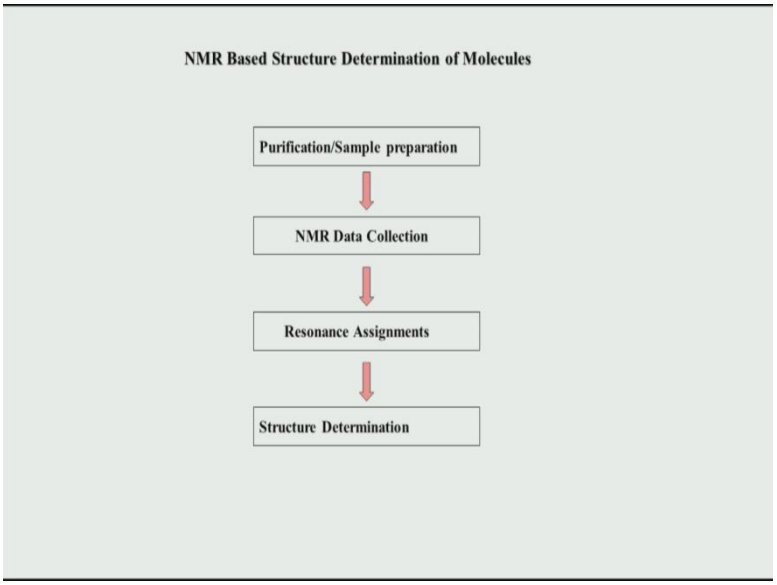
Principles and Applications of NMR Spectroscopy
Professor Hanudatta S. Atreya
NMR Research Centre
Indian Institute of Science Bangalore
Module 7
Lecture No 33

So from this lecture onwards we will look at more advance topics in NMR and we will start first with how we can use the understanding of different spectroscopy techniques in NMR and apply it to the structure determination of molecule, specifically we look at a peptide because peptide molecules illustrates the complexity involved in determining structures of biomolecules or molecules in general NMR, so let us look at that first.

(Refer Slide Time: 0:55)



So now you can see this, a general flow chart which typically people use when we solve the structure by NMR. First thing is that you need a heavy sample, so you need to purify your sample and as we saw and the purity is very important parameter when it comes to looking a structure determination of molecules. Typically you say that it should be more than 95% pure, the reason for this is that impurities present in your sample will have extra peaks or correlations in 1D and 2D NMR and that can lead to ambiguity or droughts in solving structure.

So, therefore if it is more than 95% pure, it is very unlikely that the small impurities will give any strong correlations in the spectrum. The second is also have the sample we typically take about a mili-molar sample this is for case of biomolecules and natural products but if you have an organic synthetic synthesis molecule we can look at higher concentration you may go

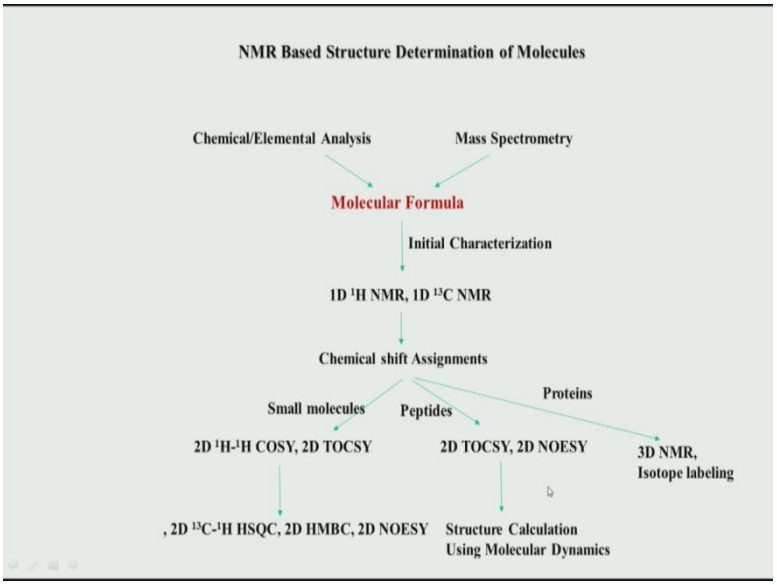
even up to 10 milimolar, but in a standard NMR spectrometers nowadays, one milimolar is sufficient for good spectrum to recollected.

