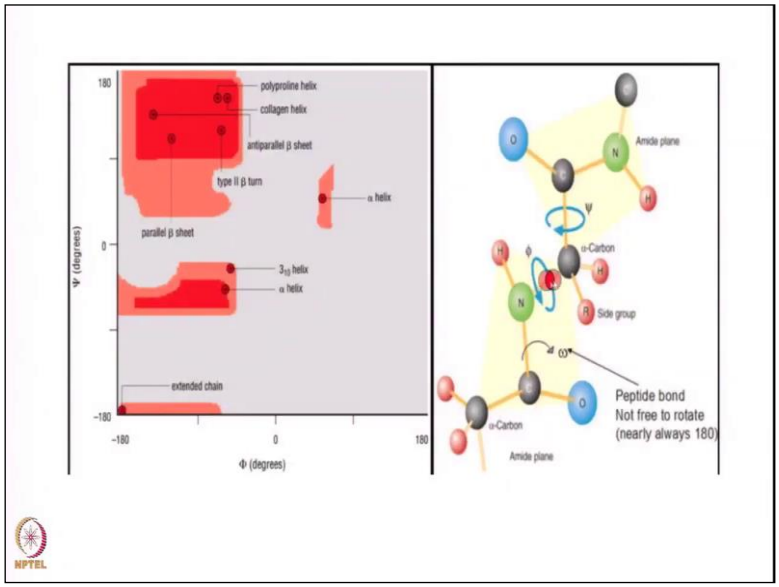
NMR spectroscopy for Structural Biology
Prof. Ashutosh Kumar and Prof. Ramkrishna Hosur
Department of Chemistry
Indian Institute of Technology - Bombay

Lecture: 32
Determination of Structure and Dynamics of Proteins - 2

So, let us continue with the description of the protein structures at a more fundamental level. We said that the relative orientations of the amino acids depend upon some torsion angles described as Phi and Psi in the previous slide and we will examine more about that now.

(Refer Slide Time: 00:38)



So, here we have the peptide chain running like this, schematically indicated here. So, you have the alpha carbon, there is a carbonyl here. So, NH the chain is running like this NH-Cα-CO-NH-Cα-CO it goes like that. All the atoms are indicated here. So, now this is the alpha carbon and then you have the peptide group. Now, these 4 atoms, this is the carbonyl, this is the double bond here, and this is the NH here.

Now, how many torsion angles are there? You have in between the dipeptide ring you have one torsion angle here, another torsion angle here, and third torsion angle here. These ones are called omega, phi and psi. So, N-Cα torsion angle is called phi, and Cα-C' or Cα-CO torsion angle is called psi. Now, what does this rotation mean?

Now consider these 4 atoms. Suppose I take this carboxyl carbon, nitrogen, and NH proton, what will be the orientation of this bond with respect to this bond? If they are exactly in the opposite direction, if they are on the same side of these 2 atoms then we say it is a cis configuration. If this carboxyl and this NH protons are on the same side of this bond then we call that a cis configuration.

But if it is this, is called a trans configuration, and the trans configuration means that this omega torsion angle which is indicated here is 180 degree. Because that is a more stable structure and by and large that does not change at all. Why is that so? Because of this carbonyl carbon, this is the double bond. So, this bond the N-Cα-N-C'-N series bond, also has a certain kind of double bond character.

When there is a double bond character, there is no free rotation possible here. So, by and large, the free rotation is restricted and the most stable structure is the one where you have these 2 going in the opposite direction. In the same way, you can also say if these are in the opposite direction, this curve this Cα and this Cα will also be in opposite orientations.

So, therefore this is the peptide bond that we call therefore the peptide bond is planar and now what is called the phi torsion angle and this is the rotation around this bond. So, which atoms do we consider? So, in this just as we consider here the 4 atoms we consider these 4 atoms this carbon this nitrogen this carbon and this carbon.

If you consider these 4 carbons, what are the relative orientations of this carbon and this carbon with respect to this bond? So, if they are in the same orientation, then it is the cis configuration or if they are in the opposite sense is called a trans configuration. Now in between things are also possible because of the free rotation possible here, because this is not a double bond here.

