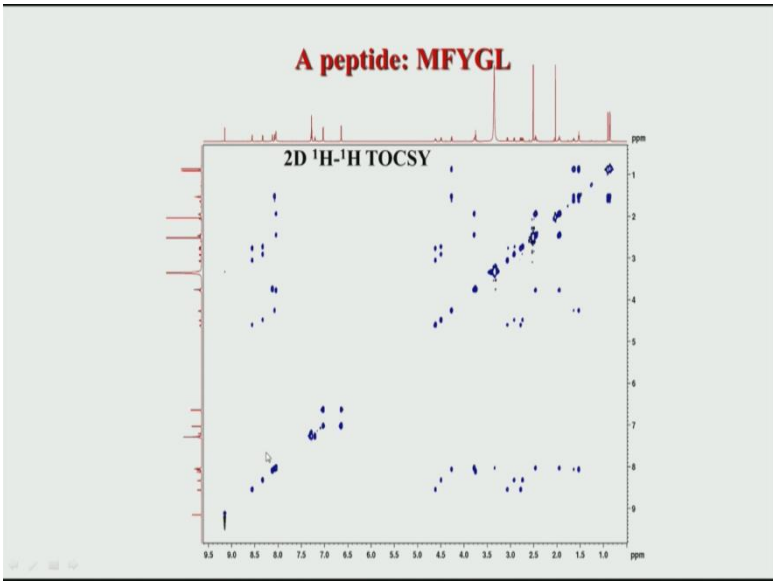
Principles and Applications of NMR Spectroscopy
Professor Hanudatta S. Atreya
NMR Research Centre
Indian Institute of Science Bangalore
Module 7
Lecture No 35

(Refer Slide Time: 0:40)



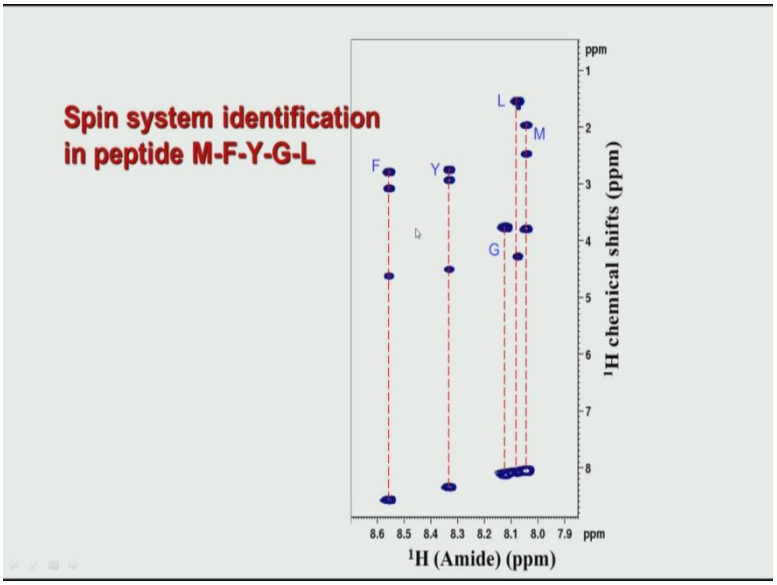
In the last class, we started with the assignments of this peptide MFYGL, we now go through it and see how you can assign this peptide using 2D NOESY and TOCSY spectrum as we have seen earlier. So this is the 2D TOCSY spectrum of the peptide as we can see the correlations from NH region. This is the NH region which is comes somewhere between 7.5 and all the way you see all the correlations to the aliphatic that is the methyl's and the methylene regions.

So, this is the important region in both NOESY and TOCSY that we will be using for assignments and another region of this spectrum also has information, but we will that is not we are going to use for assignment purpose that is from the sequential assignment purpose. So first thing is that we have to identify the spin systems so that is what we saw in the last class that for a given NH region, so if I draw a line like this I should be able to figure out what amino acids is this corresponding to.

Similarly, all the vertical lines all the NH is corresponding to the different amino acids. So, how to identify the amino acids? It will be based on the spin pattern the pattern of peaks for the aliphatic part and we have seen that all the amino acids have a very characteristic peak

pattern which will be useful for assignments, so let us go through that so this is the region which we are going to focus on which has all the NH H alpha or H protons Correlation.