So, re-put the sample in the magnet and then we record a series of NMR data. So, this is typically 1D NMR, 2D NMR and in case of biomolecules like proteins 3D NMR, but we will restrict this course only to 1D and 2D. Now, once we have record the data we have to analyze the data, we have to process the data and we have seen that how do we process data and how we analyze and based the analysis what is the objective of the analysis?

The objective of the analysis is mainly to achieve the third step which is resonance assignments means every peak in your spectrum, be it 1D or 2D the goal is to assign this to a particular atom in the molecule okay, so the every peak should be assign. Now, after the once assignments are done, then we use then we go on to structure determination. So, remember structure determination here what I mean is not just simply the structural formula of the compound like we saw in the case of 1D and 2D NMR.

We looked at few examples where we had only the chemical formula molecular formula and we tried to figure out the different functional group connectivity and look at the structural formula but that is not what is meant here, what is meant here is the complete 3 dimensional structural information, and that comes from very important experiment it called NOESY or ROESY because remember we saw that each peak in a ROESY or NOESY spectrum gives information of the distance between atoms and what is 3 dimensional structure it is just in but distance information between different atoms in 3 dimensional space is what we are looking for, so that is what is meant by structure determination, so this are basically the all steps involved in going from sample to structure by NMR.

(Refer Slide Time: 4:17)



So, where do the different experiments that we learnt comes into picture. So it comes into picture in the following manners, so let us say that we have the molecular formula of the compound, how do we get that that is by different approaches like chemical or elemental analysis? It gives the amount of percentage of nitrogen, oxygen, carbon, hydrogen, etc, and mass spectrometry gives me the mass of that sample and combining these two information, we can generate a possible molecular formula and it is what is done in chemistry and we will not go into detail here.

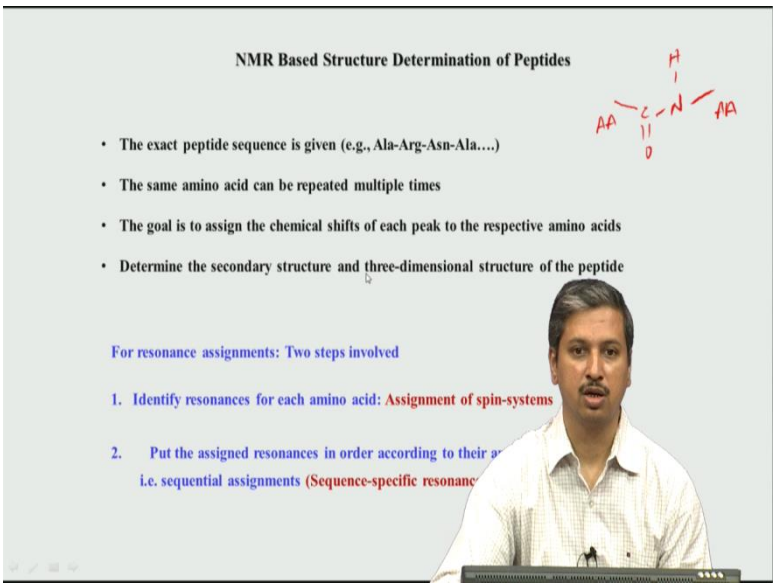
Now, given a molecular formula this is example which we have seen. We can first initially characterized molecule by simply 1D and carbon and proton NMR and at this step all you are doing is you are simply recording the spectrum and then identifying what are the peaks what how many peaks are they commensurate means are they equivalent to what you expect from based on this formula and so on so forth. So this step is not going to use structural detailed information and is only meant for just initial characterization.

