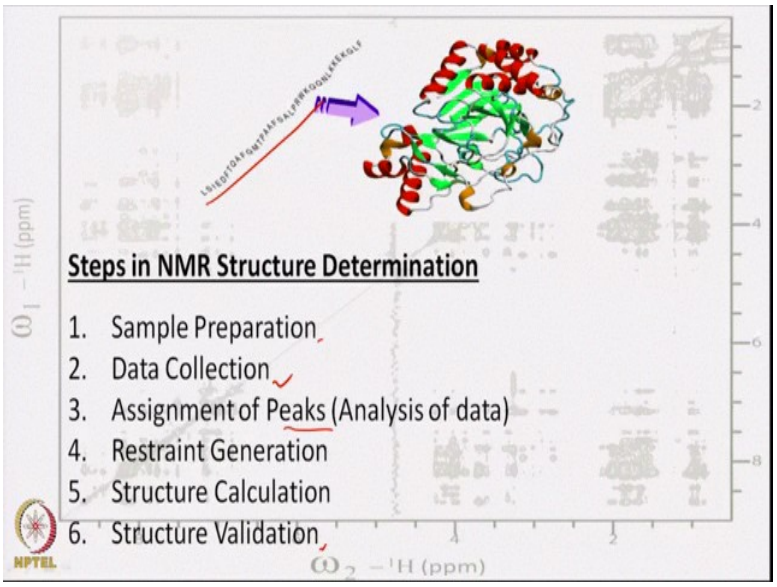
NMR Spectroscopy for Chemists and Biologists
Professor Ashutosh Kumar
Professor Ramkrishna Hosur
Department of Biosciences & Bioengineering
Indian Institute of Technology Bombay
Lecture 56 - Structure determination of peptides by NMR - I

Hello everyone! So welcome to today's lecture. Now we are almost approaching towards end of this course. So what I thought in last five lectures, I will be giving you some glimpse of NMR spectroscopy in structural biology, how to determine the peptide structure by NMR. Then I would like to see how we can understand the protein-protein interaction by NMR spectroscopy. And finally, I will give some glimpse how to get the shape and size of this, like size of the protein molecule or peptide molecule by use of NMR spectroscopy. So this will be topics that I will be covering in next five lectures.

(Refer Slide Time: 1:10)



Now let us start first today's lecture. So what we want to do today is if you have a primary sequence which is given in terms of say amino acid, can we read this amino acid sequence here? Use this information now of amino acid sequence. We record various sorts of NMR spectrum and can be translate those information into three-dimensional structure. So that is this total goal. As you know we have three techniques that are primarily used in structural biology for getting the atomic resolution structure of any biomolecules.

One is crystallography where you need to crystallise the protein. So like what you do, concentrate protein, then put it in the crystallisation condition. If you get a crystal this crystal must diffract and after diffraction you can determine the structure of the protein. Then one is

NMR spectroscopy where you need to solubilise the protein and you record to vary, you need to record various sorts of NMR spectrum where you do the like, do whatever today and we are going to discuss. And then at the end using all those NMR parameter we want to determine the three-dimensional structure.

The third one which is emerging is called cryo-electron microscopy where you do not need to crystallise but still you just sputter that protein sample on plate or seat and then you like put the electron beam and that created various deflection pattern and that is used for structure determination. So NMR is a technique which gives structure and dynamics as you have studied the relaxation, relaxation essentially gives the dynamics. In a separate course we will be later, maybe sometime later, how to use NMR extensively for getting the structure and dynamics of a protein. But today I will just give you glimpse what are the steps involved in NMR spectroscopy going all the way from the primary sequence of a protein to the three-dimensional structure.

So steps in the NMR structure determination is first one, 'sample preparation'. How to prepare a clean sample which is amenable for NMR? Whether it is peptide or it is protein. So for today's lecture let us just concentrate to peptide. Now you have to synthesize the peptide either by bacterial expression or by chemical synthesis. Then you need to purify this peptide so that we have maximum purity of this peptide, more than 98 percent or 99 percent purity needs to be to have that purity of that peptide.

Next step will be because we are doing solution state NMR, so next will be solubilizing that peptide in appropriate buffer; many peptides are not soluble in water. So you have to choose an appropriate buffer to get the structure. So mostly it should be, if you want to get a solution structure, it has to be soluble and therefore we need to choose an appropriate buffer for sample preparation.

One sample is prepared, then next step will be data collection. So we need to collect the various sets of NMR data as we had discussed in past. We have to, we need a data collection for sequential resonance assignment, like we need to do, record data for like COSY or TOCSY or double-filtered-quantum COSY, if you have COSY. Then we need to record the X-nuclei correlation in spectra like ^{13}C HSQC or ^{15}N HSQC, that I am going to discuss later. So that will be data collection.

