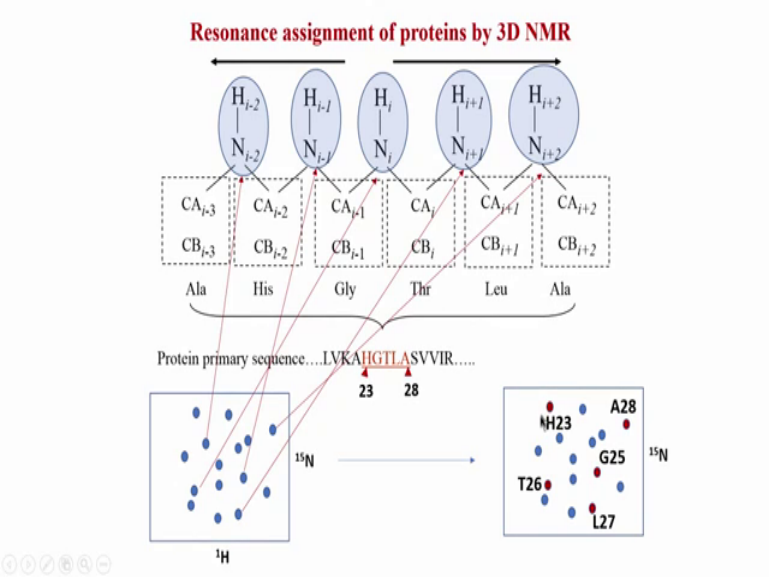
Multidimensional NMR Spectroscopy for Structural Studies of Biomolecules
Prof. Hanudatta S. Atreya
Department of Chemistry
Indian Institute of Science, Bangalore

Lecture – 30
Resonance assignments of Proteins – Part III

So, we will continue with Resonance assignment of Protein by 3D NMR. So, we saw in the last class that if you have assigned a stretch of amino acids and you have identified what type of amino acid these are, the next step is to mapping it on the protein primary sequence. And if the assignment is correct and our identification is correct, then it should find a correct match. So, we can see here that we have matched up to 5 or 6 amino acid in the protein primary sequence, but this is not the complete sequence yet, we have not completely assigned the protein, we have only identified 5 or 6 amino acids assignment. But that is good enough as a starting point and slowly, we can go expand this in both directions. So, let us see how we can do that.

(Refer Slide Time: 01:14)



So, this is shown here again. This amino acid what we identified which we matched on the sequence, let us say is some number 23 that is this amino acid is number-23 and this amino acid is number-28. So, if you look here carefully, what has happened is we have actually assigned here it shows 6, but actually we have assigned 5 NH, because for this i

minus 3, we have not got i minus 3 NH which should be here why, because we for that I need i minus 4 and so on.

So, we can go for a when I say i minus 6 amino acids actually, we assigned the N H of only 5 which is showing here ok, so that is why we labelled from i 23 arbitraries, let us say this was something 23, 24, 25, 26, 27, 28. Now, how do I know this number, this number of course comes from the protein primary sequence. The protein primary sequence tells us that this is number 23 to 28 and that is something we have assigned based on H N, CA, CB.

Now, if you look in the HSQC, these are each of these was one peak, each of this is one peak correct. So, now let us say that this arbitrary N i H i, which we started was somewhere here or N i minus 1 is somewhere here. And N i minus N i minus 2 is here N i minus 1 is this peak, this N i which we started could be this and this is something here, and this is something there.

So, you see we have basically we know this positions, because this is coming from the spectrum, but we did not know earlier which peak in the sequence corresponded to this. We did not know which sequence protein in sequence in the protein, which amino acid of the protein was this peak, but by doing this linking and matching or mapping on the sequence, we have now achieved this identification. Now, we can say that this peak number here is basically H 23 sorry this peak here, so this red colour is one it should be here.