(Refer Slide Time: 1:58)

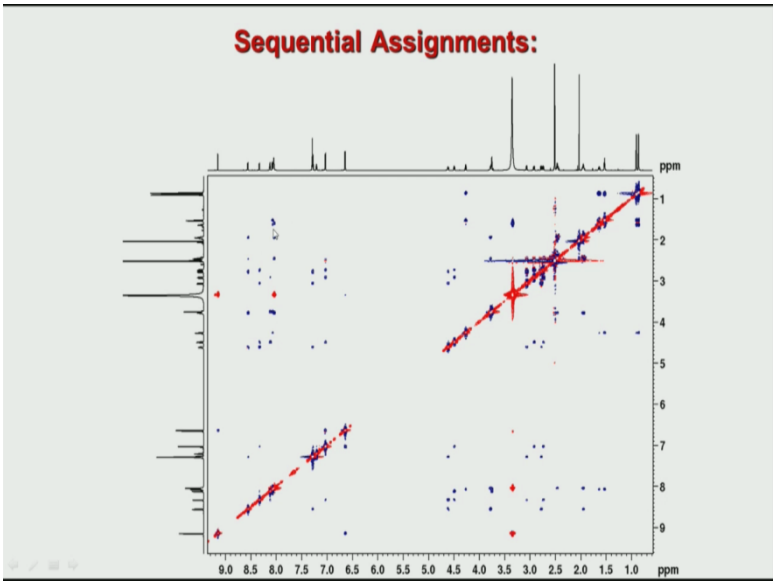


So, this is a spin system assignment is illustrated here. So as we discussed, we go from the amide region that is the NH and we go all the way vertical up and then we will identify the peaks and I look at them try to identify what amino acid is this particular spin system. In this case, this looks like a phenylalanine. We have one more AMX remember we discuss this is called an AMX, but a nevertheless this looks like an phenylalanine, if you go in this line this also looks pretty much like a phenylalanine, but it could be the next amino acid which is tyrosine but remember we do not know which is which right now. Although, I have labeled it phenylalanine here and have written tyrosine here, it does not mean that we have assigned it correctly.

In reality when we actually start with an unknown spectrum with a known peptide sequence you do not know which is which, so we will see how we can establish that identity that this line is indeed tyrosine and this line is indeed phenylalanine. Then we look at 3 more line because we have 3 more amino acids and we can see that this particular in a amide line corresponds to glycine because it has only 1 peak and that comes in the region of 3.5 to 4ppm and we from (())(03:18:08) from the database we know this should be a glycine. The next line here should be a leucine, the reason being that this peak is alpha which is not really characteristic alpha regions gain we should remember that they are not really characteristic because they vary with the structure but what is important is a beta, gamma and the methyl region.

Here there is one peak which is not shown here but it is there in the spectrum, it was at a low peak level, so it has not been seen here, so these are the gamma peaks and delta peaks. So this is the leucine's spin system which we can assign and this looks like a methionine again based on the characteristic pattern of peaks which we have seen in the last class. The many amino acid all the 20 amino acids have this characteristic pattern. So, this is how we can say that there are 5 spin systems in this molecule and each of them corresponds to the spectrum or each of them correspond to the pattern which we expect, but given that again we have not assigned sequence specifically meaning the we do not know whether this is really tyrosine or phenylalanine. We have to establish that link with the neighboring amino acids and that for that we need to go for 2D ROESY or 2D NOESY.

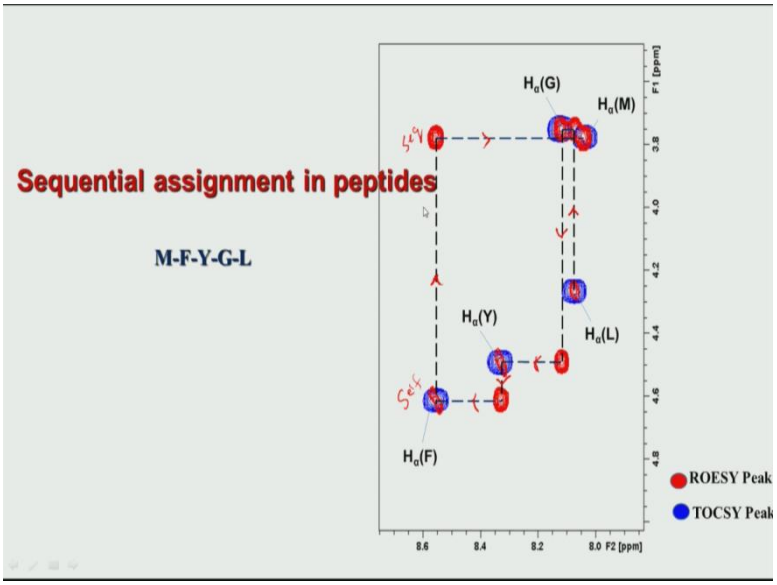
(Refer Slide Time: 4:38)



So, this is the 2D ROESY spectrum of the peptide and you can see that similar to what we see in TOCSY, but when we look closely we will see that you get additional peaks for each line and we have seen this in the last class where we looked at example of alanine and leucine where we saw that in ROESY we not only see the spin system for of TOCSY for a given amide. Suppose I draw a line here for a given amide and see not only the cross peaks on TOCSY, but I would also see additional peak from the previous that is sequential i - 1 amino acid. So, for example, if this line if I draw a line here if it that corresponds to some amino acid I, which I do not know what is i but let us say some arbitrary amino acid i then on this is if I compare with TOCSY I will get correlations of the same amino acid from TOCSY, but I will get additional peaks as well which will be we corresponding to i - 1.

