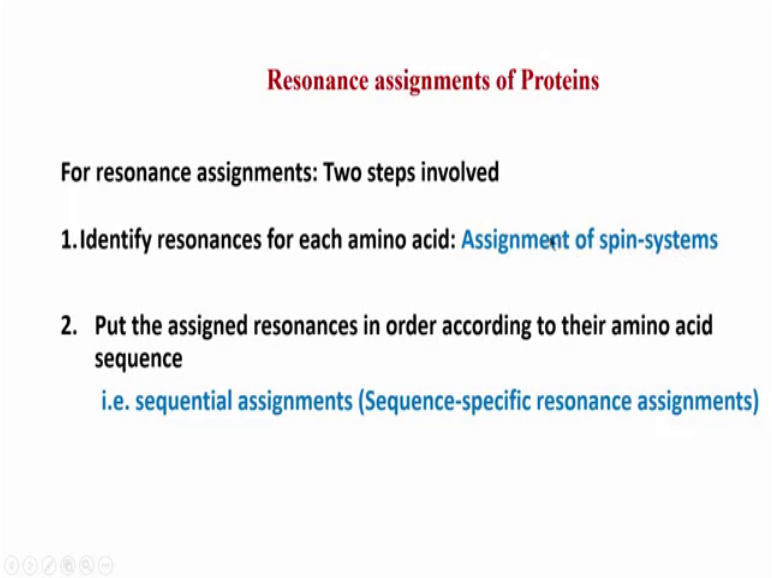
Multidimensional NMR Spectroscopy for Structural Studies of Biomolecules

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

Lecture – 29
Resonance assignments of Proteins – Part II

So, in the last class, we started looking at Resonance assignment of Proteins. As I mentioned these are the 2 steps, which are involved in any protein resonance assignment number 1 you have to identify the spin systems. So, this is basically called as assignment of spin system or identification of spin systems.

(Refer Slide Time: 00:43)



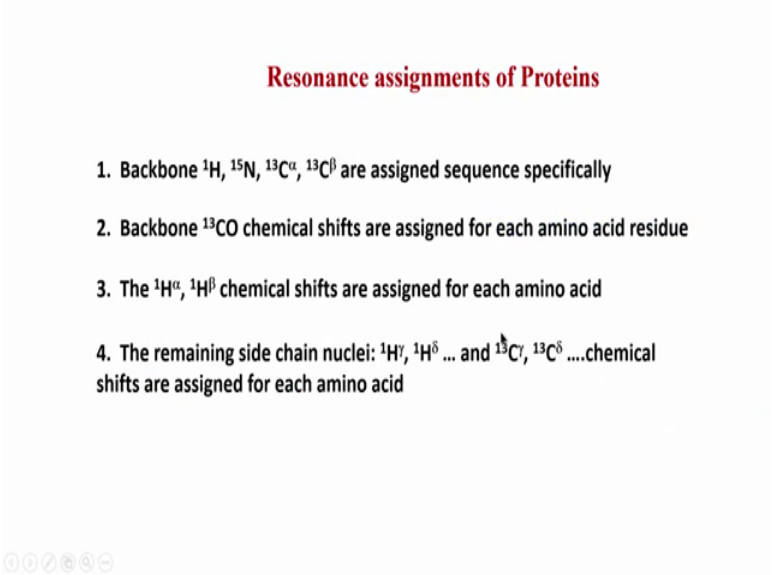
So, how do we identify a spin system? We identify based on the pattern of chemical shifts in that spin system like we saw in TOCSY, similarly if you have a chemical carbon chemical shifts you can use that to identify that the spin system is there or not.

So, basically here it means that suppose I have 4 lysine, 6 arginine, 7 aspartic acid and so, on in my protein I need to figure out whether really those number of amino acids are there in my spectrum or not. So, that is the concept of spin system identification, where we try to find out what are the amino acids which are supposed to be present are they really present in the spectrum or not. Once we identify the spin system, the next step is to put the spin systems in the right order. So, this is important here remember we have only

identified at this stage, that there are 4 lysines, there are 4 arginines, there are few aspartic acid and so, on.

But we are not arrange them in the right order and how do you arrange them in the right order is the whole idea of sequence specific resonance assignment which we will see now in this class, but the whole idea is that we need to know the sequence of the protein so, that we know how to order them because if you do not know the amino acid sequence, we will not be able to put them in the right order. So, therefore, first we have identify and then put them in this correct order.

(Refer Slide Time: 02:12)



So, now this assignment of proteins actually typically happens in 4 different stages or 4 different steps, the first step is that we assign for every amino acid the proton, nitrogen it is amide proton or a basically the backbone amide proton N 15 C 13 alpha and 13 beta assign sequence specifically for every amino acid in the protein.