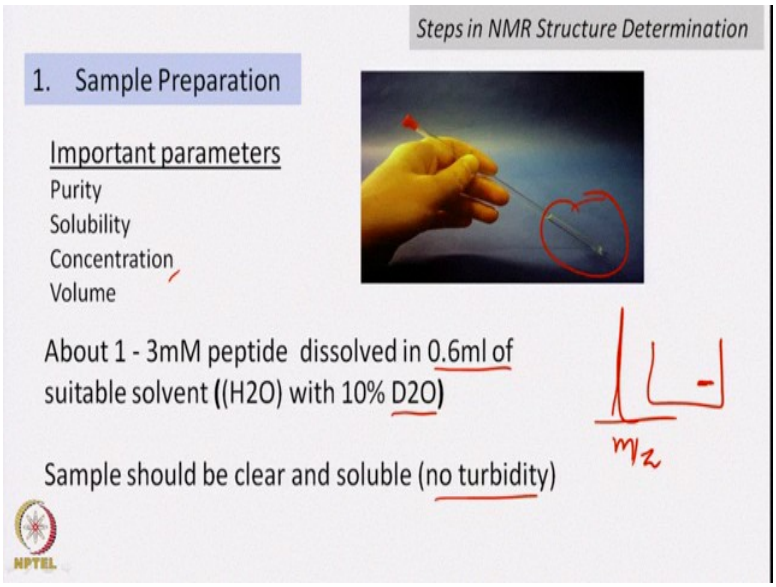
Now, once your data is collected next step will be to assign those data. So, like here if you look at in the background I have given a spectrum. So you need to identify each of this peak coming from the 2D dataset. What this peak means and today I am going to give you some example. So that will be as assignment of the peak or analysis of the data. After you assign these peaks, so you have studied earlier we have discussed that on the NOSY dataset you can generate the restrains, means looking at the NOSY peak you know the distance between two protons and that is called restraint.

So by measuring the intensity of those peak, one can generate the restraints. And now, once restraints are generated all this can be used in structure calculation. So we need all those assignments and all those distance restraints, restraints and then one can use those restraints in a structure calculation. Now, once your structure is calculated that you get a reasonably good structure, you need to validate it is a structure, that means it is not violating any phi and psi torsion angle in the Ramachandran plot. So that is called a structure validation.

So once you did it and after that all the, like after this process your structure is ready. It is a probably it is a good quality structure after validation, then we can deposit in something called protein databank PDB. So that is a single repository for the all structure determined by various scientist and you can look at the coordinates of each of these structure. So let us go all of these one by one. So if you ask me what is the timeline for that, so sample preparation depending upon, if you are getting sample from your friend or you are synthesizing yourself, so peptide synthesis can take a few days and purity if you like one-day.

Then you can dissolve it and collect data for 2-3 days. Assignment might take like if you are doing manually it might take 2 weeks to 1 month or even longer, depending upon how expert you are. Now, once assignment is done like major problem is solved so now distance restraint takes few hours. And then structure calculation, it is iterative process you do step-by-step. So that may take one day or so and then validation in 1, 2 hours. So, like total if you are doing de novo if you are a student who is starting it may take 2, 3 months, 4 months for structure determination. But if you are expert you can do very fast and with now automated software it can be really done fast.

(Refer Slide Time: 7:33)

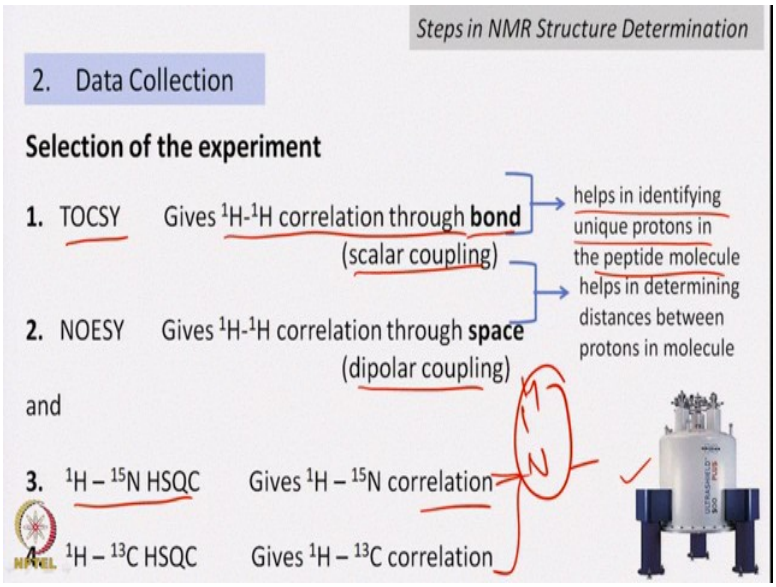


So let us go one by one, so sample preparation as I said sample preparation, sample has to be soluble. First it has to be pure and then it has to be soluble. So purity you can determine by various methods like you can run a mass spectrum to see that you are getting a single pure peak. So mass spectrum what I means, you can do by MALDI or even you can do by just running a, like a SDS-PAGE gel where you should see that you are getting a single band, single banding SDS-PAGE gel or you get a single peak corresponding to this peptide in MALDI spectrum. So m/z has to be a single peak.

So if you get that, then you know your peptide is soluble. Now next is solubilizing, so you have to choose an appropriate buffer and the concentration has to be decent enough. Now as we have now known that generally NMR is one of insensitive technique. Therefore whereas in other techniques you require just micromolar like 10 micromolar, NMR you need few hundred micromolar or maybe even millimolar concentration. So typically for small peptide I would say 1 to 3 millimolar of peptides should be dissolved in 0.6 mL of suitable solvent, mostly in water and you need 10 percent of D_2O deuterium oxide, just for locking the magnet so that magnet does not drift during experiment and you get very clean spectrum.

So that means if your sample is clear, then it has to be really really like transparent. So there should not be any turbidity and sample should be clean and clear. Now if you have done this, then you are ready for doing the experiment.