So, this determines the relative orientations of this amino acid with respect to this amino acid. So this is one torsion angle and then the second torsion angle is this, this is called the psi. This is also a single bond, this is Cα-C'-Cα-carbonyl. Now, we can look at what are the relative positions of these 4 atoms: nitrogen, Cα, carbonyl (C'), and this nitrogen.

If you take these 4 atoms, with respect to this bond whether this nitrogen and this nitrogen are in the same direction or on the opposite direction. The same direction will be the cis

configuration, other one will be the trans configuration. You can have all values in between. So, when we say the gauss confirmation. Typically we know from the stereochemistry, these are different possibilities.

So, you have different kinds of possibilities here. So in principle, they can all go from 0 to 360 or we also sometimes call it +180 to -180. Conventionally whether you want to define from -180 to + 180 or 0 to 360, both kinds of conventions are used and initially one has to figure out which are the possibilities. Now, what determines these choices?

Of course, steric contact is an important factor there. When you consider all these various rotations or all the orientations possible some of them are not because as a result of when you do this, there can be clashes in the various atoms the side chains of one amino acid residue can clash with the side chains of the other amino acid residue in which case that kind of a combination of the torsion angle will not be possible.

So, this was investigated in great detail by G N Ramachandran and there he calculated the energies of this dipeptide considering the different possible orientations, and different possible values of phi and psi and generated the values as the 2-dimensional energy map. What is shown here are these combinations which are possible and not possible.

On one side, you have the fine torsion angle on the other side you have the side torsion angle. And then considering the energies, he said below a certain energy it is allowed and above a certain energy is not allowed. If that is because if you have a 5 kilo calorie per mole if you have that kind of energy you look at the Boltzmann statistics then you will find that 5 kilo calories and above it is almost impossible to get any probability for that sort of a structure.

Therefore he eliminated all those combinations which lead to the steric clashes and this is popularly known as the Ramachandran plot. You see all this space which is here this is disallowed only the red ones which are indicated here these are the areas where you can have a combination of phi and psi. These are called phi-psi maps or Ramachandran plots and these are the various values which are indicated here.

You have the (0,0) in the middle here and then you have the (+180, -180). That is the convention used and, of course, that is also the trans and here you have the Gauss conformations

there. Now you see if your phi and psi angles are in this domain in this area this will lead to the formation of a helix structure. Then we talked about the helical structure. If this leads to the formation of the helical structure, there is a fine distinction made: there is a so-called α-helix or the 3_{10}-helix in this one.

In the α-helix, there are 3.6 residues per turn, in the 3_{10}-helix there are three residues per turn. This is somewhat more compact and you have that kind of structure also possible and that appears above this absolutely outside this range of the α-helix. So, Phi-Psi torsion angles are indicated for the helical region. On the other hand, this is a much larger area is allowed area. So, you have the β conformations there.

In this here you have the parallel β-sheet, if you have these values here, and you have in this area you have the anti-parallel β-sheet and there are also certain other kinds of helix. These are more like it is called polyprolines. If you have a sequence which has only prolines, then it forms a structure which is similar to the β-sheets and although it forms a helix, the combinations of the torsion angles are in this manner.

This polyproline is more like a beta structure. So, therefore this comes in this area. Then you also have what is called the collagen helix. Collagen was the structure which G N Ramachandran was working with. This came out as a result of his investigation of the collagen structure. The collagen forms a triple-stranded helical structure. So, it has very special combinations of torsion angles and you have three strands there intertwined and that relates to a stable structure.

And that results in a helical structure but the torsion angles there are in this in this area, therefore this is called a collagen helix. So, you have the anti-parallel β-sheet, the parallel β-sheet and then you have also a type 2 β-turn. So, this is the type 2 β-turn. Type 1 β-turn also appears somewhere here. So, you have these various kinds of beta turns which are appearing as the secondary structures and the combinations are possible.