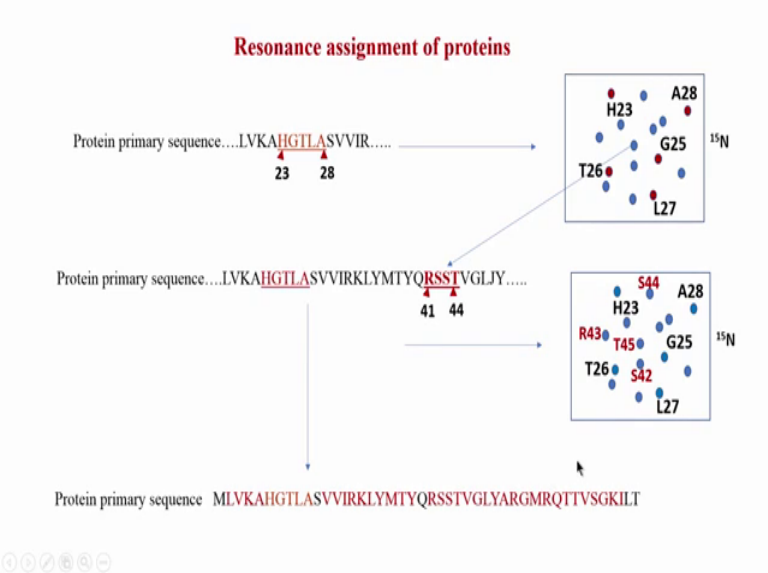
Similarly, this peak here which is shown here is A 28 why, because it corresponds to this i plus 2 which was alanine 28. So, like this for the all the set of 5 amino acids, I have we are now able to put them on the HSQC with their corresponding amino acid type that is G and number that is 25. So, this kind of label is now put on the HSQC, because now we have assigned or identified that each peak corresponds to which amino acid. This is for these 5 amino acids, you are not completed the job. There are lot of other blue colour peaks or that is unidentified peaks, which still have to be assigned.

So, we can start from now, once we know that this is assigned, these peaks are these red colour peaks are assigned. We can start from some other peak here, I could start from here or I could start from here. And again continue and do the same exercise. And that exercise may give me another set of 5 or 6 amino acid in the protein. So, this is how

basically we do, we start from one arbitrary some peak in the HSQC, go in the forward direction and backward direction up to 4, 3 or 4. So, we may identify 5, 6 or 7 amino acids.

And that we try to find amino acid type, and then put it on the sequence and see whether it matches. If it matches, then we have we can say that i whichever was arbitrary is now correctly some number belonging in the sequence. So, ones that set of 6 or 7 amide proton pairs are assigned like this in the HSQC. We start again from some other peak, and again continue we may end up with somewhere in some other part of the sequence.

(Refer Slide Time: 05:22)



So, this is shown here that in the first case, we saw that we got some five amino acids which we put it on the sequence sorry on the HSQC. Now, I can start with some arbitrary another peak, and then continue the assignment. I may probably find some smaller stretch somewhere else in the sequence, which is a 4 amino acid. So, this will give me rise to some other set of connections and that other set of connections will match to the sequence in this number.

So, now I can say I have identified two of them two sets, one stretch here, one stretch here and I can map it on the sequence now. So, you can see this is a new stretch 43, 44, 42, 45 which now I can put it on the sequence ok. So, I should write is as 41, 42, 43, 44, so this is a little this typo ok. So, now I can do like this and continue forever, I mean for

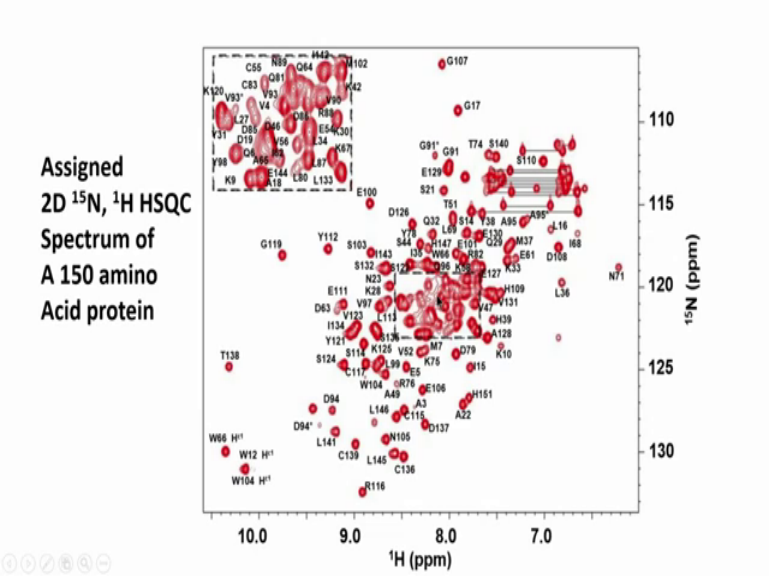
all the different parts of the sequence, because again there are many peaks I would have not assigned; so, I can continue those assignment.