(Refer Slide Time: 9:24)



So, next up will be just going ahead and now recording the data. Now what kind of data for a peptide? You have to start with something called like as you have discussed this total correlation spectroscopy. Now total correlation spectroscopy as we know it is a like you correlate from say H_α to H_β to H_γ to H_δ. So that gives proton, proton correlation through bond.

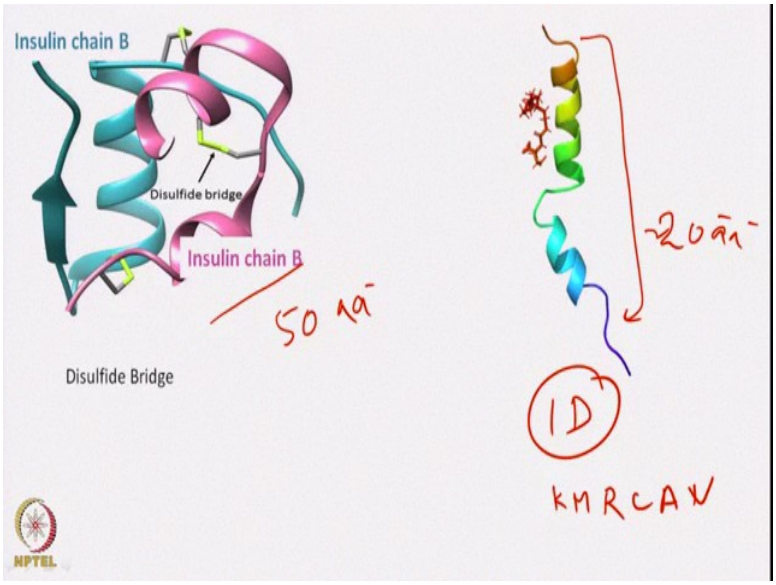
So depending upon, if these protons are correlated you get a spectrum in the TOCSY spectrum. So that gives through bond correlation. And that transfer of magnetization if you recall it correctly it happens through a scalar coupling. So in TOCSY you can go few bond order. Then the other one is NOESY, nuclear overhauser spectroscopy. This gave the correlation through a space and that is mediated by the dipolar coupling. Now these are two proton-proton detected experiment. So then and now you can do at natural abundance something like a HSQC spectrum. So proton-nitrogen HSQC spectrum. So this gives you proton-nitrogen correlation.

So if you remember little bit of biology in each of the amino acid there is a one NH correlation from the backbone and there are few inside chain. So essentially your HSQC will give the number of peaks equal to number of amino acids in your peptide sequence.

Then after that one has to record the ^{13}C HSQC for like a sidechain correlation, like H_α C_α , H_β C_β, H_γ C_γ gamma and all those. So this will give you resonance assignment. So now TOCSY helps in identifying you unique protons in the peptide and NOESY helps in

determining the distance between different protons. So essentially all those will be done on a decent magnet, like 500 or above for a peptide.

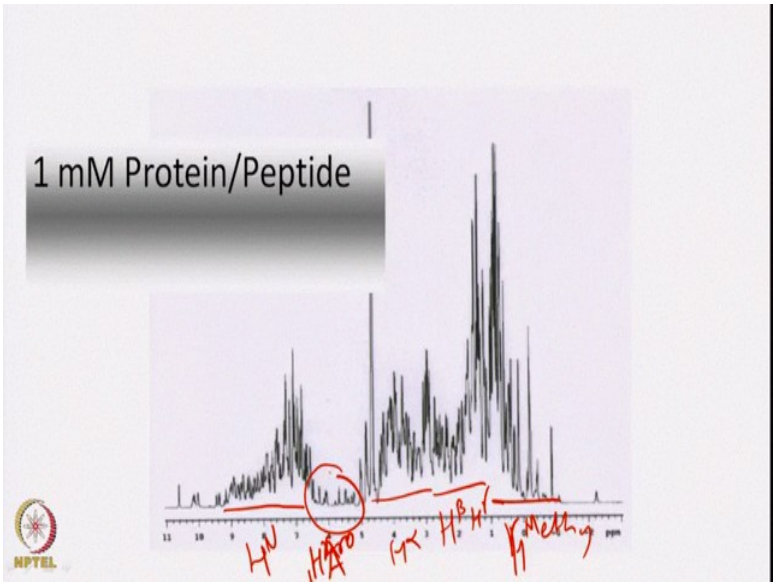
(Refer Slide Time: 11:23)



Let us see, so I want to determine the structure of a small peptide, like suppose here I am taking typically around say 20 amino acid peptide and this is insulin. So where you have a slightly longer like this is 50 amino acid and suppose this is around 20 amino acid peptide. So the first thing that I said, we need to know the primary sequence. So 1D structure or whatever primary sequence, so we need to know amino acids sequence for each of this peptide.

So if we know that, now that is the first prerequisite. NMR does not give you the primary sequence; you need to know from other techniques what is the primary sequence of a peptide.