So, therefore ROESY helps to identify a dipeptide i and i - 1. So, based on that one can actually go from i to i - 1 then you go from i - 1 to i - 2 and so on like we saw in the last class and that helps you to establish a link between 2 neighboring amino acids and that is what is important when we want to do sequential assignments. So, let us say in this case in this or how we can do that. So, for sequential assignment what we do is we take this Halpha amide correlation region and this where we will as we saw in the last class we can go along the backbone. We say use we call this as a sequential backbone walk so we walk along the backbone as we will see how we can do that for this peptide.

(Refer Slide Time: 6:34)



So, this is shown in this particular slide here. So, what is shown here is in the red color what the peaks in red color are coming from ROESY and what you see in the blue color are peaks coming from TOCSY. So now if you notice, suppose I again take an amide vertical line along this, this is the portion of the whole spectrum. So, this is only the H alpha remember if you going from 3.5 to 4.5 or 5ppm and it is portion of this spectrum and this is amide region. So, for a given amide spin system I am seeing a blue color H alpha in TOCSY which in definitely now we can called it as self means its own. But along the same line I will see an additional peak which will be sequential means i - 1, so I am getting i and i - 1 in this spectrum for every amide. So we can see in this case I am getting again the same, here also I am getting the same and here also I am getting the same.

Now, with this information how do I start the assignment? So, what we do is typically we start from an arbitrary spin system whichever you like. So, let us say we start from a simple glycine because suppose I do not know which amino acids are which peaks and glycines are

easy to identify because we have characteristic signal peak and that can be starting point so I let us say I start from the glycine. So, I have identify the glycine in a previous TOCSY spectrum and that glycine is this line here the black line and here you can see this is a glycine peak because there is a TOCSY peak also coming there so together means the TOCSY this is a self-peak. If I walk along this line vertically I see a red peak which is not there in TOCSY, it is only that peak here it is not TOCSY which means it is a sequential peak to glycine.

So, now I have identified the peak H alpha this horizontal remember is H alpha. So, I have identified the H alpha belonging to the sequential, and what is sequential to G it is tyrosine. So, now I have got the tyrosine peak here, now I will walked along the same line because along the same H alpha and there where I see a blue color peak that will become the self of tyrosine, remember TOCSY is self. So, on that particular peak I go to ROESY and then I walk along this line again here and I see another peak which is not in TOCSY, but only in the ROESY. So that should now be the sequential peak of tyrosine and that if I now go along the vertical line, I go to the TOCSY and get a blue peak which now is a self of this sequential peak, so what is sequential to Y?

Sequential to Y is phenylalanine so therefore this blue peak is a self of phenylalanine. Now, in the ROESY I go along the vertical line again and I see there is one i - 1 and that could be now the sequential to phenylalanine with his methionine. And methionine I go now to find its amide, I will go straight along this line, I see that there is red peak here which is over here this one there is a TOCSY peak here. So, we have to go from ROESY to TOCSY and in that TOCSY I am seeing the self of the methionine. Now, if I walk along this line up or down, I do not see anything which means methionine is now the last amino acid the N-terminal amino acid which is correct and based on this sequence. So therefore, we may say that methionine is the last amino acid and we have drawn from G to in this direction left and come up to methionine.

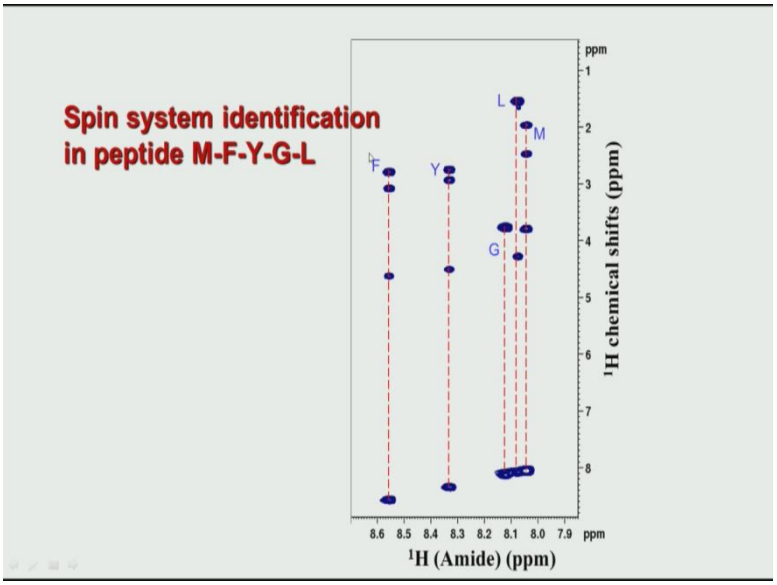
Now, how do we identify leucine or further if there is any amino acid we again start from the same glycine where we started from. Now glycine we were looking at i - 1, which was this but if glycine is i - 1 to L. So, that means at some position where self is there along that line I should get the peak matching with G because for L, G is i - 1. So, that means I have to search for a vertical line along the amide region such that I will see two peaks one of course will be the self-based on the TOCSY, but the other peak should match with the sequential H alpha that is with glycine H alpha.

