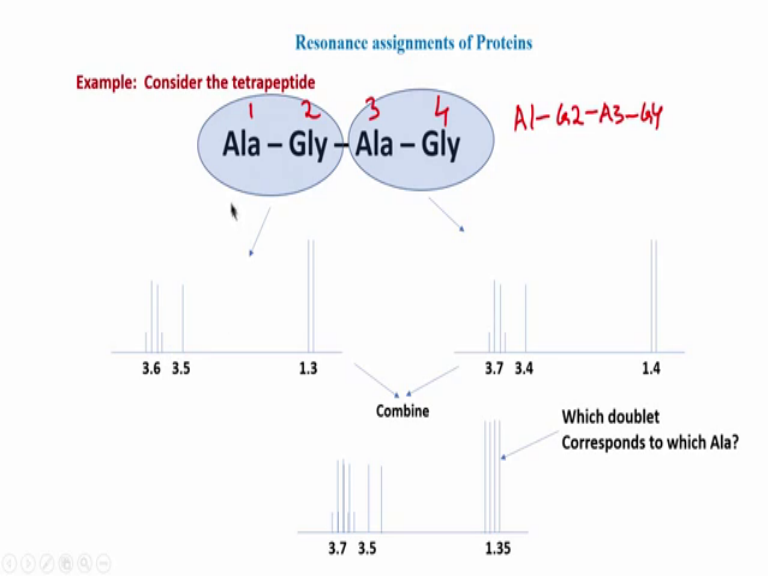
So, aromatic carbons are coming in they come in this region, then again there is a gap and then the carbonyls that is amide carbonyls they will come in this region ok. So, this is typically the range for the protein chemical shifts as well as carbon is concerned now let us see for nitrogen. So, these are basically the 3 nuclei which we are going to look at in the chemical shift assignment. For nitrogen typically it comes between 100 to 135ppm ok.

So, the glycines typically the glycines you know they come somewhere in the up field regions they come between 100 and 110 ppm, but other amino acids are all together. So, there is no amino acid specific information in nitrogen. But if you will see later that in carbon as I said already here for example, serine and threonine only come here for c beta. So, if I see a C beta chemical shift in this zone, then immediately I can close my eyes and say that belongs to either serine or threonine. How do I know it is C beta? First of all and it is not C alpha there are some ways to sub distinguish in some experiments.

Similarly, C beta here of alanine comes in a very specific zone between 15 to 20 ppm and again that any C beta which comes there can be immediately assigned to alanine. So, C beta chemical shifts do carry information of the amino acid type, but N 15 unfortunately does not carry. So, based on chemical shift of N 15 accept glysines is very difficult to tell which amino acid that is. Now remember also asparagine and glutamine side chains also contain amide CO NH 2 groups. So, the amide hydrogens of asparagine and glutamine side chain will come in this zone between 100 and 115 ppm. So, they come at the right hand side top of HSQC.

So, when we count the number of peaks of HSQC, we normally have to take into account that there are side chains also possible. So, we have to identify them and see them ok. These are the typical ranges of chemical shift for proteins as far as proton carbon and hydrogen is concerned. So, one should have an idea of where do which peaks comes where so, that helps to assign the protein.

(Refer Slide Time: 10:27)



Now, let us say we will take an example of very simple system to understand what is it mean by assignment what do we actually need in assignment. So, let us say we have a tetrapeptide something like this, alanine glycine alanine glycine. So, this is kind of a repeat the dipeptide is repeating. So, it is a very strange sequence, but can come in a any protein also.

So, let us say we have this tripeptide and now we want to assign the chemical shifts. So, now, you see here there are two alanines and two glycine; which means I should not simply tell that this is a chemical shift of glycine, I should also be able to identify which glycine is which chemical shift, not just saying the chemical shift is glycine some peak I should be able to say that is a chemical shift of glycine number 2. So, chemical shift of glycine number 4, this is alanine number 3. So, we can actually number this peptide 1, 2, 3 and 4 ok.