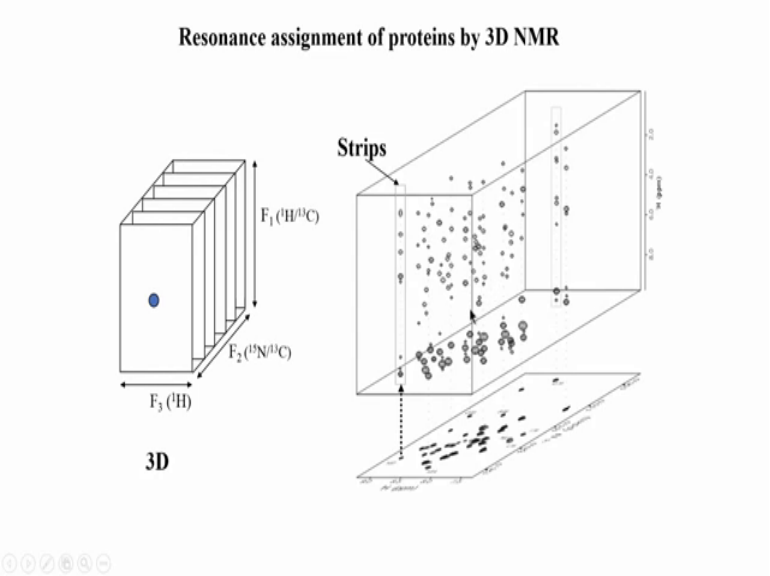
The next step is after we have assigned this, we assign the carbonyl chemical shifts because this once you know which amino acid has which carbon which proton amide alpha, we can then go to carbonyl chemical shifts and that are assigned again for every amino acid. So, once this is accomplished then we go to the side chains for side chain we need to first identify the backbone H alpha and the side chain H beta chemical shift of which amino acid. So, this step is important because this is our entrance into the side

chain. So, up to here the first 2 only concerns the backbone chemical shifts, but now we are entering the side chains.

So, for entering the side chains you need to have first the alpha and the beta protons of every amino acid. Once we have that then we can assign the rest of the side chain that is a gamma delta epsilon protons and similarly gamma delta epsilon carbons. So, that is the last step and after that we can consider that the assignment is complete.

So, these are the 4 steps. So, let us see how we go one by one in proteins through each of them.

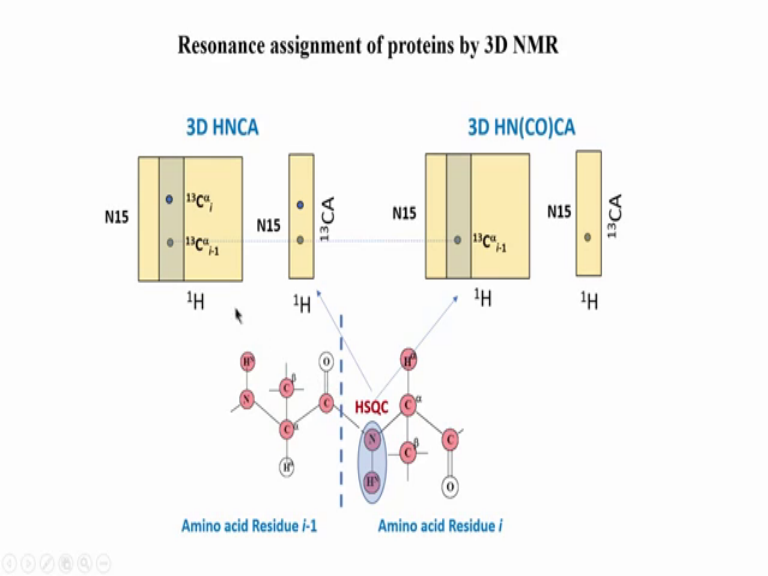
(Refer Slide Time: 03:52)



So, we will first start with the backbone chemical shifts. Now in a given 3D NMR spectrum we have to as we have been discussing, we have to assign them inform I mean you have to analyse them in the form of strip plots. So, this is example of a strip given for a 3D experiment. So, typically what we do is we start from a base that is projection. So, the base projection typically is N 15 HSQC where each amino acid here is one peak.

So, what we do next is, we take this peak and then go to the 3D and then in the 3D we look at the chemical shifts in the third axis. So, this is how we will see how we will analyse and assign the protein.

(Refer Slide Time: 04:33)

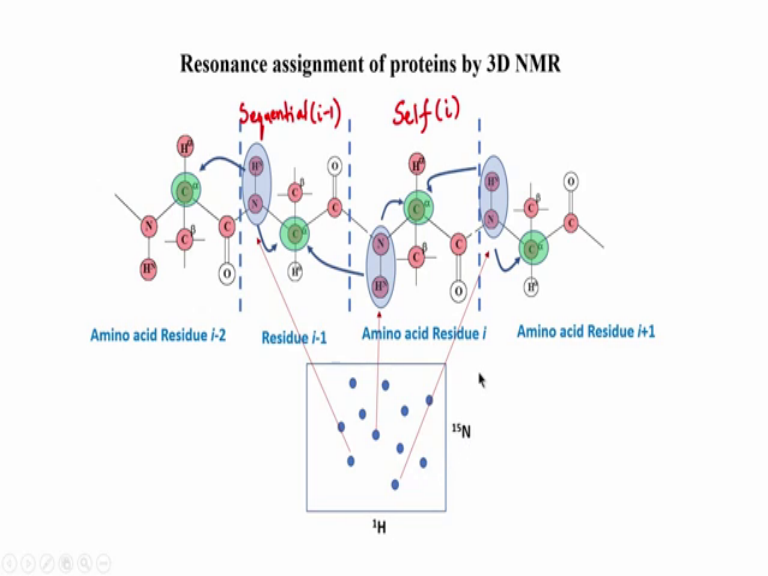


So, now we looked at a very simple experiments a triple resonance earlier that is HNCA and HNCOCA. So, we will use this experiment as a illustration of how assignment is carried out to start with, then we will move on to more difficult or complicated experiments.