Once you have done that we need to now, accomplish or achieve it is chemical shift resonance assignments for that we need to go to higher dimensional that is 2D NMR. So, for small molecules, organic compounds typically this set of experiments are sufficient to establish the resonance assignments. So, this is what we all of these experiments we have seen in the in this course. If you go to peptides which is a link of amino acids, again this sort of experiments what we discussed also will be sufficient. In fact, just 2 dimensional experiments NOESY and TOCSY or TOCSY and NOESY are sufficient to establish or accomplish this type of chemical shift resonance assignment. And once you have done that

the next step is to do now the full fleshed structure calculation that is determination of 3D structure using NOESY.

So, the information of NOESY is used and the cross peak intensities are converted into distance information and those distance values are given to a program computer program, which carries out what is known as molecular dynamics calculation and the output of that is the final structure, so this is typically for peptides. When we go to proteins, which are bigger molecule where 2D NMR is not going to be sufficient, we have go to the even larger high dimensional spectroscopy that is 3D NMR and not only that, we need to do special sample preparation by doing what is known as carbon 13 and 15 labeling, but we are not going to go into details of that particular aspects, so the whole course is only focused on small molecules and peptides. So, now we will focus or how we can assign and get the structures of peptides with 2D NMR.

(Refer Slide Time: 7:37)



So, this is how the process that goes on, so first thing is you need to have exact peptide sequence with you, that is you should know the sequence that is how the amino acids sequence is there in your sample. So, let us say you have this remember there are 20 naturally occurring amino acids in our in biology in nature and we are also they can be unnatural amino acids.

A very synthetic organic compound can be inserted, but all of these are connected to each other by a peptide bond so, what is the peptide bond? This is well known to all chemists basically peptide bond is looking this kind of a connection. So, this is one amino acid I will

note it as AA, the second amino acid and this two are link together by the peptide bond. So this is what we are going to focus on in this next few slides or next this part of the course where we look at the structures of peptides. So, this is the peptide bond but you can have unnatural amino acids but we are again, we will take an example of a natural amino acid containing peptide containing acid amino acid and there are 20 of them, so this could be the sequence example, but remember you can also have the same amino acid repeated multiple times.

So, look at this example, I have a peptide which as alanine, arginine, asparagine, alanine (Ala-Arg-Asp-Ala-…) so on so forth, so Alanine has come twice. So, there is no limit or no restriction as to you how many times the same amino acids can repeat, so this is like a polymer so we are looking at a biopolymer okay. So, this is where the complexity starts if you have multiple amino acids of the same time same type then the spectrum it is complicated to interpret. So, let us see how that can be accomplish, so what are two goals in our whole project now, the goal is to assign the chemical shift of each peak of the spectrum which we get in the spectrum to the respective amino acid the peptide, okay.

So, for example, your alanine that is which has a CH alpha and beta which we saw in the example in the last class, then there it has alpha-proton, alpha-beta proton, alpha-carbon, beta-carbon. So, I have 4 carbon and proton and of course you have the carbonyl and amides, amine which will not considered right now, we are looking at only the alpha and beta carbon proton and that alpha-carbon, beta-carbon protons where do they come in the spectrum? What peaks chemical shift are those? Those things have to be assigned at this first step. When once you have know assign the chemical shift of every peak in the spectra to the respective amino acids, then you can go ahead and start looking at the structure. The structure means there are 2 types of structures in peptides and proteins. We say one is secondary structure that is alpha-helix beta-sheet and that the final structure you what we say is 3 dimensional structure of the peptide.