And as a result in the end, I would have assigned many of the residues. This one show in red colour are those which are now assigned. So, how did we assign, we went by small small stretches one stretch here, one stretch here, one stretch here and get the but you can see here that we can never achieve 100 percent assignment. For example, the first amino acid in a protein normal, it does not get assigned, because it has missing peaks, first few amino acids can also like that.

Then somewhere in between the peaks may be missing, somewhere in between the peaks may be missing, so therefore we cannot continue the from here this side or come back from this side. So, we will not be able to assign. Similarly, somewhere here, so there could be a proline which is not shown in this sequence, but let us say I have a proline somewhere here that also will remain as a cause of break.

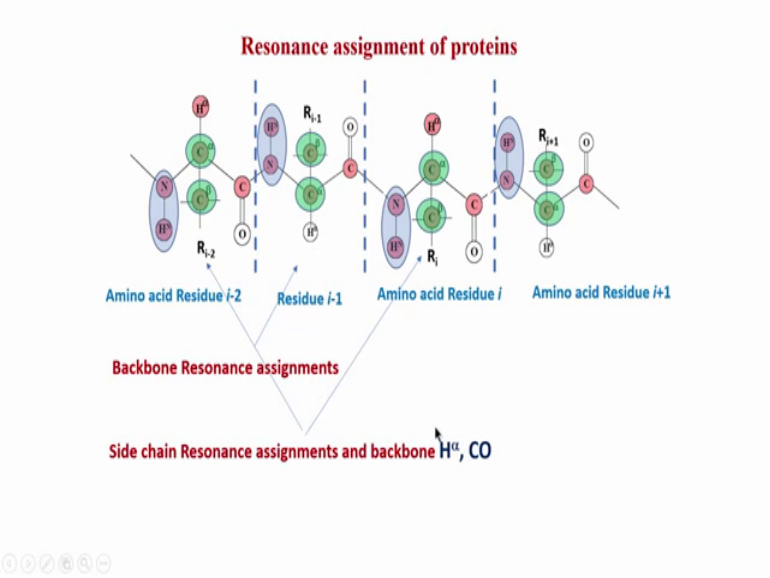
So, break in the assignments come because of missing peak and prolines, but typically we will never be able to achieve complete 100 percent assignment of a protein. There will be normally 90 to 95. And if you get 95 percent, it is pretty good for structure determination of proteins ok. So, this is how we basically assign, so we can see I will show you the next slide a real example of a real protein which is 140 150 amino acid protein.

(Refer Slide Time: 07:47)



And you can see that all the peaks are now labelled with the corresponding amino acid. So, this is it from a real protein and you can see how this is how practically is done ok. So, now the main thing what we have achieved with this is that we have assigned HSQC spectrum with the correct labels. So, this is very important, because this forms the first step, before we go to the side chain or next level of assignment.

(Refer Slide Time: 08:23)



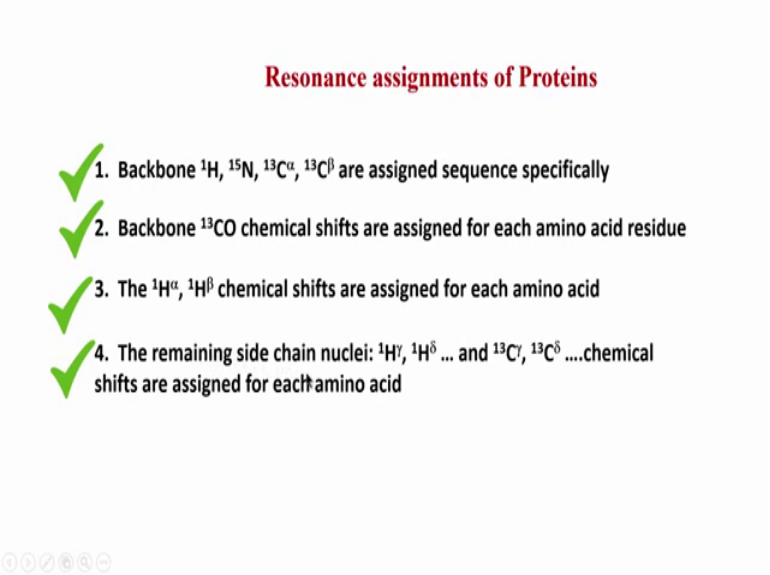
So, let us see what we have achieved with this, we have achieved the assignment of NH, because NH is what we in each of them is the NH for each amino acid that is what we have achieved. And what we have achieved is the C alpha and C beta assignment, because we use that for connectivity. So, remember this was shared between these two, this is shared between these two, so the connectivity helped us to assign.