So, I can write it in a single letter code like this, it is A 1 G 2 A 3 G 4 G is for glycine A is for alanine. So, now, we have to assign let us say this tetrapeptide how do we do that? So, let us see first test spectrum 1 D spectrum. So, let us say that these two dipeptide have this following chemical shift. Suppose I make this dipeptide separately in a peptide synthesis way I may get a chemical shift spectrum like this. So, this tetra this is quadrate belongs to the alpha of alanine and this doublet corresponds to the methyl group of alanine and this singlet is from the glycine H alpha.

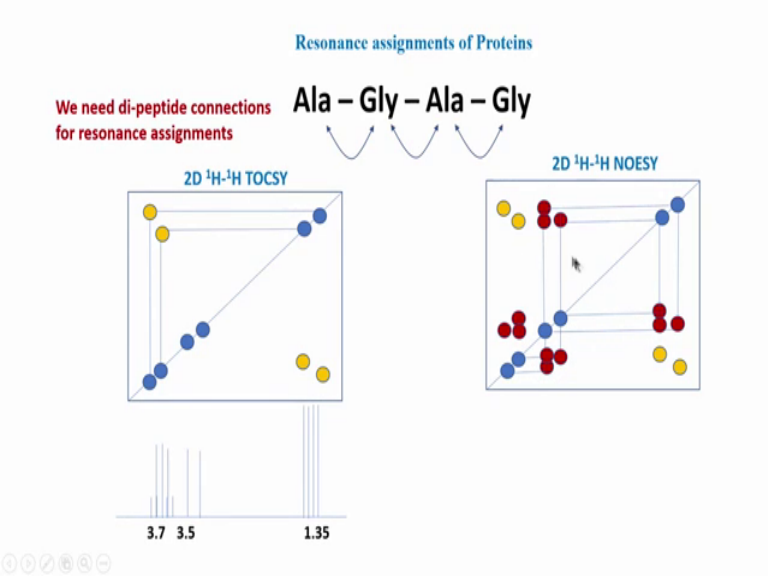
This is just a schematic, this is not a real spectrum this is just a schematic drawing what you would expect to see if you had a alanine glycine dipeptide. Similarly this alanine glycine may be when you attached to this not separately, when you attach it may have a slightly different chemical shift ok. So, I am not taking the same peptide now, I am just showing you that this chemical shift of this alanine glycine may not be same as this alanine glycine, because the chemical environment is now different there are tetrapeptide. So, therefore, it may have different values.

So, when you combine this all 4 together you may get finally, a spectrum which corresponds to the full tetrapeptide; where you see here there are two glycines, there are two quartets here and there are two doublets here. So, the question now is which doublet corresponds to which alanine ok. So, that information is not here, because when I record a spectrum of a mixture not like this this was done just to show you, but in reality when I make a tetrapeptide and I record a 1 D spectrum, I will simply get something like this. So, I may, will not be able to tell which alanine doublet here which of course, I know there are alanines here, because I know the sequence and I need I need I can see that there is alanine in my spectrum.

But which alanine peak is which alanine number 1 or 3 that information is not there here you cannot do anything to extract it from here. Similarly glycine there are two glycines I just do not know which is G 2 or which is G 4, again that information is not present in this 1 D spectrum. So, that is why this whole game of business of assignment resonance assignment becomes important because just looking at a 1D you do not get that information at all which is which. So, even from 2D you will not get. So, we can do 3D. So, let us see at least with a 2D can we assign this tetrapeptide, how do we do that? So, let us say that we have a TOSCY spectrum. So, what is needed in a resonance assignment of protein is that we need dipeptide connection.

So, this is something let me go back to the previous slide. So, here you see we had the amino acid we knew there is a glycine aniline here. So, individual amino acid information is present in this spectrum. I know there are two glycine there is an two analine so, on. But I do not have a connection or correlation between the neighbouring amino acid between the neighbours. And that is what is actually needed in a assignment procedure.