So, this is the basic requirement or objectives in looking at the structure of a peptide. So, when it comes to resonance assignments, there are 2 steps involved or it can be accomplished in 2 steps. Number one is that given the spectrum find out the resonances of each amino acid okay. So, let us understand what this means? It means again let us take an example of this peptide shown on the top alanine, arginine, asparagine, alanine. So, each of these amino acid is now a spin system because remember because of this peptide bond here, the connection

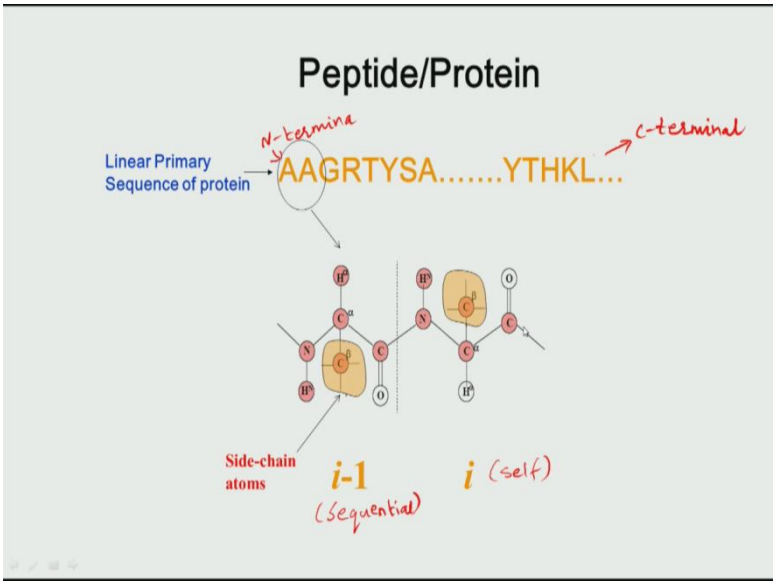
between this amino acid to this acid or j-coupling is Brocken because of this intermediate long distance. So, the proton here is very far away from proton here, so there is no coupling between them so basically we can say that each amino acid is like an isolated spin system.

So if I have 4 amino acids in the peptide, I can say it is a tetra peptide means there will be 4 spin systems in the spectrum, okay. So, in a TOCSY spectrum example you will expect to get 4 different spin systems, but alanine is twice here 2 times. So, there will be 2 spin systems for each one for each alanine. So, the first step you were given the spectrum you have to identify the spin systems. Okay. So, we have to say that if you recollect in the last class we saw the HSQC TOCSY experiment, where we saw that the valine spin system, has a set of spins and connected to each other in a TOCSY.

Similarly, leucine at another set of peaks connected to each other. So, we could differentiate separate or distinguish valine from leucine spin system by TOCSY, same thing has to be done in this case. So once we have identified these spin systems then you have to find out which spin system comes in which sequence or order, to look at this again in this sequence the first amino acid is alanine and second is arginine, third is a asparagine and fourth is alanine.

So, that means the 2 alanine spin system which I would have got from my TOCSY, I should now then decide which spin system comes this side that is the fourth one here example and which spin system alanine comes in the beginning. And that is not easy by just looking at this spin system because the spin system as such does not tell you anything about its location in the sequence. It only tells you there is an alanine present or there are 2 alanines in your sample or in a peptide. It does not tell you which alanine spin system comes in the last or in this case the fourth or which spin systems comes here. So, that kind of arrangement of arranging the spin systems in the right sequence is called as sequence specific resonance assignment. So, this is a very important step, only when these 2 steps are completed then only we can say that a molecule or peptide has his goal is achieved that is the goal of chemical shift assignment is achieved .

(Refer Slide Time: 14:41)