So, therefore C alpha C beta also were assigned in the process, so that is what is basically shown here, we call this as backbone resonance assignments. Actually, strictly speaking C beta is not a backbone, but never the less we can you call it in a loose way that this is we have achieved backbone assignment. But, not the complete backbone, you see the CO is not assigned yet, H alpha is not assigned yet and the side chains are not assigned yet. So, this is basically only one-fourth of the job has been done, but that is the most important part of the assignment.

So, the next is we have to assign the side chain of atoms, carbon and proton for every amino acid and also we have to assign this to which are not assigned. There is H alpha

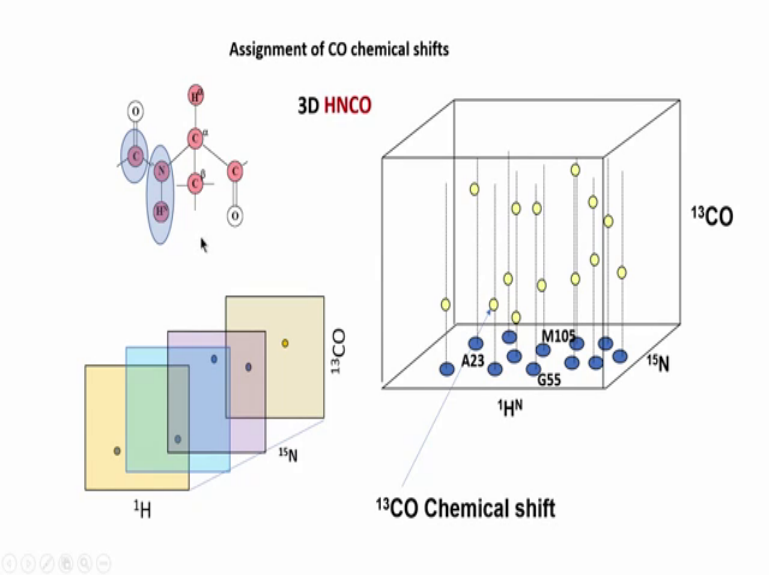
and CO. If you see here, we have only connected NH with C alpha and C beta, but we did not connect to CO and H alpha. So, how do we assign that let us see in the next few slides, how do we assign H alpha and CO chemical shift of proteins.

(Refer Slide Time: 09:53)



So, for H alpha ok, so this is basically saying that what one-fourth of the job is done. The first step which is a backbone these assignments of these atoms have been completed. So, now we have to go to the next two steps CO and H alpha and H beta. So, let us see how CO can be assigned using 3D NMR.

(Refer Slide Time: 10:13)



So, CO can be very easily assigned from HNCO. So, this is something we have been looking at earlier also that HNCO gives you the chemical shift correlation between the amide pair of one amino acid to the CO of the previous amino acid. So, based on that if I know the NH amino acid which number for example, alanine-23, glycine-25 and so on. If I know their NH values, I can simply go to the spectrum and find out the CO of the other amino acid the previous one.