(Refer Slide Time: 14:51)



So, any protein assignment procedure which you way will come across remember it always will try to find the dipeptide connection that is the heart that is the importance important point in any assignment procedure.

So, I would recommend you to pay careful attention to this point that we have to we need a dipeptide connectivity, then only I can go from one to the next to the next and walk along the sequence otherwise these are all independent amino acids, I will not be able to get simply without if there is no connectivity information given to me, between neighbouring amino acids. So, neighbouring means one I suppose this is I then this is I minus 1 or if this is I this is I plus 1. So, I need some information of connection between these two between these two so, on. So, which experiment will give me that information TOSCY will not give me that.

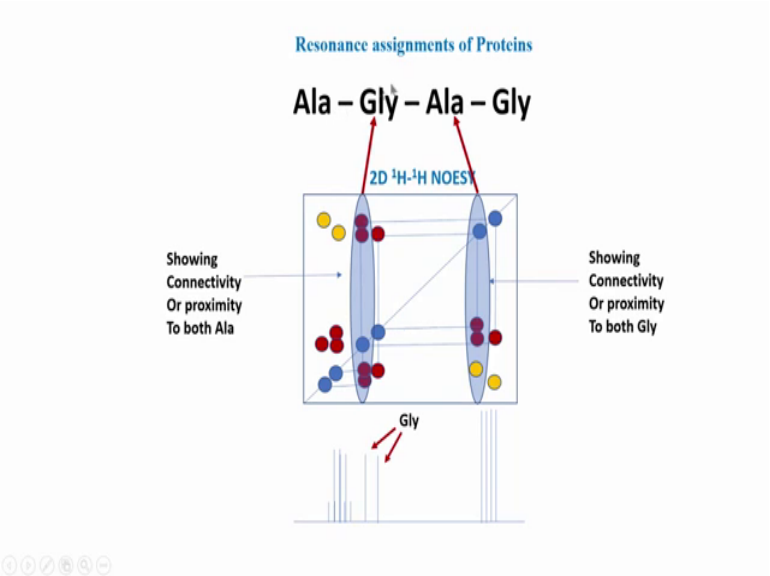
So, we will see in TOSCY what do we get? In a TOSCY we simply get information of a connectivity within one amino acid. So, this is an alanine. So, this is what I showed at the 1D there are alanine here, there are two alanines here. So, the one which is connected to this is corresponding partner it is connected with a cross peak. So, this is one spin system this is something we have seen in the 2D part. Similarly the other alanine also we will connect within the alanine ok. So, this is an alanine with alanine its own connection and glycine of course, there are two glycines there is no connection because there are only single protons.

So, you see here this TOSCY is not helping me much, it is only giving me a connection within a given amino acid which is fine which I need to know at least which amin[o]-alanine which peak here is connected to which here. So, that at least that information I am able to get it here which I earlier had not got from 1D. So, that is fine. So, I am able to get some information from TOSCY and there is a transpose peak will come here which is now shown here. So, this is exactly equal and opposite and the same connection information is shown here getting here.

So, what do you do then to get the interpeptide connection which I mentioned? For that we need this NOESY. So, a 2D NOESY will now help us to do this assignment now let see how it helps. First we need of course, TOSCY to find out which peaks are the self peaks self means the same spin system so; obviously, that is these peaks are corresponding to this here, this corresponding to be that here. So, the color looks same. Remember NOESY is not a colour spectrum like this in reality in re practice it will come all in the same colour, but I have just shown in different colour to help you identify the peaks.

So, you can see here now these two peaks are coming from TOCSY. So, all the remaining peaks are from NOESY of course, this diagonal is same as TOSCY diagonal. So, now, let us pay little more careful attention to this red colour cross peaks. So, there is something I will show you in the next slide.