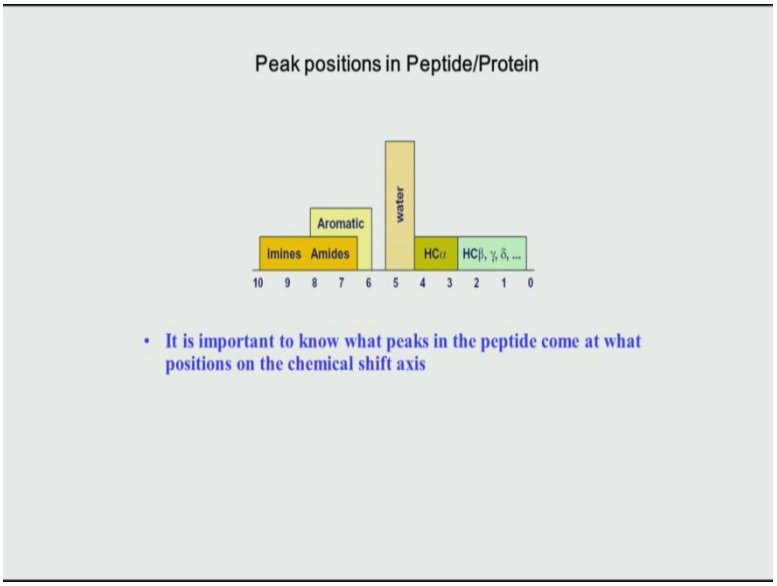
So, when once you do that, then comes the next step of secondary and tertiary structures. So, let us see how we can achieve these two steps shown here that is of assigning spin systems and then followed by sequence specific resonance assignment. So, this is example again of a peptide or a protein sequence which runs like this. So, typically we denote this side as N ter-, so let me write that now, so we say this is called N terminal residue and this is some this side is C terminal. So, this is the notation which is used in NMR and therefore, we will use that notation to talk about the assignments. So, now if you see generally if you go to each amino acids and this is the dipeptide, dipeptide will be like this. So, let us look at this carefully, what we are seeing here is the individual atoms and you can say this is N, this is one amino acids this part is shown separated by a dotted line, so this portion is one amino acid and this portion is another amino acid.

So, if you go by this notation, we say ‘i’ means this amino acid the first amino acid the first in the sense any amino acid we will use the word arbitrarily ‘i’ and then the one previous to with that is before it in the sequence will be now called i - 1. Okay. So, when I say when in the next few classes we will discussing i, I - 1 like that you should immediately it should come to your mind but i means the amino acids any amino acid which is self, so we use the word self, so let me write that down, we say this is self and we say this sequential okay, so we use the word self and sequential to denote these two different amino acids. Okay. So, this general this is a peptide bond which is basically linking connecting the 3 amino acids. So, in protein and peptide NMR jargon terminology we have to define some more terms; first is backbone

atoms, so this zigzag chain running through the sequence that is little go of this side and this side and which as H Alpha including the H Alpha, we use the word backbone atoms.

And the remaining which is this beta onwards, so we will have gamma, delta, epsilon on every amino acid that part portion of the amino acid is known as side chain. These two differentiation is done because peptide NMR and protein in specifically protein not so much in peptide, there are different set of NMR experiments which are particular to only the backbone atoms or those which gives correlations for side chain atoms. So, where is therefore, a segregation in terms of the nomenclature and we distinguish these 2 sets of atom backbone and side-chain.

(Refer Slide Time: 17:46)