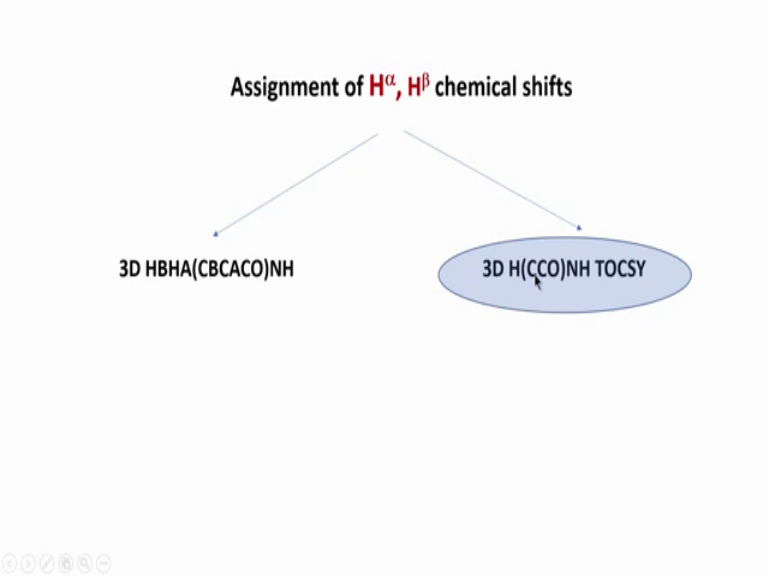
So, this is what basically shown here that for every NH chemical shifts, I can go in this direction third dimension. And just read the CO value, because CO value simply coming from this position in the third axis ok. So, this is basically what it shows here is that by simply going to the third dimension starting from HSQC, I go to the third dimension which is in the HNCO, I simply get the chemical shift directly from the CO dimension.

So, in this manner I can complete the CO assignment of each and every amino acid, which has been assigned here ok. So, for example, G55 will give me the G54 sorry the 54th amino acid, I do not know which is 54, but whichever is 54 I will get the i minus 1 for a 105 I get 104, for 106 I will get 105, for 24 I will get 23, for 23 I will get 22 and so on. So, you see even though it is giving me i minus 1, still I am able to get i minus 1 on the next one.

So, I will be able to get all the CO's, except of course the last amino acid, because the last amino acid will be there would not be anything beyond the last amino acid. So, last amino acid will not be i minus 1 to anybody. So, for the CO value of the last amino acid will not be possible with HNCO, but it is only one chemical shift value which I want to get, the remaining I will get.

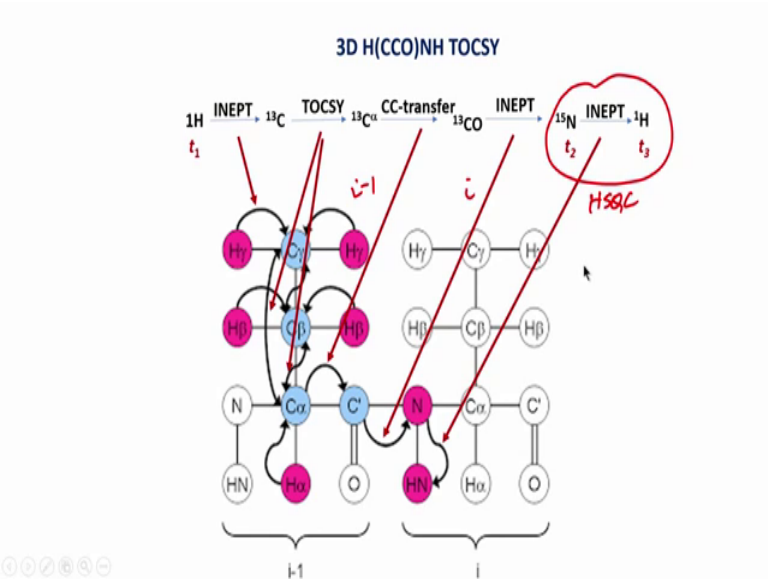
So, we normally we do not bother about getting 100 percent of chemical shifts, we try to maximize how much ever we can get. And if you get 95, 96, we are actually sufficiently strong to carry forward ok. So, this is how CO values are assigned in proteins. So, now let us look at this like third one, so, have achieved these two now. So, let us see how H alpha and H beta now, we can assign for each amino acid.

(Refer Slide Time: 12:49)



For assigning H alpha, and H beta, there are two 3D experiments we can use. So, 3D it is called 3D HB H beta H alpha CBCACO NH. So, this is a triple resonance experiment, so or we can use another triple resonance experiment H CCO NH TOCSY. So, what we will do in this course is we will focus on this only, we will not go into this experiment. You can find details of this in many test books, we will let us only look at this, because this is more popular and very sensitive experiment.