So, all basically we have to do is start from the glycine amide here the vertical line but now go horizontal you went vertical to go to i - 1 but I will go horizontal to go to i + 1. So if I go horizontal to go to i +1. i this is self-peak that is the blue color of TOCSY and I see that it matches with another peak here which is there is no TOCSY, there it is a ROESY peak. So, that has to be now if you see in the vertical line that has a TOCSY peak here.

So, that means this vertical line this amide has to be the next i + 1 to glycine because only then it will show the glycine as i - 1. So, basically what is the conclusion? We can say that if I go vertically along an amide I find i - 1, if I go horizontally along an the ROESY I will get i + 1, so i - 1, i + 1. So, therefore I got an i + 1, and if I go vertically now here this is the leucine peak because for leucine glycine becomes i - 1 okay.

So, now you can complete this walk along the backbone. So this is like a walk along the backbone and you can say I can start from here and go up to here and you can see I can trace a path, which is going from i to i -1 and it ends finally at methionine. So, this is the sequential assignment process that we will normally carry out in this manner. So once sequence specific assignments are over so, we can see now I can say that this was tyrosine and this is phenylalanine that information I did not have when I went to TOCSY.

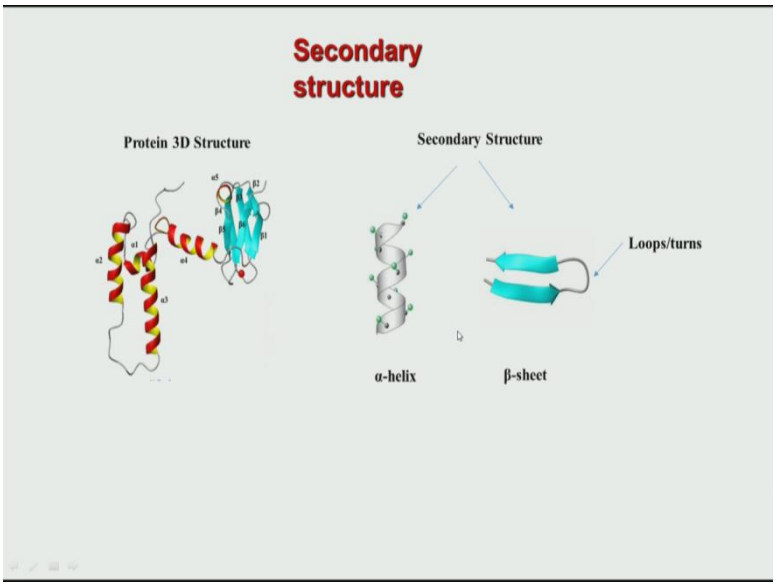
(Refer Slide Time: 13:06)



So, if you go back to TOCSY here we did not know which is phenylalanine which is TOCSY or which is a tyrosine and this was an arbitrary assignment given but it is verified only when we do the sequential assignment where we actually find a link between the amino acids. So, this is the whole process of sequence specific assignment in peptides. A similar process

happens in a proteins but in proteins this 2D experiment can be little messy because let us say you have 100 amino acids then you will get 100 such blue color and red color peaks, which becomes very complicated. Therefore, for proteins we have to go to 3 Dimensional NMR experiments which we are not going to go into this course but the basic philosophy idea behind sequential assignment is basically the same that you establish a link between i amino acids i and i - 1.Usually, it is done through the amide correlations.

(Refer Slide Time: 14:04)