(Refer Slide Time: 18:04)



So, now look here. So, basically you have two glycines here ok. So, we know there are two glycines how did we know that? Because in TOSCY in a 1 D remember I and just told you there are two glycines which from the simple chemical shift we can say, but still I do not know which glycine is this and which glycine peak is this. So, G 2 and G 4 that assignment is not done up to now.

So, how can we achieve that is what we will see now. So, now, let us focus on this glycine here. If you look at this glycine what do you see? You see that if you go or vertically or down be either way. So, let us go vertically it is showing me connection or correlation to both the alanines, these two are alanines right beta proton. So, it is showing me to both the alanine similarly when I go to the alpha side, it is showing me to both the alpha protons. This is alpha proton these two alanine these two are beta protons. So, at means this glycine is connected to both the alanines, that we where is this connection coming from? It in NOESY where does how does the correlation come? It comes because of proximity means less than 5 to 6 Armstrong.

So; that means, this glycine which is this particular glycine here, it showing me connect proximity or distance close distance to alanines, which means it could be only this glycine. Because only this glycine you see is close to both the alanines whereas, this glycine is close to this alanine, but it is far away from here so; that means, now I can assign that this peak this glycine peak is definitely G 2, because it is showing me correlation to both the alanine. So, you see the neighbouring information immediately helped me to say that this is glycine number 2.

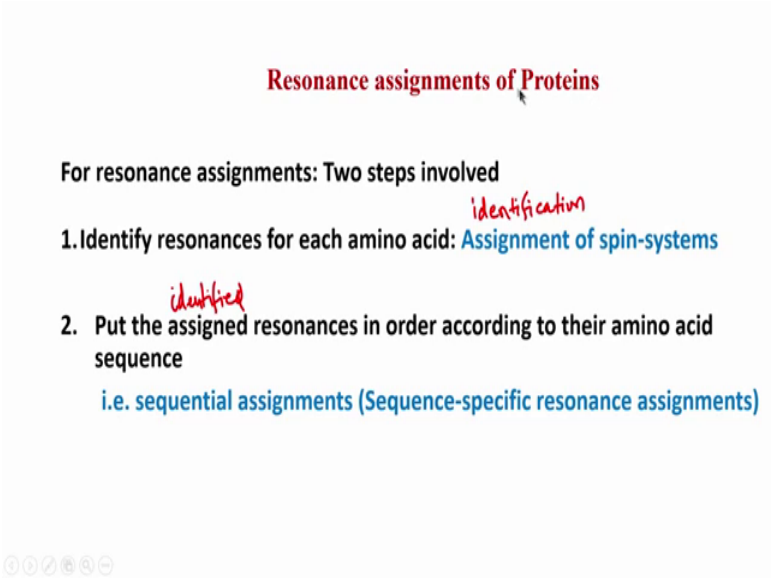
Otherwise which glycine is which I could not have mention it deduce just based on TOSCY. Similarly now look at this alanine this alanine peak here this alanine peak here is showing me to both the glycines you see this alanine so; that means, this alanine is the similar to this glycine, is showing me connectivity or proximity to both glycines; that means, it has to be this alanine only because it is the one which is sandwiched between two glycines. So, it is equally closed here equally closed here. So, the H beta of this alanine is having a NOESY cross peak with this H alpha of this glycine and this H beta of this alanine is having a cross peak two the H alpha of this glycin also because they are equally close.

So, you see this is how? Now you can clearly say that this peak which alanine we here again we had two alanines, I mean there are doublets for each alanine. So, we do not know which alanine doublet is which now I can say this left side doublet is clearly alanine number 3. So, now, what is remaining? One glycine remains another alanine remains so; obviously, one glycine which remains is this and this alanine which remains is this ok. So, basically that is how we assign a protein of this was a very simple example, but what I you can also look at this part here, here also if you see this alanine showing to two glycines. So, it is this alanine and this glycine is showing to 2 alanines which is this glysine.