So, first is there HNCA what is the information do we get? For every amide nitrogen and proton chemical shift that is for every residue very amino acid which has this chemical shifts in the third access that is the third dimension we get the C alpha of the residue i minus 1 and the residue i for the same amide proton chemical shift. So, that is shown here in the strip format.

Now, how do we know which is i minus 1 when which is i that comes that information comes from HNCOCA. So, this is shown here schematically. So, we have NH and what we have done is we have correlated the HSQC. So, this is each strip here this strip here and this strip here is basically an HSQC pair because nitrogen and amide correspond to one pair of this residue amino acid. Now for each amino acid in this experiment we are getting i and i minus 1 for each pair of nitrogen proton chemical shifts we are getting i and i minus 1 it is where correlating to this and this ok.

(Refer Slide Time: 05:58)



So, basically that is what we saw in this here, your correlating in HNCA the chemical shift of the nitrogen and proton with the C alpha of i which is i minus 1 residue and C alpha of ith residue.

So, this kind of a connection that is where you connect to both plus is own chemical shift and my the previous one is very important in assignments. This is something we saw in the case of TOCSY as well in the NOESY or a peptide which I showed in the last class and if you take a peptide like alanine glycine alanine glycine, there also we saw that in the NOESY what helped us to assign was the inter amino acid connectivity. Similarly here this is the inter amino acid between the 2 amino acid connectivity, which is very important for assign without this connectivity we cannot assign because only having the self has no big value for assignments we will see that shortly now.

So, now say for example, for every peak in HSQC, I will basically get the correlations to i and i minus 1. So, I can move this to the previous amino acid to i minus 2 i minus 3 and so, on I can also move this side i plus 1 i plus 2 and so, on how do I do that? I will do by joining or connecting this C alpha now with the next NH here because remember this C alpha can be connected to the next NH, because it will be the i minus 1 for that amino acid. Similarly this C alpha we can be connected to this NH because it is the self. So, member the nomenclature we will call this as self and this as sequential ok. So, let me

write down this for to be ensure. So, this will call it as self that is i and sequential which is i minus 1. So, this is the nomenclature used.

Now, I can go to for example, this peak let us say any arbitrary peak in this spectrum look at be this particular amide. So, that gives me correlation to this now I do not know which amino acid is this I have just started with some arbitrary amide proton chemical shift in the HSQC. Now if I go here, I can now see which this peak which peak here which of this peak here corresponds to this how will I know that? That peak whichever is that peak we will give me correlations to this C alpha ok. So, this C alpha is basically now shared between this and this because for him for this particular case it is i and for this particular case i minus 1.

So, they share this C alpha between i and i minus I i and i minus 1. So, I can therefore, have a common connectivity here. So, basically this NH once I find that it could be some peak here, let us say it is this peak here which need not be nearby it can we k here. Once I have found this peak now I have found the i minus 1 chemical shift, but again I do not know which I number we are looking at just any i. So, I am not saying that we already know which what i is and what number is i, we are just saying some arbitrary I and i minus 1. Now once I get i minus 1 it will give me correlation to i minus 2 because that will be the HNCA correlation.

Now, again if I go further I can get i minus 2 NH by correlating it to this C alpha which is again from HNCA. So, I basically need some peak which corresponds to this NH in the HSQC, which will now give me correlation to this. We will see this again as we go further similarly I can go on this side on this side I can again find NH pair, this is one pair which will be some peak here which will give me correlation to i and also i minus 1. So, you see this is the common connectivity here now v it is shared between i and i plus 1.

So, I can find now some NH somewhere here which could be anywhere, let us say it is this NH and that NH turns out to be such that it has a connectivity to this C alpha which was connected to this NH. So, these 2 now are connected to a common C alpha and that could be i plus 1 therefore, and this is now i plus 1. Now this further will connect to i plus 2 and then i plus 2 will go to i plus 3 and so, on. So, you see what we have done

here, we have we have connected inter amino acids 2 amino acids by this common NH. So, in other words one C alpha is shared by 2 NHs you can look at it that way also ok.

So, this is basically how we will see now we will go further and see how assignments are done by connecting one NH to another NH through C alpha. So, this is the basic point here we have connecting 2 neighbouring amide proton pairs, through they are common C alpha to which they are correlated ok. So, we will get the complete stretch will be able to connect a stretch of amino acid from i minus 2 to i minus 4 or so, on to i plus 1 to i plus 2 and so, on.