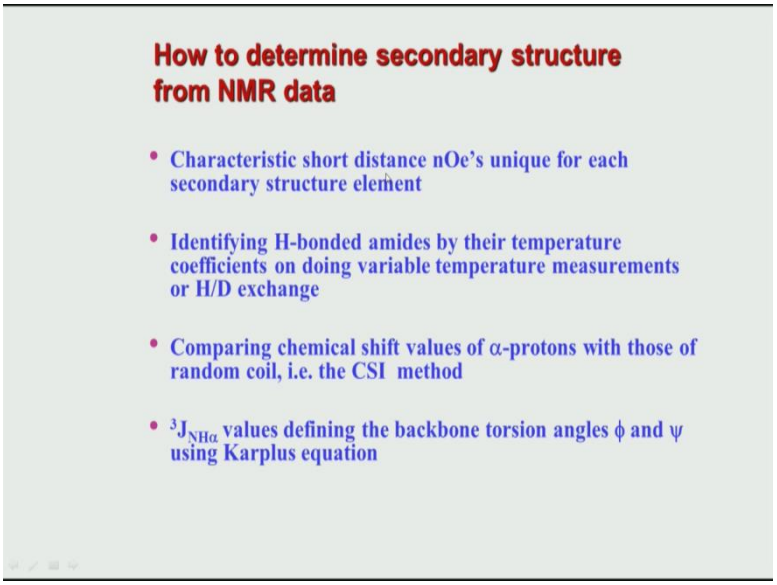


So, now once a sequence specific resonance assignment is completed, the next process in the structure determination is to identify the secondary structures in proteins. So, remember in proteins and peptides we know call this as tertiary structure and the elements which make up the tertiary structures we called it as secondary structure. The secondary structure is typically alpha helix, beta-strands or sheet and you have what is called turns or loops. Loops are basically, region which does not have any structure, so we do not call them as regular secondary structure but turns have very specific geometry. So, this kind of different parts of the protein structure have to be identified in the peptide that means we need to know which part of the peptide has alpha-helix, which part of the peptide has beta-strands and which part of the peptide is loops and so on.

So, the example which we have taken is a very short peptide so in that peptide you do not expect a structure. So therefore, we are not going to look at real example now, but we will go through a theoretical way that how this secondary structures can be illustrated by NMR. So, that is what is written here that before 3D structure is completed or computed the secondary

structure elements in the peptide or protein are elucidated, so how do you do that? There are verities of methods in NMR to do that so let us see go through them.

(Refer Slide Time: 15:43)

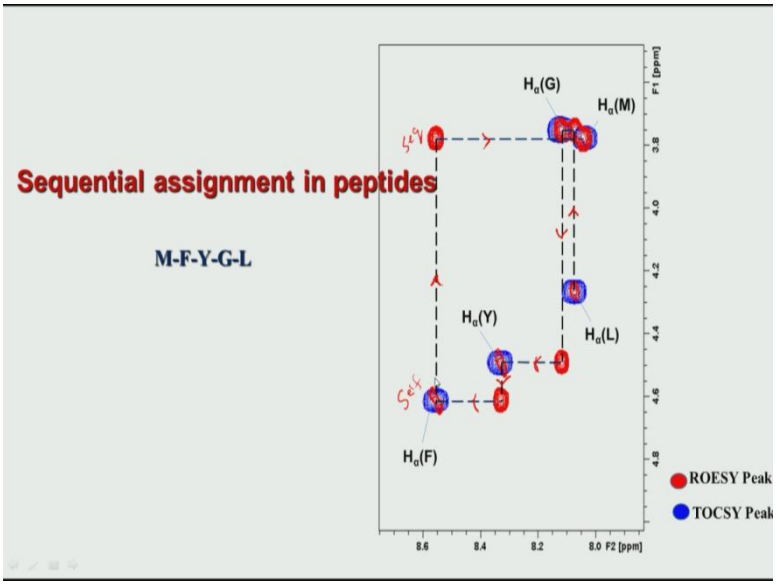


So, one of the approach is to look at what is called as short distance nOe's means short distance cross peak between two residues, which are close to each other in a NOESY spectrum and typically what happens as we will see that in case of alpha-helix , beta-sheet they are very characteristic like we saw the peak pattern.

Similarly, there is distance pattern between residues which are very characteristic of alpha-helix or beta-sheet. So, based on that kind of distance or connectivity information in the NOESY spectrum, we can find out whether that portion of the peptide or protein has alpha-helix helixical nature or beta-strands so we can we will see that shortly. Another approach is by hydrogen bond that you identify the hydrogen bond pattern in peptides and typically what will happen is, loops which are very flexible, they do not have a strong hydrogen bonds but alpha-helix and beta-strands are very hydrogen bonds.

So, by looking at a some hydrogen bonds pattern, we will see later on in the advance section of this course how hydrogen/Deuterium exchange is done, based on that one can identify hydrogen bonds in the peptides regions different regions and that can also be very useful for finding out the secondary structure. The third is to look at chemical shift index so what is chemical shift index? This is called CSI method chemical shift index. So, chemical shift index is basically is for name. So, you once you have assigned the alpha-proton so let me go back to the NOESY spectrum here.

(Refer Slide Time: 17:23)