(Refer Slide Time: 13:21)



So, what do we do here? So, we do the following experiments. So, this is again a kind of steps which I will show similar to what we did in the analysis for the 3D triple resonance. So, in this H CCO NH TOCSY we start from the hydrogen amide of any hydrogen, it need not be amide. It can be these hydrogens which is shown in pink colour. So, all the hydrogens are excited by this one proton pulse first proton pulse.

And then after excitation, we evolve that proton. So, when we evolve, we get the chemical shift of that proton. So, we are capturing the chemical shift of the proton of all this protons ok, before we transfer to C 13. So, after getting the chemical shift revolution or while getting the chemical shift revolution, we transfer the magnetization to 13 C which is a directly attached 13 C. So, keep in mind, we are looking at the one bond. And that transfer is achieved through inept. So, this is shown here this type of a transfer of magnetization from one proton to carbon is called is through inept.

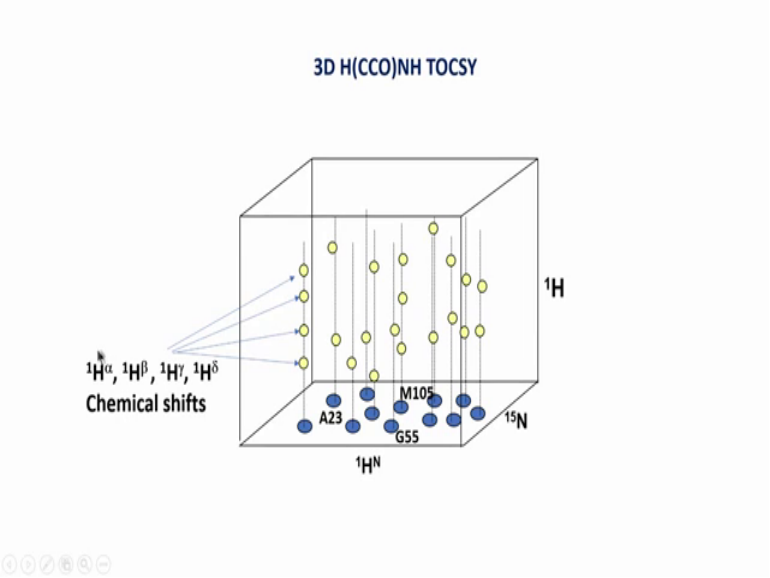
Once we have achieved the transfer to C gamma or any of these carbons, remember each of the hydrogen will give it to the respective carbon, with then apply a TOCSY, but this is now carbon TOCSY, it is not proton to proton TOCSY, it is carbon to carbon, but in a spin system carbon to carbon will transfer to each other. So, this is shown here you can see this lines here corresponds to the carbon transfer through TOCSY that means, all the carbons will transfer to each other. And finally, many of them will transfer to C alpha. So, our focus will be on C alpha.

So, whenever all the magnetization has come to C alpha by TOCSY from the carbon through the carbon, I then transfer to CO which is shown here. So, this transfer this transfer which is shown here. So, this is CC-transfer CO, this is through a carbon simple cosy transfer. Then from there I go by a inept to the next nitrogen to this nitrogen. So, remember this is i minus 1, so let me write this now this is residue i minus 1 and this is residue i ok.

So, I go to this next residue of nitrogen of next residue through inept ok. So, this is this inept here. And then finally, I go now during this after I transfer, I evolve the nitrogen, I capture the chemical shift of nitrogen, so that is that second dimension of the 3D experiment. Then I transfer it to hydrogen that is amide proton and that is the last step, and this is what is detected, during the detection period. And during the detection period that capturing of chemical shift happens and that is found the third dimension.

So, we have three-dimensions in this experiment. The first dimension gives me the hydrogen chemical shifts of all the hydrogens, and we get also H beta, we get H alpha, H beta, H gamma and so on. And then we go all the way to N 15 of the next amino acid, and this is a amide proton pair in a HSQC. So, you see this t 2, t 3, this is nothing but HSQC. So, I can actually circle this and right that this is an HSQC ok. So, this is basically how we do the experiment.