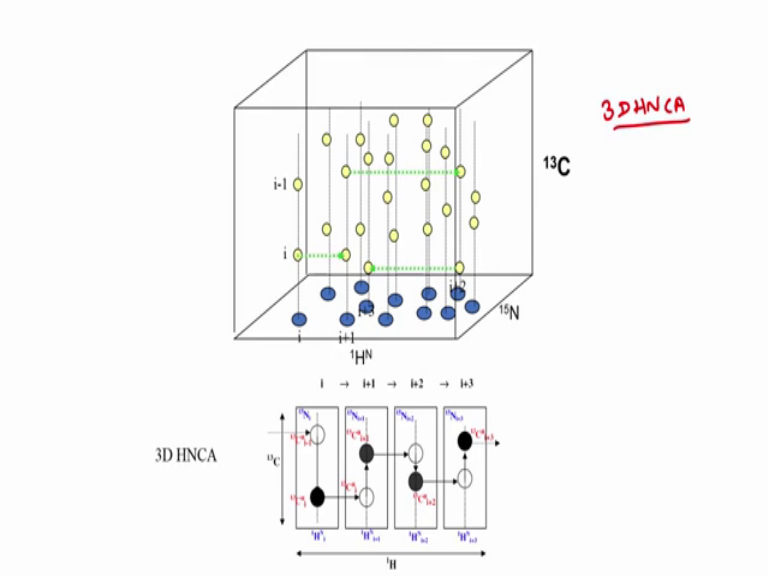
Remember this connectivity come does not go from minus 1 and 2 other end of the protein, it will only go to 5 or 6 amino acids because after 4 or 5 roughly you may encounter some breaks and what could be those breaks? In the brakes could come because there is a missing peak in the HSQC; remember I said in HSQC you never get 100 percent peaks for a protein. So, if you expect about 100 peaks for a protein you may probably get probably about 80 90 percent.

So, those missing peaks are basically those peaks which are simply do not appear in the spectrum. So, if they do not appear there is no way for us to identify that amino acid. So, for example, it may happened that this NH is completely missing in the spectrum, the it does not mean this is missing it will mean only this is missing. This could be missing because of various reasons like exchangeable fast exchange and so, on. So, it could be possible that this NH is missing and therefore, we are not able to continue further this side because there is no information on this.

Similarly, here we go upto i plus 2 and i plus 3 and then again we may find a missing peak and then again we would not be able to proceed further. So, we would have assigned from i minus 2 to probably up to i plus 3 which is about 5 6 amino acids ok. So, this is how we go we go in batches of 5 and 6 amino acids along the protein sequence and try to complete the assignments.

Now, one more thing is another [bre/break]- reason break could happen is because of proline. Proline does not have NH proton proline is a special amino acid which does not contain NH. So, if that is the case there when I go to the next amino acid I will not get any peak in the HSQC, because there is no and the proline cannot have any peak in this HSQC there is no NH pair in proline and therefore, that constitutes a break.

(Refer Slide Time: 13: 27)



So, now let us continue and see how this assignment is done. So, we can see this is the basal plane projection of a 3D HNCA. So, in this what you are seeing here is a HSQC spectrum of some protein. This is a very schematic drawing and can see each of this peak corresponds to one amino acid; now on this plane you get the 3D of HNCA. So, which we will look something like this ok. So, let me write down we are looking specifically at 3D HNCA ok.

Now, in 3D HNCA for every NH peak we get i and i minus 1 again remember we do not know which is i which is i minus one. So, we need HNCOCA, but let us say we have HNCOCA which is not shown here, but based on that let us say we have identified that this is self in this particular case and this is sequential like that for every case. So, we will not we will assume that we know which is i, which is i minus 1 because we have used HNCOCA which is not shown here.

So, let us start from some NH any arbitrary NH here. So, let us say I start from here, let us call it as I when I start from here I will get 2 peaks which is i and i minus 1 in the C alpha; now I will search in this whole 3D cube where this I now appears as i minus 1 to i plus 1. So, for i plus 1 something is i minus 1 is nothing, but i. So, the sequential of i plus 1 is i. So, this i and i plus 1 they share this common chemical shift C alpha i. This is what we saw in the previous slide in the previous slide let me go backwards a little bit you saw in the previous slide. So, if you look here, the i plus 1 NH and i NH they share or they

correlate to i C alpha. Similarly I NH and i minus 1 NH correlate to i minus 1 C alpha. So, they have a common connection. So, same thing is shown here from i go to i and i plus 1 NH peaks they share a common C alpha i.

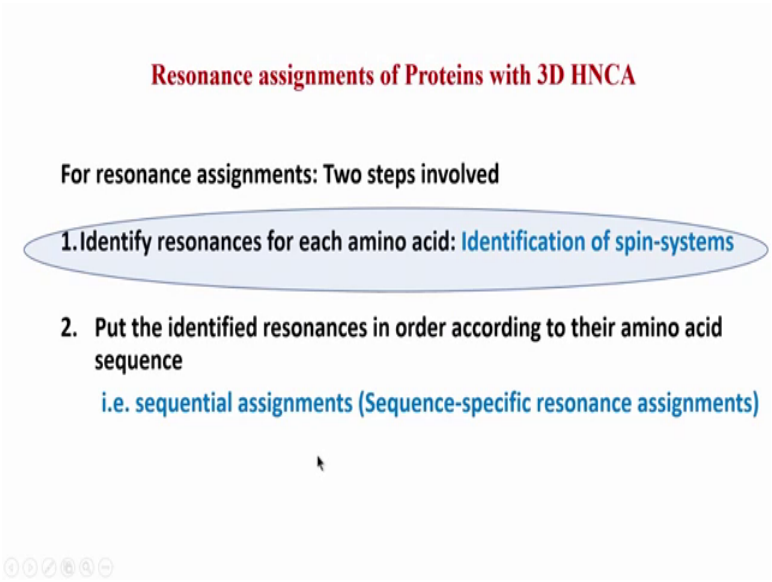
Now, in this line if I go to the remain this is i i plus this is i. So, this will become i plus 1 because i plus 1 and i is what you will see on this peak. Now this peak will be shared with i plus 2 like we saw in the previous slide. Similarly this i plus 2 will be shared with i plus 3. So, like this we can basically go along the chain by simply connecting the common C alpha chemical shifts ok.