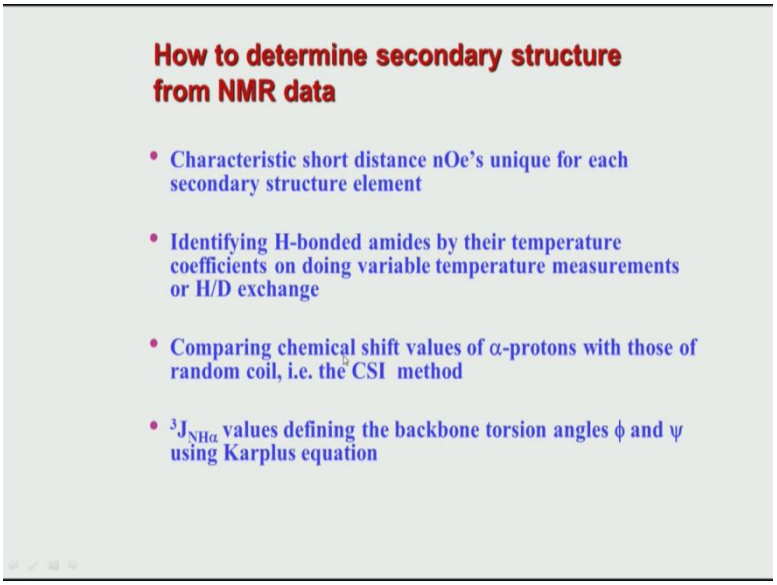
So, this blue color peaks what you are seeing are called H alpha so if you look at horizontal line this is 3.8, this is 4.3 and so on. Now, this particular chemical shift value for alpha-protons depends very much on the secondary structure. So what is done is, you take a random coil peptide, for example, you take G in a peptide like XGX where X is any amino acid. So, that in that particular peptide the glycine will be in a random coil. So, you can take a G-G-G sequence for example. Now, for each of the 20 amino acids which we have in proteins we can look at these random coil peptides and for each of these amino acids we can find out where the chemical shift of that amino acid will come if it is in the random coil conformation or random coil structure.

So, for each of the 20 amino acids this is run systematically by taking tripeptides X-G-X-G, where X is the same amino acid we have interested in, and these values are actually stored in the database. So, there are several databases where you can find these values for random coil values of each amino acid, so now what you do next is this value which you are seeing in your peptide observed in your peptide you subtract this - the random coil value for that amino acid, for example, if you are looking at leucine you subtract the H alpha of the leucine in your peptide, for example, here it is 4.23, then that you subtracted from the random coil H alpha value of leucine available in the database.

So by doing the subtraction we generate a number we call it as secondary shift. So secondary shift is a word given for that value which is basically the value seeing in for the peptide - the random coil for that amino acid value for that amino acid value. So, like that you can generate secondary shift for each every amino acid based on the value given in a database.

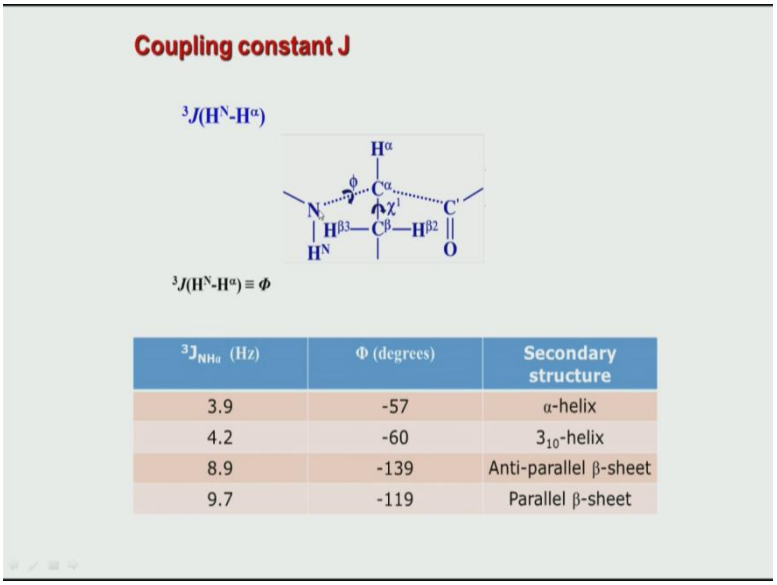
Now it turns out that the secondary shift is very useful indicator of the alpha-helix or beta-sheet that is why we call it as secondary shift, it is useful indicator of secondary structure.

(Refer Slide Time: 19:56)



So, by looking at that secondary shift value there is method known as CSI chemical shift index which helps to find out based on the deviation of the random coil from the observed value that is based on the secondary shift values the method gives you an idea whether it is alpha-helix or beta-strand or beta-sheet. So, that is a very standard approach these days for both peptides and proteins and very useful for finding out local secondary structures. Another approach is to look at J three bonds j-coupling and this coupling is easy to measure. It is the backbone NH H alpha coupling, which helps you to find out the torsion angles. So, if you recollect in our j-coupling part of this course, we saw that you can actually find out if there is a relation between 2 bond J-couplings and the torsion angles via the kar+ equation, so which we see in this slide here if you can look at this HN H alpha coupling this 3 bonds.