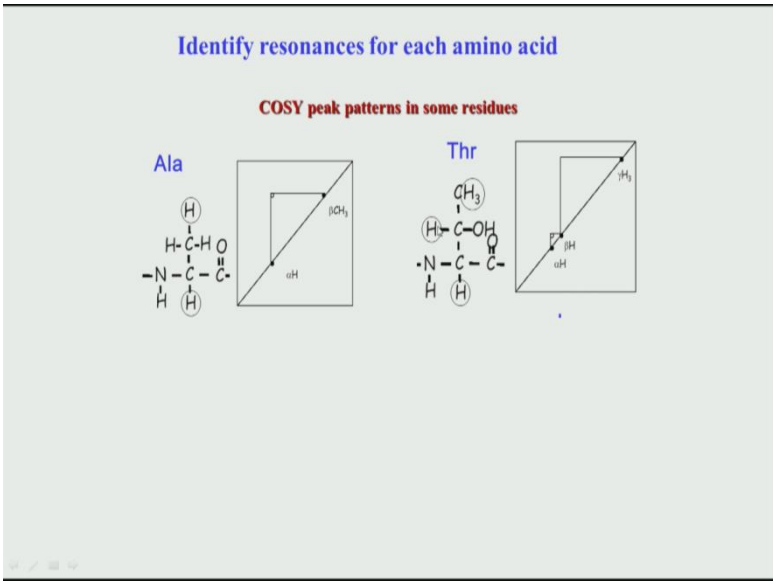


So, now let us move ahead that now suppose I record a 1D spectrum of a peptide, so what kinds of spectrum in expect to get and what are the different chemical shift values? So, this is just a schematic, down the line we will see a real NMR spectrum of a peptide 1D NMR, but right now this is a slight test telling you the rough positions where the different functional groups in the peptide will come. So let us start from this end the 0ppm here comes the methyl, if you recollect methyl peaks comes here and the aliphatic portion that is the gamma, beta, delta, hydrogen, carbon atoms, so hydrogen atoms they come in this column this position for every peptide. So, this one thing we should keep in mind this whole picture which is shown here is conserved means same for every peptide very rarely you will get unusual values of course this is for naturally occurring amino acids.

It may happen for a natural amino acid or some special amino acid it may deviate from this zones so this demarcations, but by and large for all peptides you expect to get correlations or peaks for different functional groups in this zones which are marked here. So between 3 to 5ppm you typically get peaks for all amino acids which have the H Alpha of all amino acids and then comes the water, so typically peptide is a biological molecule and same with protein, so we record those spectra in water or of deuterium that is D2O. In that case, the water comes at this position around 5ppm. And then between 6 and 7 or 7or 8ppm comes the protons from the aromatic amino acids. These are typically phenylalanine, tyrosine, tryptophan, histidine and these are the 4 amino acids in proteins. So, this is similar to what we normally see in organic compound that if an aromatic spin system you get peak somewhere between 6 to 8.

Amides that is the CONH, the peptide bond, the NH of the peptide bond also comes in this particular regions and then comes more special like imines and so on, but our major focus now is focused on these 4 functional groups that is amides, aromatic, CH Alpha, H Alpha and remaining aliphatic side chains okay. So, this is important to know what fix comes where for resonance assignment, so now let us look at us kind of a pattern peak pattern for amino acids.

(Refer Slide Time: 20:41)

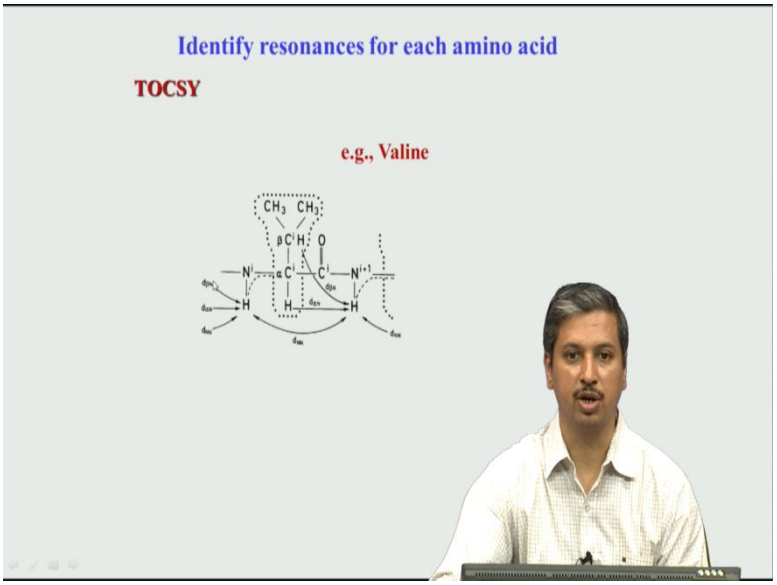


So, again remember we are going to restrict our focus only to the 20 naturally occurring L amino acids. L amino acids are what we see in nature in all peptides in proteins and therefore, we will focus on those 20 amino acids. So, we will look at a few examples of what is a peak pattern, so now given that we have 20 types now, we have let us look at how an alanine COESY pattern. This proton to proton COESY will look like and how a peak pattern of some

another amino acid threonine will look like. So this shown here, Alanine is a very simple and this is what we see in many textbooks then that is alanine alpha and we have already seen this is correlation to the beta proton, okay.