So, that is how basically assignment is done, but as I said after this you may probably get a break, you may not be able to see i plus 4 because i plus 4 may be missing in this spectrum become missing because of either peaks are weak or that there is a proline i plus 4 could be proline. So, we do not know what is i here we started from some arbitrary i and then worked along the backbone. Similarly I can go to i minus 1 i.

So, I can start from this, this will be shared with some other i minus 1 and so, on. So, you can see this is how the connectivity is shown in this picture. So, we can see that for a given in a this is a strip plot format, in a given for a given NH and amide proton pair I will get i and i minus 1 here we have shown filled circle here and open circle for i minus 1. Remember this distinction comes from HNCOCA which is not shown here, but I am assuming that we have recorded an HNCOCA, you have already looked at it and then based on that we know which is i and which is i minus 1.

Now, this i is a shared with i plus 1 it is what we saw in the previous slide and this i plus 1 this i is shared. So, this will be this is i this is the other peak in this line becomes i plus 1 because it wear on the amide of i plus 1. This i plus 1 now it shared with the i plus 2 and i plus 3 2 is shared with i plus 3 and so, on. So, this is how we walk along the backbone this is called sequential backbone walk, it is a very common terminology which is used in protein assignment which is will be talking about more.

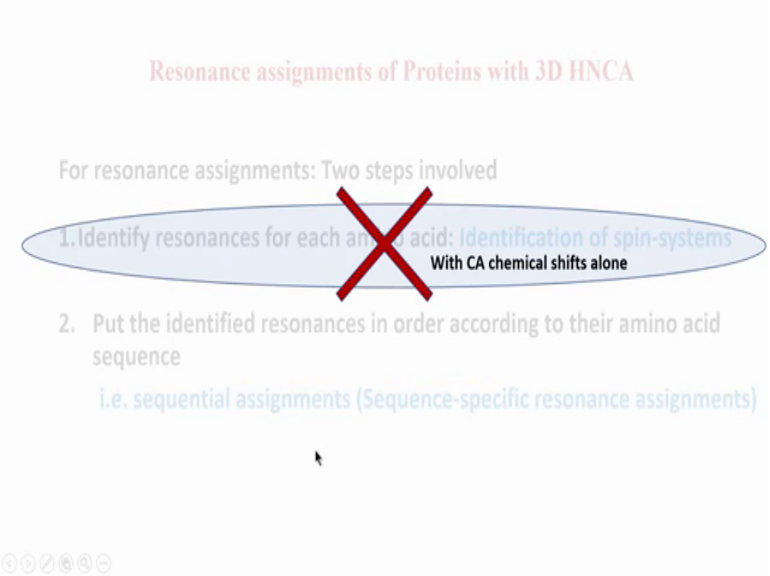
(Refer Slide Time: 18:03)



So, now if you see what we have achieved in this basically this whole procedure is that, we have achieved actually the second step we have identified the sequence in order ok, but we have not got the spin systems why? Because look at it here I am only talking about some i, I do not know what is this I what spin system is this i is it a glycine, it could is the could this be a lysine, could this be an arginine, could this be an aspartic acid we do not know any of that information from this exercise.

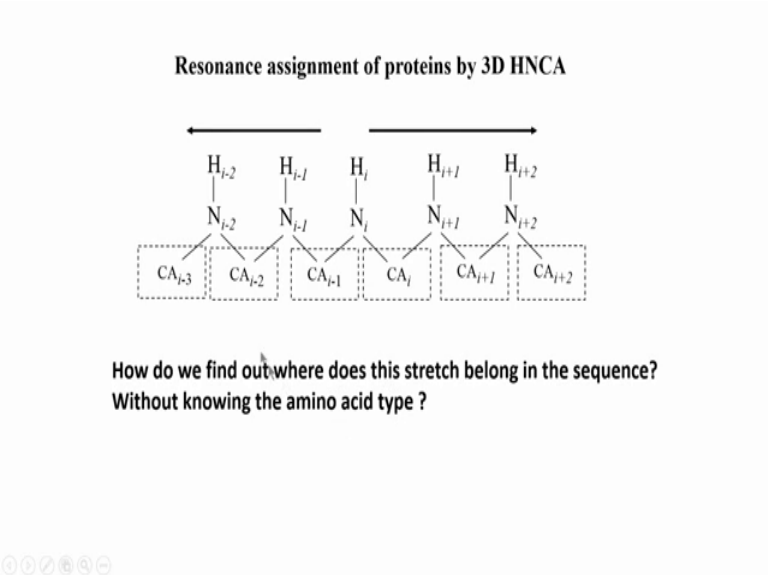
So, this whole exercise is only giving us connectivity pattern between amino acids. So, this is one problem with using HNCA only, and HNCA and HNCA you see course is that we do not get the amino acid type information. We come do not come to know which amino acid is which peak because we just start from a i we assume some peak C alpha, but C alpha has no information of what amino acid it is many C alphas of many amino acids they all local i it means they come in the same spectral region.