(Refer Slide Time: 17:04)



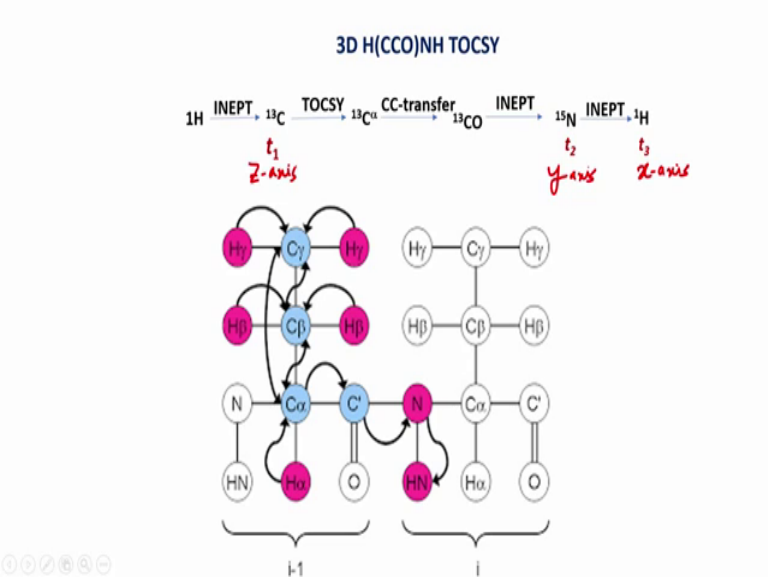
So, let us see further how the peak pattern will come so in a 3D H CCO NH. So, we start from this HSQC plane at HAQC peak. So, for every peak in this HSQC, if I go to the next dimension third dimension, again this is proton now. So, this is coming from this here the proton was the third dimension. Although it is t 1 here, if I call this as first dimension, this as second dimension, you can say z-axis. This is x, this is y, this is z, so, the z-axis is proton.

So, there is what is shown here and for each of this side chain. So, I can say that this is H alpha, this is H beta, H gamma, H delta, I can actually get the chemical shift value of the side chains, and H alpha and H beta backbone by going from for each peak in the HSQC. So, for example if this is alanine, I will get only H alpha and H beta. Remember alanine does not have H gamma, and H delta. Similarly, if there is methionine, I will get H alpha, H beta, H gamma like that. So, for different amino acid type, for example glycine I will get only H alpha, because glycine does not have H beta and so on. So, like this for each

amino acid, I will be able to get the chemical shift of the side chain starting from that base peak which is the HSQC.

And then going in the third dimension, but this is only the proton chemical shifts we have got, we have not got the carbon chemical shift ok. So, we complete the 3rd step of getting proton, but our 4th step is still remaining that is we need to still get the carbon chemical shift ok. So, we got the H alpha, H beta and also we got H gamma, H delta, but this is remaining.

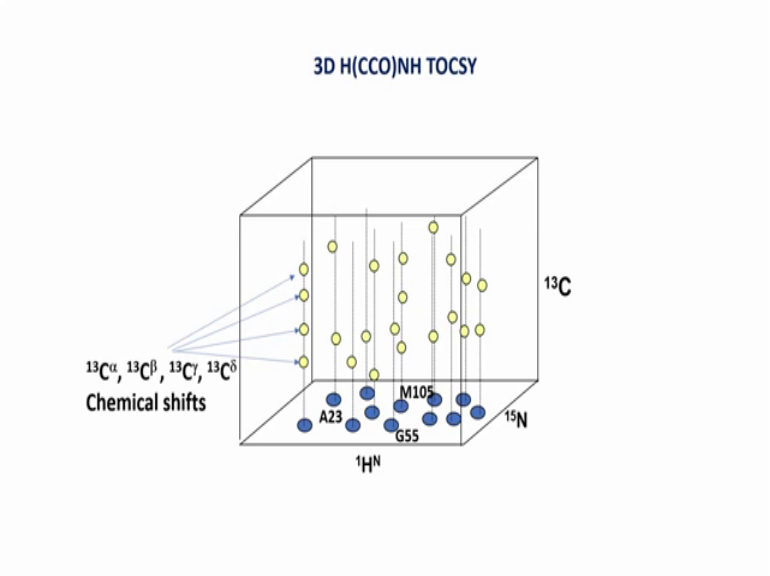
(Refer Slide Time: 18:51)



So, how do we get this one now the remaining ones again we have to use the same experiment HCCONH TOCSY, but here now you see here what we do is we do not frequency label here now we can frequency label this carbon. So, you can actually shift the frequency labelling means chemical shift evolution of from this hydrogen to this carbon and capture the chemical shift of this carbon.