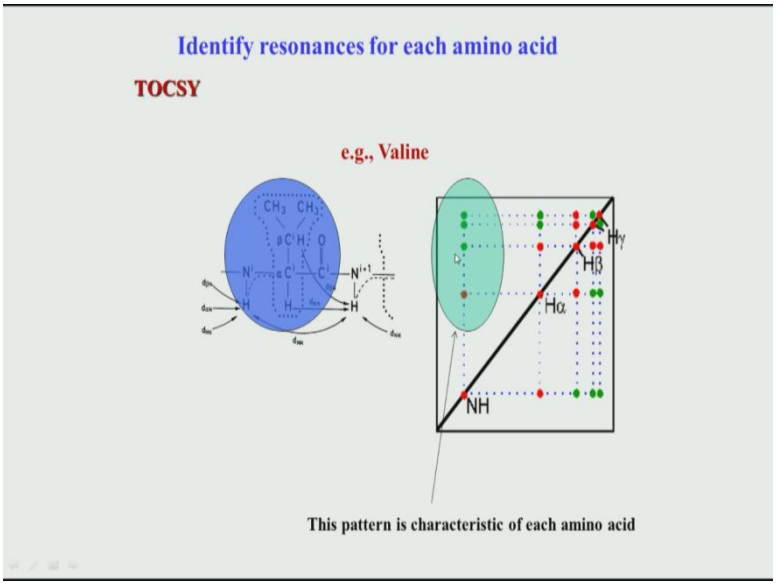
So, remember this axis both are hydrogens, so beta proton is coupled to the alpha proton by 3 bonds j-coupling. Similarly, we can see for threonine you get again 3 bond correlations proton to proton. There are 3 protons here. One is alpha, beta, gamma, so, alpha goes to gives correlation to beta, and beta gives correlation to gamma okay, similarly for many amino acids like this. Now let us see how the TOCSY spectrum will look like for let us say some spin system. So, remember whenever you talk about TOCSY moment we talk about TOCSY the word spin system comes into picture.

(Refer Slide Time: 22:30)



So, as we discussed in each amino acid in a peptide is decup or not coupled to the neighboring amino acid by j-coupling, so therefore, you can say that they are isolated spin system. So, example in a valine this amino acid I would expect correlations between only this proton to this proton and all of them but not to any neighboring amino acids.

(Refer Slide Time: 22:51)