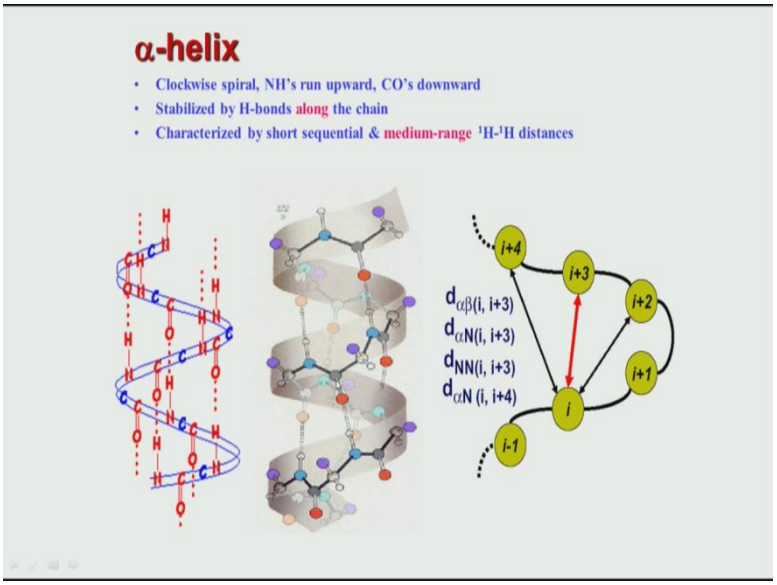
(Refer Slide Time: 20:55)



$^3J_{NH\alpha}$ (Hz)	Φ (degrees)	Secondary structure
3.9	-57	α-helix
4.2	-60	3_{10}-helix
8.9	-139	Anti-parallel β-sheet
9.7	-119	Parallel β-sheet

So, this is one, 1 bond, 2 bond and 3 bond, so this 3 bond coupling gives you information on this torsion angle phi and based on the value of 3 bond j-coupling you can see that the phi be lies in different regions. So, if the j value which can be measured there are different methods to ways to measure j value, one of the standard method is you just look at the doublet of this amide peak. So, we will see in a 1D so in the 1D when you look at this NH peak in this peak the proton peak, it will be split into a doublet because of j-coupling to H alpha, which is 3 bonds away. So that splitting can give you j values and that values now indicates whether it is an alpha-helical region which is based on this angle or depending on and if the value is high that is of 9 to 10 (())(21:45) hertz that value gives you tells you whether it is now in a beta-sheet, so based on this value also you can find out the secondary structure of the peptide.

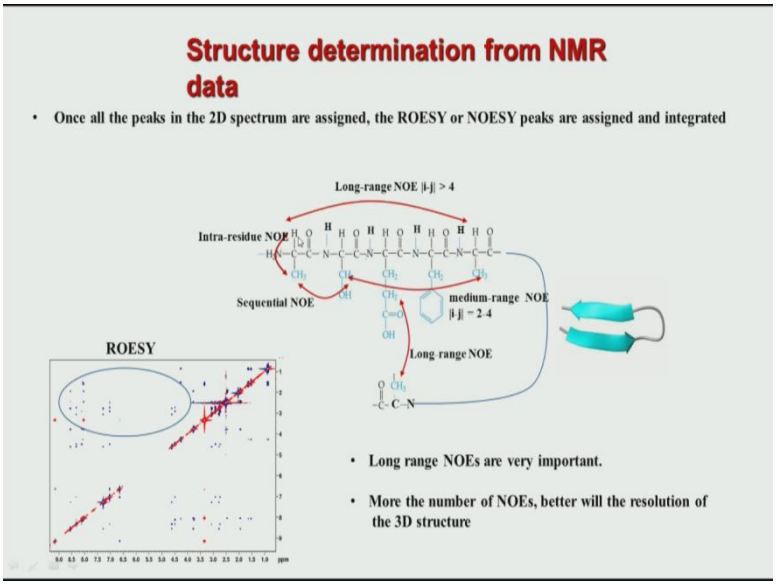
(Refer Slide Time: 21:58)



And now, this is another way to find out secondary structure as we saw that is called by nOe's. So what is nOe? So this is what we shown here, suppose this is an alpha-helix okay, so it is a part of an alpha-helix, Alpha runs like this. Now, in this i is a residue number so if it is an alpha-helical structure, these amino acids that is i + 2, i + 3, i + 4, they come close to this residue i. So, within residue i specifically it is alpha-proton to beta-proton. So, if you see the alpha-proton of i to beta-proton of i + 3, alpha-proton of i to beta-proton of i + 3, there is a very close distance within 5 armstrong and that helps us to find out if there is a secondary structure is helix or not. Similarly, for helix there are some more this is i, i + 4, i, i + 3, i + 2 and so on. So, basically it is i, i + 3 and i, i +4 and different alpha and amide.

So, N means for amide proton so between alpha-proton of i and amide proton of i + 4 if the distance between them is less than 5 Armstrong and how do you know it is between less than 5? By simply looking at the NOESY spectrum, you should be able to see a chemical shift correlation between the alpha-proton of i and the amide proton of i + 4. So, such a correlation immediately tells you that these two atoms are within 5 or 6 Armstrong, and if that is so then they belong to an alpha-helix because only then alpha-helix this residues will come and take a turn and they will give you information of the secondary structure. So, these are the different information which we can see from secondary structures.