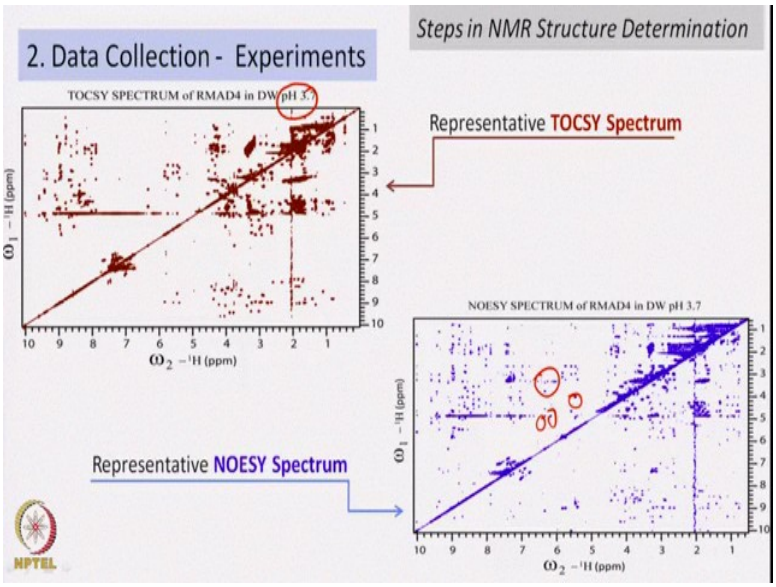
(Refer Slide Time: 12:10)



Now, once we have the primary sequence then we want to get the structure. So simply if you recall by solubilising this peptide into water, so if say record 1 millimolar you get a proton spectrum something like this. So from now we know that these say, these are NH protons, here you can have aromatic proton. So these are H-N protons, this is H aromatic proton, H aromatic proton, aromatic proton. These are our H_α proton, here are H_β, H_γ proton and this probably are H methyl proton.

So we know, like these are the range of the peptides. So now if you look at the peptide or protein spectrum like small size protein of a 50 amino acids, 60 amino acids or peptide more than 20 amino acid, you see how many lines we have. Now these lines cannot be resolved by 1D. So 1D ends in chemistry, now you would require also 2D in chemistry. But for biology minimum requirement is your 2D spectrum. So 2D, that is why I said we start with a TOCSY experiment.

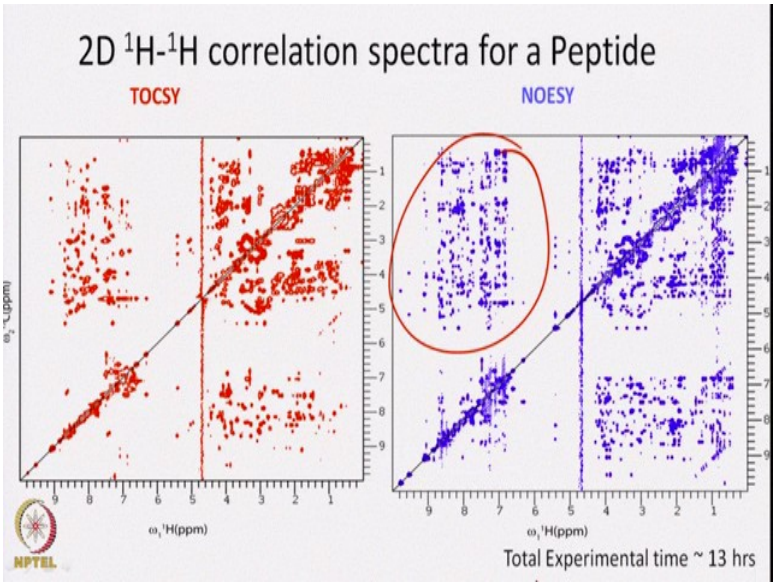
(Refer Slide Time: 13:22)



So that is the complexity of peptide. So we go from 1D to 2D. Now this is say, we are, we have recorded a spectrum of 37 amino acid peptide at double distilled water at pH 3.7. Now you ask me why pH 3.7? I just said this peptide has to be soluble. So we found that this peptide was soluble at lower pH therefore we have to reduce the pH to get a decent spectrum. So one can see here we are getting lots of correlation and now you can like here if you look at here all these are correlation coming from H-N to H_α to H_β to H_γ and that is the TOCSY spectrum, a representative TOCSY spectrum.

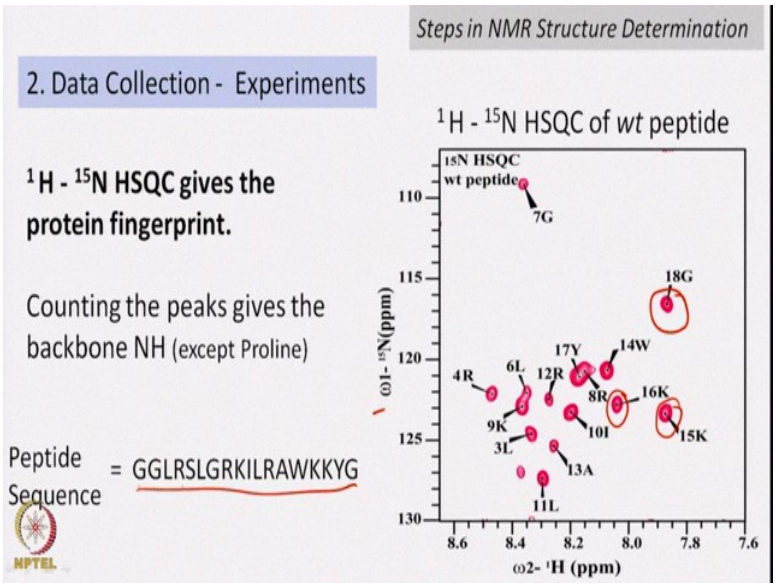
And then you record a NOESY spectrum, you get many more peaks. So now each of these dot is a correlation. So, next job after solubilising the peptide, after recording the NOESY and TOCSY spectrum to analyse this spectrum, assign what these peaks in NOESY spectrum means. Before that we analyse any of this peak, we have to analyse what these peaks means. So, from which spin system like what is the amino acid, where the peak are coming, what is the number of that amino acid which is lysine number 10 or lysine number 15 or a alanine number 12 or something like that, so that is called resonance assignment of the peak. So we need to do that.