So, the same information either you go from this beta proton side two alpha proton or you can go from alpha proton to alpha proton. So, there is no issue we can use any one of this chemical shifts to do that. So, this is a simple example to tell you that once we have a neighbouring connectivity information, we should be able to connect two amino acid with each other and with help of that we should be able to find out which glycine is which are which alanine is which in this case or which amino acid is which chemical shift from a spectrum. So, in a 3D let us see how do we do that. So, before that, so, basically if you go back to this picture and whatever we did here, we can actually see what are the two steps we did first. So, if we go back one more step.

So, here let me go back to this here we first identified that there are two glycines and we first identified that there are two alanines based on the 1D spectrum you can use the 2D also the first thing we did was is summarized here.

(Refer Slide Time: 22:47)



The first thing is we first identified the chemical shift we did not assign it we are just identified what are the amino acids present. So, I expected two glycines and two alanines because I knew the sequence. Again remember we comes back to this point that the primary sequence should be known to us without knowing the primary sequence we cannot proceed.

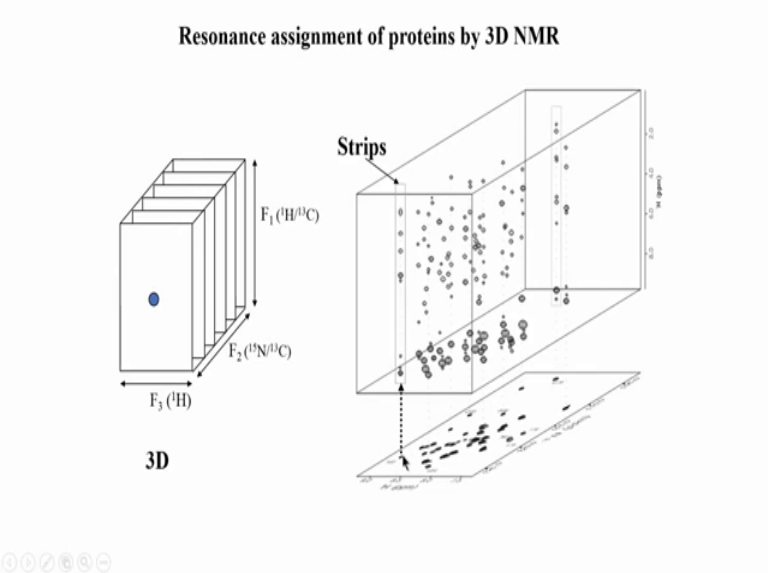
So, given that a amino acid sequence was known, I knew how many glycines to expect, I knew how many alanines to expect so, that is the first step I did I assign the spin system means I identified. So, I would also write it here as identification. It is not just assignment we are not assigned at this stage we have identified. So, it is an identification of spin system in a spectrum. So, which spectrum that was a TOCSY. So, once a TOSCY spectrum you identify a spin system, then you can go to NOESY and then actually put the assigned or identified resonances in order.

So, I can again write here identified resonances, it is not a assigned yet identified resonances in order; order means in the correct order in the sequence. So, that second step which we I showed you using NOESY is called sequential assignment or sequence specific resonance because now specifically we are telling glycine number 2 is this peak, glycine number 4 is that peak and so, on. Here we did not do that here we just simply say there are two glycine; there are two alanine, there are two other amino acid or n amino acids, but which amino acid is which chemical shift was only achieved at the second

step. So, this is as far as the 2D part is concerned means using NOESY and TOSCY and this is something we have extensively also gone through in the last course, using specific peptides we will not use specific peptide examples here we just gave a simple example.

Now, let us move on to proteins means in a larger protein sequence what you do? So, number 1 is you need to label the protein this is. Why because we need 2 D is not sufficient like we did for TOSCY a NOESY for peptides we need 3D NMR. So, when it comes to 3D NMR I have been telling this that we need to label the protein we cannot to 3D NMR without labeling.