And what are these carbon, this carbon is nothing but these carbons shown in blue colour. So, you see, by doing that I am able to now get the carbon to amide proton chemical shift correlation, so that is the trick here. The trick earlier was you frequency label or a chemical shift evolution of this hydrogen, now it change it to the carbon which is directly attached to that proton.

(Refer Slide Time: 19:35)



So, in the similar manner now for a given HSQC peak, I will be able to correlate now with the third dimension again I will get similar peak spectrum, but now these peaks are now C 13 because I have frequency labelled the C 13 during the z-axis. So, again remember I can write here this is x-axis, this is y-axis and this is z-axis. So, in the z-axis that that is along the z axis here, I am getting a set of chemical shift values, but these chemical shifts are now carbon chemical shifts. So, you can see now I have C 13 alpha beta gamma chemical shifts. So, like this I can complete all the steps now that is I have completed this step also and got the chemical shift values.

So, the trick here the main thing to understand is this particular experiment. So, in this experiment, again I repeat because this is a very important experiment, we go back. We start from we start from this hydrogen chemical shift, we got to C 13. And then from C 13, we went had a TOCSY to all the carbons. And we then transferred that magnetization to this CO through this CC transfer step. And this CC transfer is remember a simple cosy based a transfer. This is not inept here and then from this carbonyl, we did not frequency label.

So, remember we are not getting the chemical shift of carbonyl, it is only been used to go through the carbonyl, but not stopping at carbonyl. So, we are just passing through the carbonyl to the next atom which is nitrogen 15 and that is frequency labelled. And then from there we transferred to proton and that is a detection axis. So, this is a x-axis; this is

a y-axis and z-axis, this is how we got the peak pattern the chemical shifts, and that gave us the H alpha and H beta.

Now, the same thing we did for carbon. In carbon what we did was we moved this frequency labelling to this carbon. So, this from the proton to this carbon, that means, the chemical shift frequency evolution is now taking place here, it is actually taking place on this carbon. And because of that we are able to get the chemical shift of C 13 side chain all the side chain. And then through the same transfer is a same experiment, so the all the other things are remaining the same. You can see they transfer from C alpha, C, C gamma, C beta to all of them come to C alpha through TOCSY from there they are transferred to CO and that is through simple CC transfer and from CO we transfer to N and H.

So, we get the HSQC peak. Again for every HSQC peak we are correlating the carbon chemical shift here and that gives us the information of the carbon chemical shifts ok. So, this gets the chemical shift of carbon. So, this basically completes the resonance assignment of proteins. And you can see that although the task is divided into four different ways, each one is actually a very complicated task, but actually the most important and difficult is this part the backbone assignment.

This is something which takes a long time because after that we are using the information of the previous step to get the assignments. So, you see from for getting this chemical shift, we use this information available for a backbone. For getting this chemical shift we again use the information of the this resonances and for even for carbon.

So, the first step here is a most crucial step and typically the takes sometimes a few days if it is a small protein, or sometimes it may take a few months if it is a large protein. And lot of sample preparation NMR time experimental design goes into this part that is the backbone assignment. So, this is very important part. And this takes up almost 50 percent of the time compared to other assignments and because of missing peaks the life becomes more complicated.

And actually it is very difficult to assign manually these days for large proteins. So, lot of automated softwares are available for assignment. So, I would recommend you to go to online and just look at many of the softwares, which will assign the protein automatically

without any manual intervention. This actually depends on the protein to protein, but if we can give the HSQC peak and the C alpha and C beta of each amino acid i and i minus 1, then there are many software which work basically take this information and give out the sequence specific resonance assignment.

So, this completes a assignment part for the protein. In the next class we will move on to structure determination of the protein. So, we will start from looking at how this chemical shift will give us first the secondary structure that is alpha helix beta sheet information. From there we can go to 3D NOESY and get the distance information and from there we can then calculate the three-dimensional structure of the protein, so that will be the next part of the course.