(Refer Slide Time: 14:51)



If you take a bigger protein, you get even more crowded spectrum, here, just look at the NOESY spectrum. How much wealth of information you have? So this is on a, like a 50 amino acid. So if you record TOCSY on that 50 amino acid, you get lot more peak here and also a NOESY spectrum. And this can be done typically in 13 hour of experimental time. Next job is to assign these peaks, assign each of these amino acids. So that in a moment I am going to tell you how to assign it.

(Refer Slide Time: 15:25)



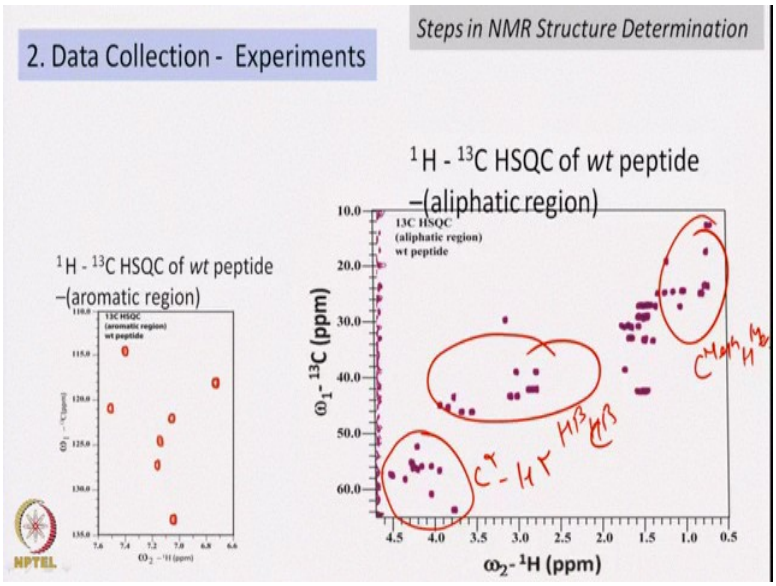
So let us take an example of assignment. But before assignment as I said we will do the say HSQC. So here I am showing you a natural abundance HSQC of a peptide which was like 18

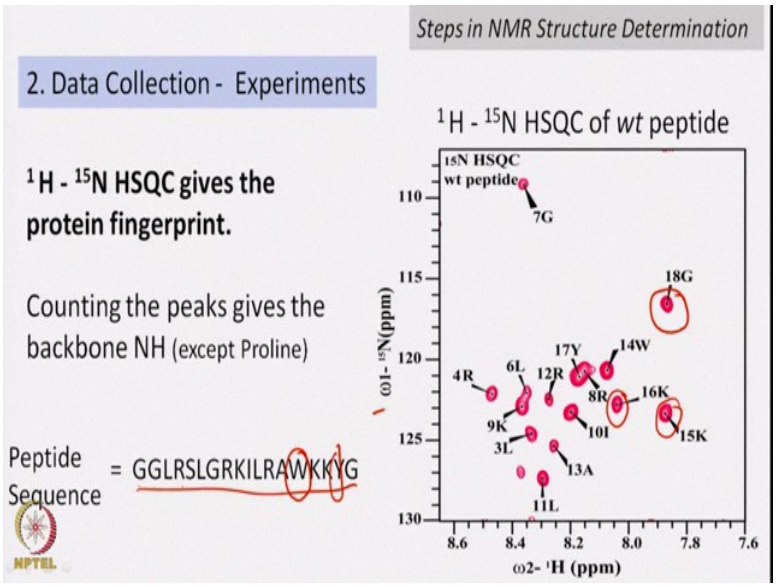
amino acid. So here if you, like if you record long enough you get each of these dots in HSQC spectrum is giving one peak. So that, this means glycine number 18, this means lysine number 15, lysine number 16. I do not know like a priory when I record this spectrum what this is.

So after the resonance assignment I was able to pinpoint this. But you get as number of amino acids as you have in your peptide or protein sequence. Now, HSQC gives you fingerprint. So if you record this spectrum, you know that each of these peaks is coming from one amino acids. So now it is, now you know number of amino acids and also you know that my peptide is good enough.

Now, so only one amino acid that will not give peak in the N15 HSQC which is proline, why? Because proline is a imino acid not an amino acid, proline does not have NH. Therefore this is NH correlation. So here is N15 and here is proton. In NH correlation proline will not have any peak. So if I my sequence here, if you look at GGLRSL all the way from G to here are giving all peaks. Since I do not have any proline, so essentially all 18 amino acids are probably giving me peak. You might miss 1 or 2 peak that is because of dynamics or some modification. So that is a ^{15}N HSQC.