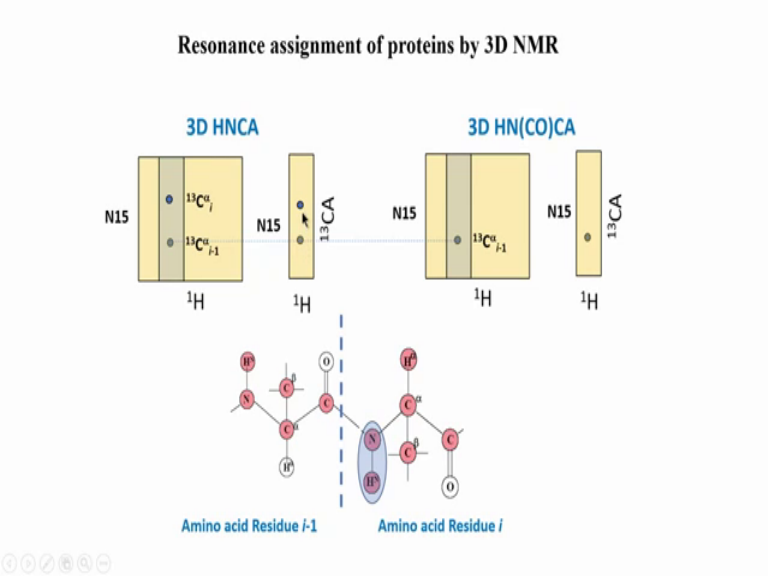
(Refer Slide Time: 25:18)



So, now let us see in 3D NMR how do we analyze? So, in 3D NMR again this is something we repeat. So, earlier also we have to now analyze with 3 dimensions and we use a strip plot. So, for example, we can see in a 3D NMR, the base plane is first identified this is typically HSQC nitrogen proton HSQC and from that HSQC one every peak in the third axis we get chemical shift in the third dimension.

So, we draw at strip the of course, this is not physically done, this is done by a computer. Computer will plot a strip from the 3 dimension based on what you choose here and the strip plot will be displayed to you on the screen and based on many strip plots in the spectrum you can then start assigning the protein. So, let us see how we can do that with 3D NMR.

(Refer Slide Time: 26:12)



So, I will give an example from this very simple two experiments HNCA and HNCOCA. So, if you recollect basically what we had in HNCA and HNCOCA is that, we in HNCA we got for every pair of nitrogen and proton chemical shift this is basically NH of a peptide backbone so; that means, for every amino acid we got two peaks in the third axis and that axis consists of carbon sulphur C alpha i minus 1 and i.

So, these are the two peaks which we saw in the 3D HNCA. So, this is a stripped version of the 3D where in the stripped version for a given pair of an 15 NMR proton chemical shift; that means, for a given amino acid, remember these two combination represent one amino acid we got two chemical shifts whereas, in HNCOCA we got only one chemical shift for a given N 15 proton pair and that corresponded to the i minus 1. So, this is shown here in diagram. So, for a given amino acid i we got correlation to C alpha I n C alpha i minus 1 in the HNCA, but in HNCOCA we got only 2 i minus 1.

So, why do we need these two experiments? Because actually with most important is this experiment, but in this experiment this spectrum I will not be able to identify which is i and i minus 1 therefore, to help me do that, I need another experiment and that gives me the information of which is I and which is i minus 1. So, I need that distinction on that discrimination between the two peaks to proceed further. So, for proceeding further, we need to connect now the neighbouring amino acids. So, we saw i and i minus 1 now we have to go to i minus 2 or i plus 1 like we did for alanine and glycine. So, that

connectivity information is something you will proceed in the next class, where we will see how we can actually join the different amino acids chemical shifts along this side or this side and then continue this assignment so, that we can identify the correct sequence and identify the chemical shifts. So, we will continue in the next class with 3D NMR based assignments of proteins.