(Refer Slide Time: 19:15)



So, therefore, I need to achieve this assignment. So, this is what is written here that with C alpha chemical shifts alone you cannot achieve the identification of spin systems. So, therefore, we HNCA is not so, useful for assignment in the sense it does not help us to get this part.

(Refer Slide Time: 19:37)

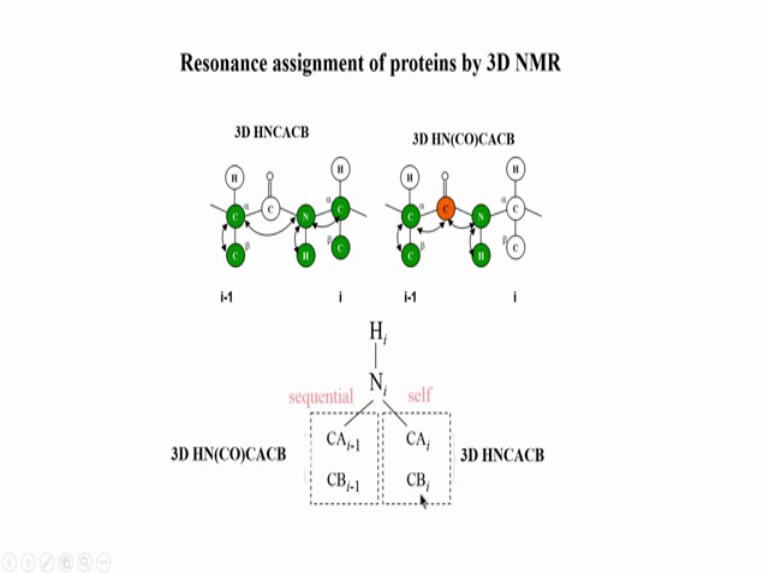


So, how do we then improve this? This can be improved with using HNCACB. So, this is again the same thing is shown here that for every amide proton in HNCAs we are sharing the C A alpha. For example, H I and n I which is one peak in HSQC and this is

another peak they share the common Ci C alpha i similarly i plus 1 is shared between these two. So, here i minus 2 is shared between these 2 and so, on.

But the problem is how do we find out where does this stretch belong in the sequence. So, we have connected a stretch from i minus 3 to i plus 2, but we have no clue no information on which amino acid type is this like for example, is this arginine, is this glycine, is this isoleucine some kind of amino acid information is not there in this assignment in this connectivity therefore, we need to know what amino acid it is, before we can say that we have completed our assignment.

(Refer Slide Time: 20:37)



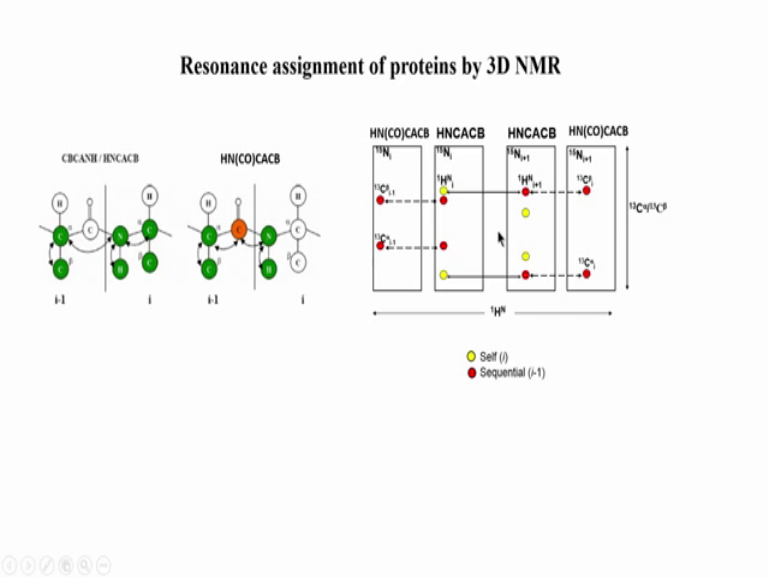
For that we need to go to another experiment that is called HNCACB and HNCOCACB. So, this is what happens in HNCACB and HNCOCACB we have seen thus here an HNCACB, the amide pair a nitrogen proton pair correlates means connects to C alpha and C beta y minus 1 and C alpha and C beta of i. So, we see the extra things happening is C beta, the C beta here is and C beta here and this information was not present in hnca. So, when we were doing HNCA and HNCOCA we only went up to C alpha, but not to C beta.

So, here we are going further one step to C beta. So, in HNCOCACB we get information of C alpha C beta of i minus 1 remember this residue is i minus 1. So, like similar to HNCA and HNCOCA, we can now say like this that for every given amide proton pair

that is every residue i, the self-chemical shifts are C alpha C beta and I get sequential chemical shift these 2 ok.