(Refer Slide Time: 17:16)



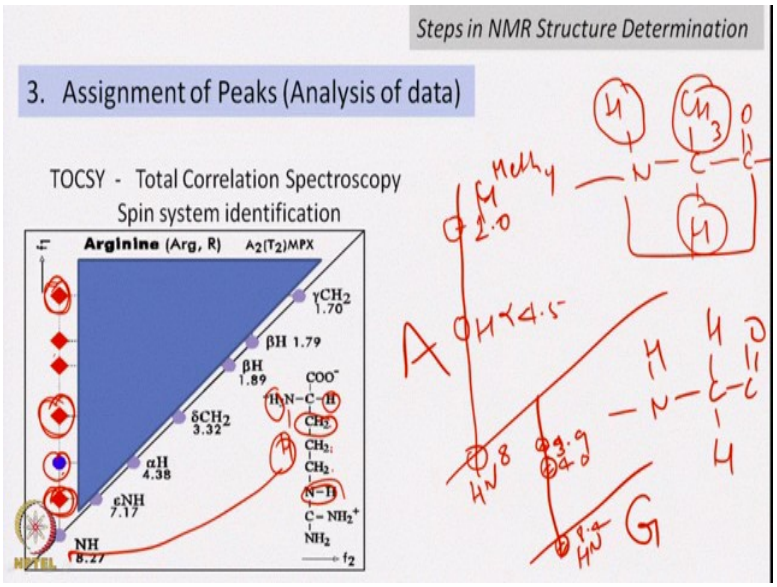


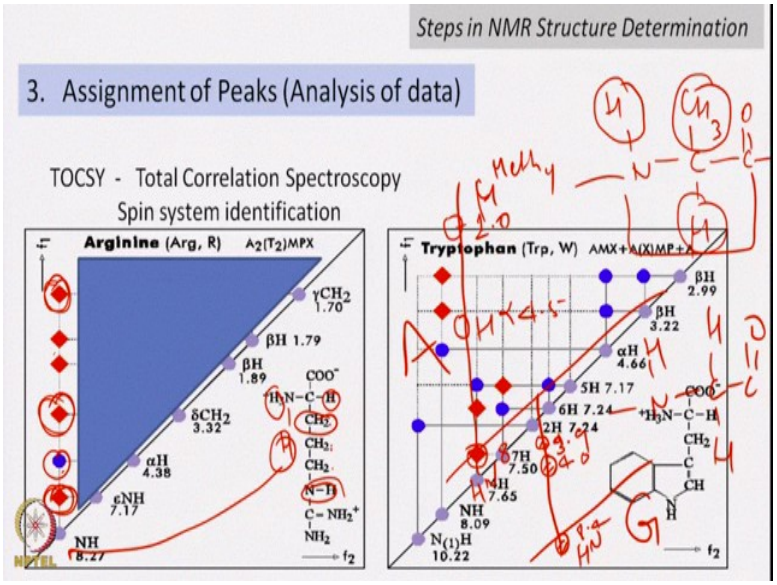
Similar I can do the data collection for ^{13}C HSQC. So if I do for the same peptide that 18 amino acid peptide that I showed you, this is the region from the aromatic, so few of the amino acids has a aromatic ring, like if you see tryptophan as a aromatic ring. And then here your tyrosine has a aromatic ring. So these two will give peaks in the aromatic amine acid. And that is what you get from the aromatic peptides.

Then from the aliphatic region all the CH_{α}, CH_{β}, $CH_{\gamma}$and all those will come. So here is your CH_{α}, like C_{α}, H_{α} correlation. Then here you have a C_{β}, H_{β}correlation and here are your methyls, C methyl and H methyl correlation. This I know because we know the chemical shift, but I do not know here which is from alanine number say 5 or something like that, sequence-specific I do not know. So for that I do, need to do resonance assignment.

So now we know sample preparation, it has to be pure, it has to be soluble and 10 percent of D_2O needs to be added to the sample for locking purpose. Now once you sample which is of good concentration and high purity, we recorded a series of experiment. So minimum experiment that we counted is TOCSY for spin system assignment, NOESY for like a distance restraint. So the sequential and long-range correlation in space. Then we recorded ^{15}N HSQC that is a fingerprint for peptide or protein. Then we recorded ^{13}C HSQC from the aromatic region and the aliphatic region.

(Refer Slide Time: 19:26)





Now next job and the most important job is to assign these peaks. So I will give you some example how to start assignment. Now assign, assignment should start with a TOCSY spectrum. So TOCSY I have here one experiment, but maybe I will give you some more for amino acid. So let us start with the simple amino acid, which is alanine. If suppose your alanine is there in the peptides chain, let us write it down alanine here and NH, this is our alanine.

So if alanine is there, now what all protons we have alanine? First is say NH proton, second is CH_α proton and then C_β or C methyl H proton. So in TOCSY essentially we should see three peaks. One for your NH proton, which will be suppose around HN. So, which will be suppose

around 8 ppm, then we have one proton CH alpha proton. So H alpha proton which is around say 4.5 ppm. And then you have around like 1.8 or 2 ppm for methyl protons C, sorry, H same methyl proton or H beta proton.