(Refer Slide Time: 23:50)



Now, what we do next is to find out the tertiary structure, tertiary structure is basically the final 3 dimensional structure and for that you need to look at now the ROESY spectrum here. And in the ROESY spectrum every peak here if you notice this spots we called them as spots or peaks. Now actually our 2 chemical shift correlations between two atoms right, so one is on the x-axis another is on the y-axis. So, between the 2 protons, they give a peak and the immediately when you see at peak it implies that there is a distance of less than 5 to 6 Armstrong between those two atoms. So that is very important because all the peaks here are giving you now the distance information. So, if you look at the sequence here, an arbitrary sequence so I may see different correlations between different atoms so let us classify these different correlations.

So, number one is suppose you see a correlation between atoms in the same amino acid. So, this is alanine and within alanine we are seeing let us say correlation between H alpha to H beta that could be somewhere in the spectrum. So if know the chemical shift of H beta of this alanine and H alpha of this alanine, remember there are 2 alanines in this sequence. So I should know which alanine is which H alpha so that comes from our sequence specific resonance assignment step. So, we know now for sure this H alpha is alanine number one or this H beta chemical shift is alanine number one so on. So, once I have suppose I see a correlation or chemical shift correlation between these 2 atoms in a ROESY, it belongs to the same amino acid, so we use the word intra-residue NOE. So, this is one type of distance information you will get and for different amino acids.

In every amino acid you will have inra-resedue NOE's and second is between neighboring residues. So, you see here I may get a correlation in the ROESY spectrum or NOESY between this H beta atom of this amino acid which is serine and Hbeta of alanine, which is i - 1 to in this residue. So, that type of peak which we see in the ROESY spectrum we use the word sequential NOE. So, again you may see pair wise sequential NOE between all amino acids, so all those will be classified as sequential NOE. The third type of NOE which you will get is slightly far away, so this is between residues 2 to 4. So you can see in this case it is three residues apart and this is between this alanine and this serine hydrogen atom. So, the beta of alanine and beta of serine you may see or you may not see because that (())(26:38) depends on the 3 dimensional structure.

But observing such a NOE's is useful and that type of NOE's are called as medium range NOE because they are between 2 amino acids, which are not far away but in medium range. Now, the last type of NOE is known as long range NOE and this is very crucial type of NOE because as you can see it is giving me an information between a very far away amino acids, so for example, this alanine may come close to this alanine by because of the alpha-helical or beta-sheet nature of because the turn, so that information is obtained from a NOE between these 2 atoms. For example, this is just an illustration you may also get a NOE between this atom to this atom or any combination, but if you see a distance between 2 atom or a cross peak chemical shift correlation between 2 atoms, which are very far away in sequence, but they are very close sorry, far away in sequence but close in space and that is i - j is more than 4, then we use the word long range NOE, so this is the long range NOE between two atoms.

Now, this is what I have shown in the linear chain means this may turn around because of its structure but suppose this peptide is continues and let us say we have a structure like this. So, there is basically a beta-strand 2 beta-strand connected by a loop as we can see here. So that is basically this whole sequence turns around and comes like this and these 2 residues come to each other close in space. So, this can happen, for example like this in this structure shown here, where you have basically two beta-strands connected by a loop to form a beta-sheet and that results in this act residue coming close to this. Now, here you may again see a correlation chemical shift correlation between 2 atoms that is beta of this and the beta alpha of beta of this or gamma of this. So, this is for example, a glutamic acid, so I may see correlation to this methyl to this methylene because of this structure here, the blue color we see the beta-strand and they are coming turning around and these two are coming close.

So, that results in a very long range distance information and this is also very useful because now it is telling us that these 2 strands are coming close to each other, you may not see that information if this is a open structure. Suppose, these 2 strands are open and they are not coming near each other than that will not come this correlation. So, by seeing this correlation in a NOESY one can say that I have got long range NOE between these 2 residues. So, this is what is done in from ROESY spectrum that you have to look at all the different NOE's and of out of all of them the long range NOE's are very important why; because they are the one which helps you to nail down or to figure out the real geometry or structure of the molecule.

The sequential and intra-residue are definitely also useful, but they are not going to be give much information because the structure of amino acids are anyways known to us, so a distance between these 2 will not change much whether it is an alpha-helix or beta-sheet, but what will change is a distance between this and far away residue, so that distance information comes from long range NOE's and that is useful. Then more the number of such NOE's which you can extract from the NOESY. So, if you see this ROESY spectrum here, every peak is a information, so the more the number of NOE's you can obtain from ROESY the more better will be the resolution of 3D structure because that is how we determine 3D structure by giving the inputs of these distances which we will see in the next class.

We will see how this any distances can be used as inputs. Similarly, we can determine the phi angles the torsion angles and together they help to determine the 3 dimensional structure of the peptide or protein.