So, this now I have took extra one extra chemical shift, which I can use for connectivity. So, this is why the reason these experiments are important is because C beta carries a lot of information of amino acid type. We will see that shortly and therefore, we can identify the spin system which was missing in the HNCA.

(Refer Slide Time: 22:12)



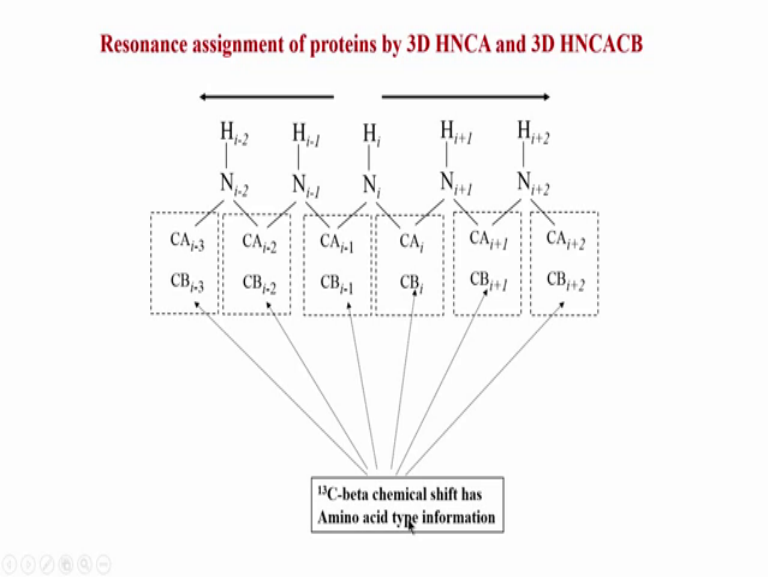
So, how does hncs to work it is similar to HNCA matching? So, you can see here in the schematic that in a strip plot from suppose I start from some arbitrary i o arbitrary means any i there I will expect to get 4 peaks in HNCACB. This is something you might be now familiar with from the previous part where we looked at 3D NMR and there we saw that you get 4 peaks corresponding to C alpha and C beta.

So, now we can see that these 2 which are now self and sequential are coloured separately. So, shown here yellow is self and sequential is red. So, these 2 now sequential means which are i minus 1 can be identified based on these spectrum and whereas, the self is different. Now this self will be shared common with a plus one this is something we saw in the previous slides we get i plus 1 amino acid share the C alpha i and C beta i. Now these 2 become the sequential for this amino acid ok. So, far this amino acid now that is sequential it is i is sequential and i plus 1 is shown here.

Now, i plus 1 if you go further we will find in another match to i plus 2 for i plus 2 these 2 become sequential. So, we see the self is becoming sequential when you go this side and the sequential will become self when you this side. So, self and sequential are interchangeably changing because whatever is self for this amino acid will be sequential for the next amino acid. So, this way we can connect now 2 chemical shifts and continue the assignment. So, what is assignment we are doing you are connect to find a connectivity a stretch of amino acid which are connected.

So, you seen the advantage of this part is that we not only get this, we also get this information we are getting both that information. That is assignment of identification of spin system also comes from this C beta and the connectivity also comes from the C beta.

(Refer Slide Time: 24:22)



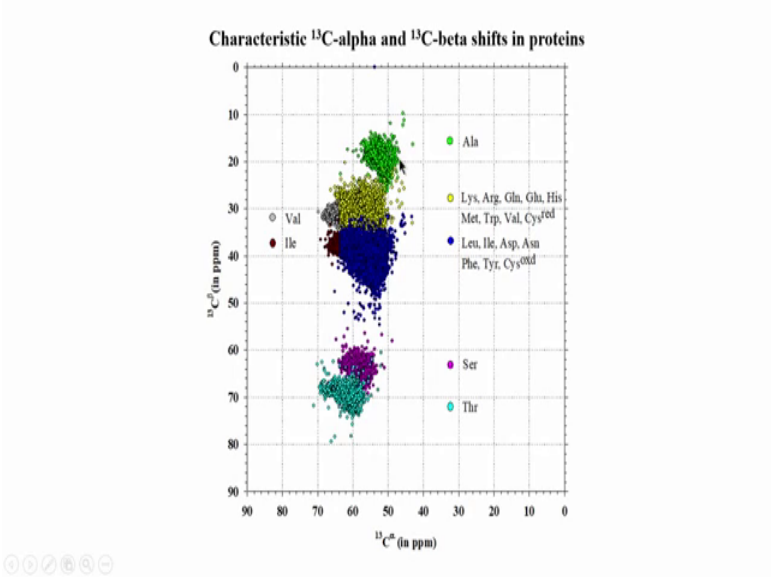
So, how does the spin system come? Let us see in this particular slide. So, we can see here that what we have done we have i minus 2. So, we have connect a we started from an arbitrary H and N i in a protein some peak in HSQC and we went this way and this side. So, when you go i plus this side, you will see we share this 2 chemical shifts between the neighbouring amino acids. So, very important point is neighbours this neighbour share a common correlation to these chemical shift.