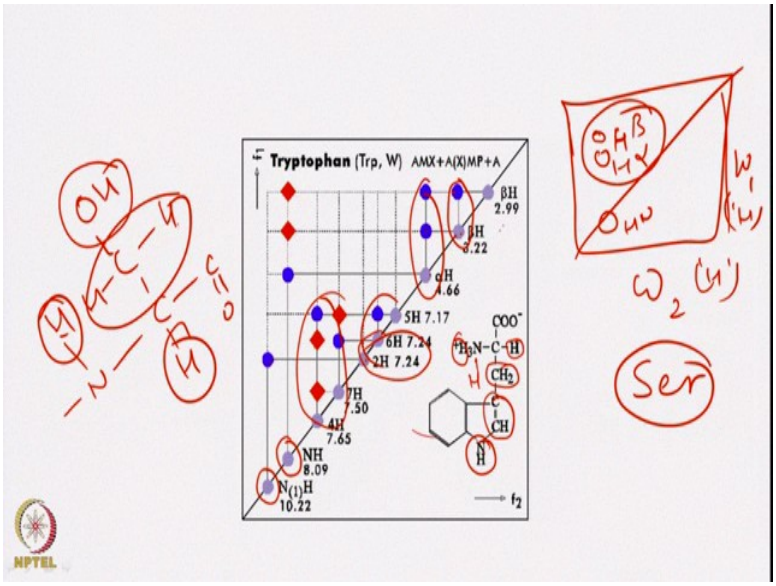
Now if I see this kind of spin system in my TOCSY spectrum, I know that this is coming from the alanine. So I identify alanine but still I do not know which alanine is. Similarly, if I take an example for say glycine. So glycine what we have here? NH CH H and CO. Now if you take glycine, so now glycine, what glycine should have? 1 NH and 2 for say H alpha 1 and H alpha 2. So this is around say HN, which will be around again 8 point something, 4 ppm.

And these two are say around 4 ppm and 3.9 ppm something like this. So if you see this pattern, you know a priory that this is for glycine and this is for alanine. So this is the way we assign the TOCSY spectrum. Let us now move to little complex system, say we have taken here the longest one say arginine. What we have in arginine here? Let us start, so I will take this simplicity NH. So arginine, because we are talking in a peptide sequence, so arginine has a NH like amide proton. Then we have CH_{α}, C_{β}, C_{γ}, C_{δ} then these are the side-chain NH and these are all three side-chain NH.

So let us see what we are getting. So here is coming our backbone NH this one, 8.27. Then let us go little back here, then we are getting CH, say alpha here, CH_{α} here, CH_{β} here, sorry, CH_{β} at 1.89 and here. And then γ are here and then δ is here. And these are coming from the side-chain, side-chain NH is around 7.7, this is H alpha. So let us repeat it again.

So at the diagonal here what I am getting here is from this one, 8.27. Then let us move at H alpha 4.38. Then H_{β} are here, 1.89 and 1.79. Then γ are 1.7 and then δ at 3.32 and then the side-chain NH which is this one is coming at 7.7. So if I see this kind of pattern in my TOCSY spectrum, I know that this is an alanine spin system.

(Refer Slide Time: 23:18)



Now let us take an example for tryptophan. This is again a bigger amino acid. So tryptophan if we look at here again for simplicity, we will take this as a HN because this is in peptide bond. So we have a HN from the backbone then we have CH_α and CH_β and these are all from the side-chains. So this is from aromatic ring and these are from this 5-membered ring and this is the side-chain NH.

So let us see now, we have two NH, one from the backbone, one from the side-chain. Now backbone NH generally comes around 8. So this is backbone NH and the side-chain NH comes around like 10 or higher than that. So here that we have. So if you look at the TOCSY correlation spectrum, so from NH you can get this CH alpha that is this one, around 4.66 and then you have β. So two β we have, one comes at 3.22 and one comes at 2.99. That is 3 ppm. So that is what TOCSY you are getting.

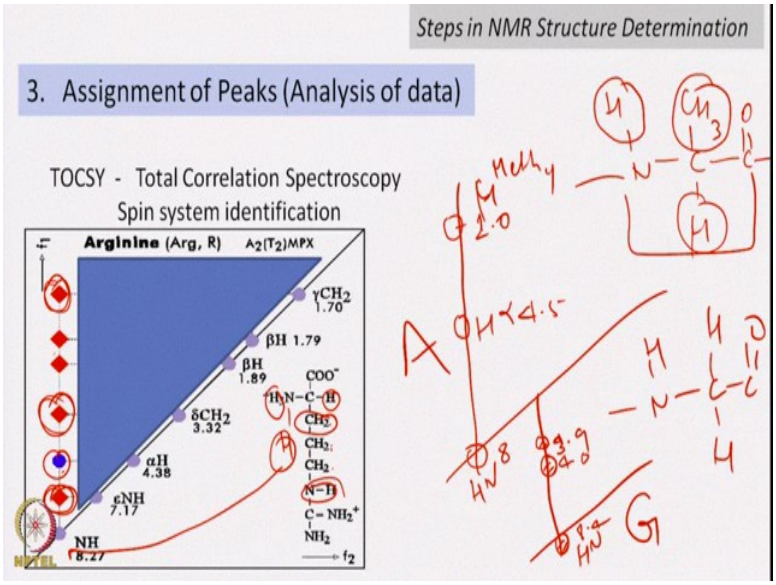
Now from this NH, we can get the side-chains. So like here, so that is here. Now from this aromatics you can again get it from all the aromatic correlation. So here you have this aromatic correlation and that you can assign it from these guys. Then again here you have for 5-member ring. So here like this was and that you can assign it, all this. So from H_α you get again correlation to H_2 H_β, these and then from H_β you get it from the other H_β.