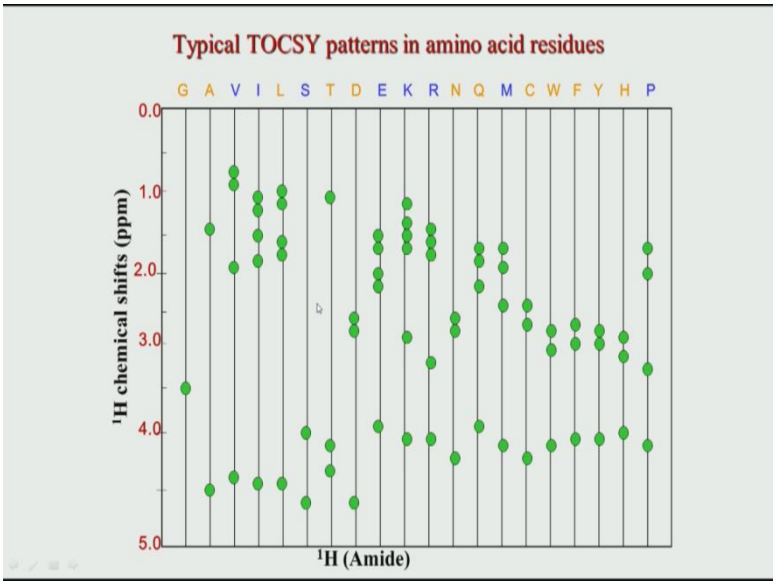


So, this is what is shown here you can see that this diagonal, this is a TOCSY peak and you can see the correlation on TOCSY NH, this is from starting from this NH to the C alpha. Remember this is 3 bond j-coupling as shown by this arrow here and then from this proton H alpha you see correlation to the near neighboring side chain aliphatic spin system. So, to beta to gamma 1, gamma 2, now, which gamma is which proton we do not know? So, that we cannot distinguish just by stand TOCSY, so we will just called arbitrary one gamma as gamma 1 and another gamma as gamma 2. So, this is known as speuro specific sporocardal protons.

This procaral ethyl's right now we cannot distinguish, but we will see there are 2 gammas. So, this is the pattern what you shown in this picture here for a TOCSY spectrum of a valine. Now, those which are in the red color are the COESY peaks, which are only next neighboring 3 bond here, but the green peaks are one for a TOCSY. So, this whole spin system is now by are connected to each other and that is how a TOCSY spectrum should look like and this is what we have been seeing in the TOCSY part of the course.

So, typically when we go to peptides this pattern what you shown in this green color oval circle is what we look for. This is a very interesting pattern, this is called a fingerprint region and we can say we will see the next slide that all the 20 amino acids have unique fingerprint pattern like this. So, no two amino acids are similar of course, compare if you look at aromatics, yes family of amino acid will look similar but they are in a very well interestingly they have very unique fingerprint, so let us see that in the next slide here.

(Refer Slide Time: 25:06)



So, what is shown here is on the X-axis just the different amino acids, the 20 amino acids are listed here on the top and on the Y-axis is the chemical shift proton values, but only shown from 0 to 5. So, what we are trying to see here as the previous slide we are looking at this portion here shown by this green circle. And this portion if you see it is only the aliphatic portion of the amino acid and that goes somewhere from 0ppm which is a top part the center which is around 5ppm. If you recollect in the previous slide we saw the center is water, so typically below water means down field of water comes aromatics and amides, but up field above water comes the aliphatic and that is what is what we shown here and you can see each amino acid has a very distinct pattern of peaks expected in this region.

So for example, Glycine of course there will be 2 peaks, it can be a 2 peaks or it can be 1 depending on you can (())(26:17) but it has only a 2 alpha protons, there are no hydrogens in the beta, gamma, delta, it is only alpha and that is what comes around 3.5 to 4ppm. Then comes alanine which is what we saw if you have been seen, there is an alpha proton which is around 4 to 4.5 or sometimes it can come at 3.5, when it comes in this region and there is beta carbon proton which is a very much up and come from 1.2, 1.5, then comes valine. Valine as we just none saw has alpha, beta 1 gamma 2 and that is what has a pattern like this. So like, this for every amino acid there is a kind of a distinct pattern which very easy based on those patterns to recognize what are those amino acids.

Now if you see here, these have only beta protons you can see all these amino acids cysteine, tryptophan, phenylalanine, tyrosine, histidine similarly, aspartic acid and asparagine. So, these are called AMX spin system because they have only alpha and beta protons and the spin

system terminology we have not seen in this course, we can see that later in the part of the course and then this amino acids like eutarmic Glutamic acid, glycine, arginine , isoleucine and so on. They have bit gamma and delta protons as well and therefore, we use a word long side chains. But as far as our interpretation of NMR data is concern, we will not worry about this nomenclature, we will simply use directly the word.

So one has to know what does each amino acid stands for and this is single letter code that is histidine, this tyrosine, phenylalanine, so I would recommend that you go through any standard textbook in chemist biochemistry to know what letters we used for which amino acids. So, now in the next class we will focus on given this nice peak pattern of each amino acid, how can that be used to assign the first step, which is spin system assignment and then subsequently the second step of sequence specific resonance assignments.