Similarly, if I go to i minus 1 side, I will be sharing this for these 2 C i minus 1 C beta i minus 1 and similarly between i minus 1 and i minus 2, they will be sharing this

chemical shift in common this CA i minus 2 and CB i minus 2. So, like this I am forming a connectivity pattern or connection between neighbouring amino acid. But again remember we have only connected we have not find out which amino acid it is simply form connection.

So, how do I find out now which amino acid is this which is this which is this and so, on it is based on this C beta value? So, written here C beta chemical shifts has a very good information of amino acid type.

(Refer Slide Time: 25:48)

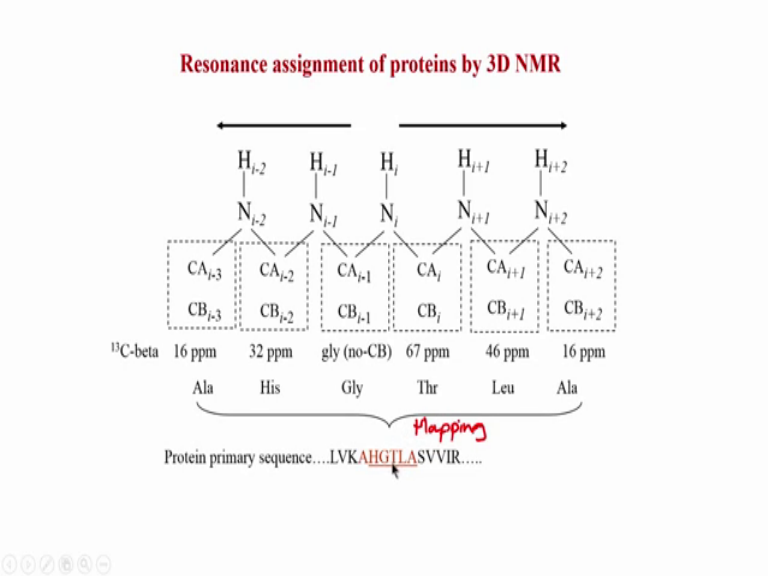


So, let me show you a statistical analysis which was done long ago, and this is the result of that analysis. Here you can see this is the 2 dimensional plot with C alpha on the C axis and C beta on the y axis. So, we can see that amino acids cluster into different groups based on their C alpha and C beta.

So, for example, let us start from here, alanine in is very distinct chemical shift as far as C beta is concerned. Look at this axis which is on the vertical axis C beta comes in this zone this was based on thousands of chemical shift from database. So, we can see this is the range of chemical shift for alanine. Similarly these groups of amino acid lysine arginine glutamine so, on they come in this category they come in this zone ok. Similarly this zone blue colour belongs to this amino acids and a few of this valine and isoleucine, they are actually very unusual they there is C alpha comes this is now looking C alpha come somewhere around 65 ppm ok. So, this is special case for valine and isoleucine.

Now, if you further go down in C beta chemical shift. So, you see there is a gap here there is no C beta here mostly it is C alpha, but C beta there is that will starts around 60 ppm for serine and then further down for threonine. So, serine threonine also can be separated. So, we can see here how nicely the C beta chemical shift uses information of what amino acid type it could. Of course, here yellow means all of them could be here. So, it does not tell us between these amino acid similarly blue colour is full set of these amino acids, we cannot distinguish within this set, but still we are able to get some information which was not available from C alpha. So, a combination of C alpha and C beta is helping us to achieve that information.

(Refer Slide Time: 27:47)



So, once I have that information, I can probably predict suppose the C beta of this case was 16 ppm I can say that could be alanine. Because that is what I saw from the previous chart. If it is 32 ppm probably it is histedine based on database similarly here it may be happen that this C beta is zero means there is no chemical shift of C beta for this amino acid or this amino acid, which means that it is a glycine because glycine does not have C beta glicine is only up to C alpha. So, we could get a glycine similarly here it could be that it is 67 ppm and based on the theory the that is could be threonine like that.

So, based on this database and analysis statistical analysis, we can predict or guess which amino acid type it is based on the C beta ok. So, once we have guessed or predicted that, you can now see that I have connected a stretch of 6 amino acids we have connected a

stretch which lie one after the other; that means, if this is i minus 2 this is i 3 this is i minus 2 i minus 1 so, on. So, they are continuous not only they are continuous this is the specific set of amino acid.

Now, can we try to find that in the sequence? So, I am trying to find these amino acid in the sequence in this order only now order is coming from here. So, I have connected in the right order after connecting them in the right order I have found the chemical shift type, amino acid type of these spin systems and now I am trying to find where they can belong in the primary sequence. So, remember again we have given this sequence we already know the sequence. So, we are trying to map. So, this is called mapping its a very important word in assignments mapping.

So, we are trying to map the stretch of amino acid that we have assigned on the sequence and then if there this correct assignment if this what I have done is correct then I should be able to find that sequence in my protein sequence also. So, my protein primary sequence should contain this stretch of amino acids, which I have connected and assigned. If it is found then I can say that yes my assignment is over for these amino acids.

Again remember we are not talking about the remaining ones only this 4 or 5 or 6 or 6 amino acids we have assigned now. So, we will now see how we can continue this assignment in the next class, and we will look at how we can complete the full protein chain starting from one end to the other end.