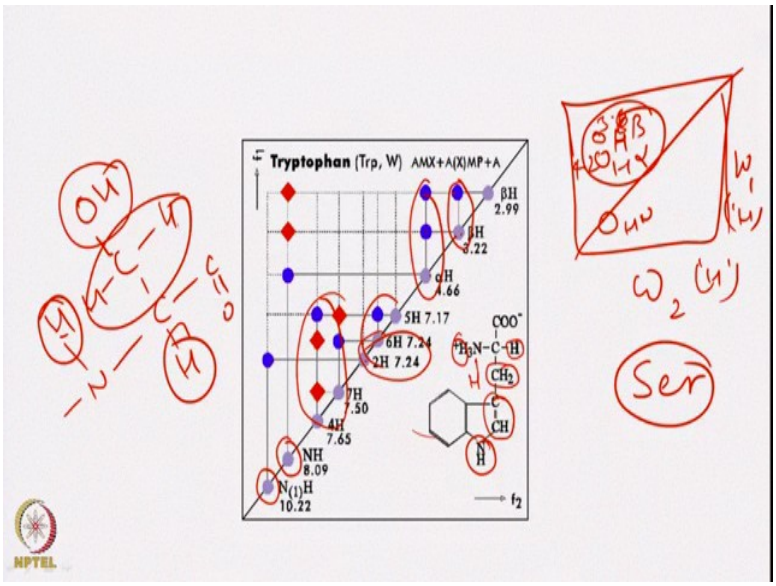
So this pattern if you look at, now by doing this pattern we know that this is tryptophan. So that assignment you are doing and then we are getting the rest. Now amino acid type assignment for each of this. Just to make it more comfortable let us take another amino acid

say serine. So what serine is? I will write serine in a peptide bond CH, CO and say OH. So now we have a 1 here, 1 here and 1 here.

So let us see how it comes, $\omega_2 H^1$ and $\omega_1 H^1$. So here we are getting HN peak for your backbone this one. And then we are getting so here, so say H_α and H_β. Now H_α and H_β in this case will be quite close, why? Because here, if you remember O is electronegative wave and that will actually create this reading of beta proton. So this is H_α and this is H_β. So that is come quite close, but in case of methyl you saw that all that, if this was CH_2 in many cases see CH_2 here is around 3 ppm.

(Refer Slide Time: 26:42)

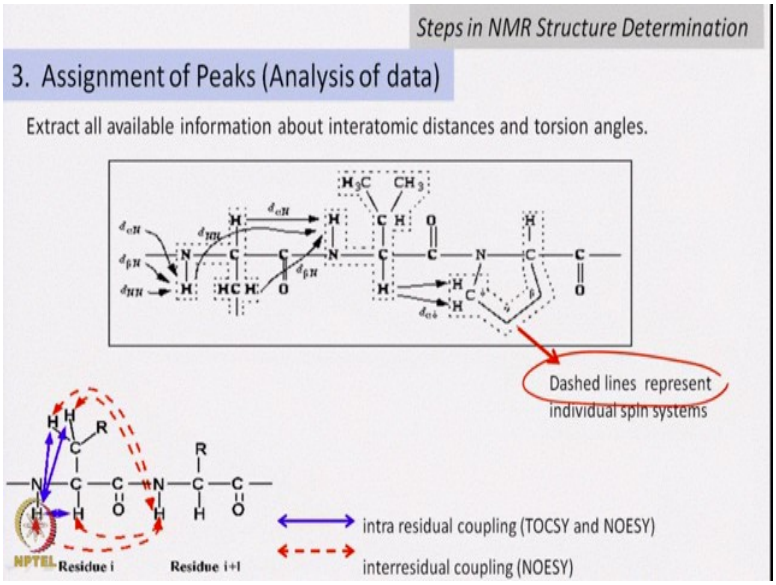




In the previous case, we looked at here the H_β are 1.79. In case of serine again, we looked at that it comes, it will come around 3.6 ppm and this will be or 3.8 ppm and this will be 4.2 ppm. So that will come quite close. Now if you look at we are getting unique pattern from different amino acids. So like alanine as I said 8.34 ppm, 4.2 ppm and around 2 ppm, so if these three peaks are there you know alanine.

Serine again you are getting H_α and H_β quite close. Glycine you are getting 2H alpha around 4 ppm. So looking at this pattern, you will know that what kind of amino acid is, what type of amino acid it is. Tryptophan you see here are the pattern, so H_α, $H_\beta 1$, $H_\beta 2$ by looking you know that this is tryptophan. Similarly like we had looked at the arginine long TOCSY correlation.

(Refer Slide Time: 27:43)



So by this way we identify different amino acids. Now individual spin systems are identify here. Now, once the spin systems are identified, now we have to find the correlation with next amino acid. Because here, we, I found this is tryptophan, this is arginine, this is alanine, this is serine, this is glycine or whatsoever. But which alanine, which glycine, what is the number of that alanine? Alanine number 5 or alanine number 12, glycine number 7 or glycine number 11?

For that I need to know what is the near that and then we can solve this puzzle. So what I will do, I will start from this point how to use now NOESY spectrum for knowing the nearest neighbour. And that will help us identify the number of these amino acids. And we will

continue from there and then we see how to do complete resonance assignment identification of the system. So I will start again with this point. Thank you very much and looking forward to see you in the next class.