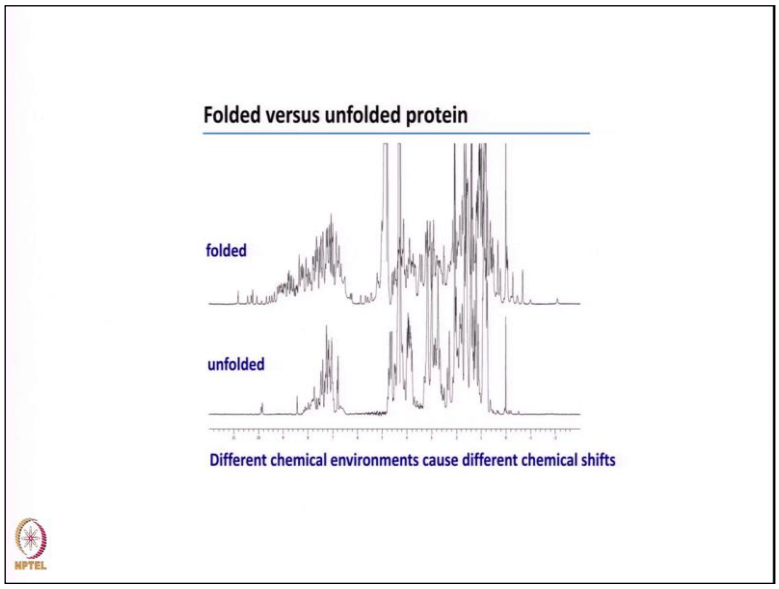
These are slightly extended areas. The thick ones are the ones which are very strongly favoured and the ones which are slightly thinner the shaded ones are less favoured, they are possible, it is not that they are not possible, they are possible, but they are less favoured. The thick areas

are the most favoured and then you see here you have a so-called left-handed helix that is present here.

And this particular region is α helix but it is a left-handed helix, these ones are right-handed α-helix. This is the left-handed α helix and you see the most important thing is much of this 2-dimensional space is disallowed. You cannot have combinations of this. If anybody determines the structure of a protein with this Phi-Psi torsion angle falling in this area then the structure is not acceptable.

Because it will lead to steric clashes and certainly will not be a stable structure.

(Refer Slide Time: 10:50)

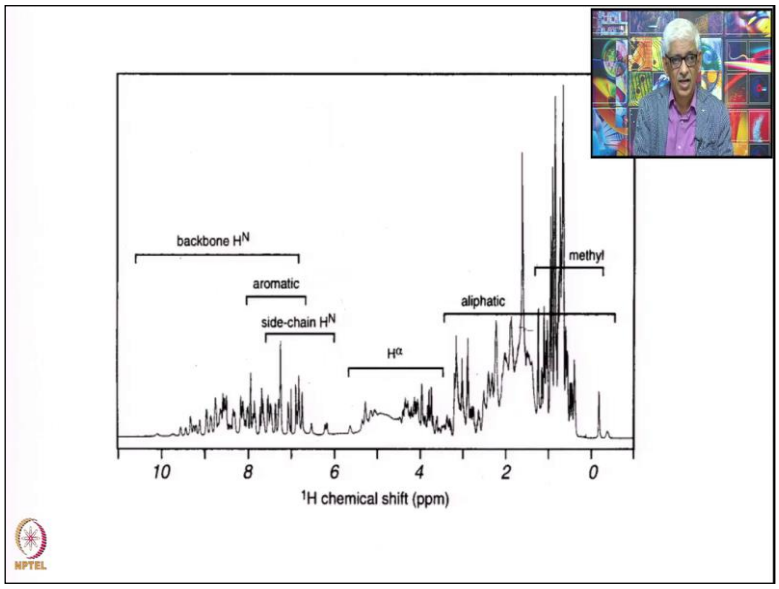


Now, having looked at the various possibilities of the structures and we want to go into the NMR application. How to determine the structures and what are the characteristics of the protein spectra depending upon what is the nature of the structure? And here is the typical illustration of if you have a protein which is very well folded that means it has regular combinations of helices, sheets, turns, etc everything is present.

Then you may have a proton spectrum, this is the proton spectrum goes from 0 ppm to 11 ppm, is a beautifully spread out peaks here and in the case if the protein is unfolded in the sense that there is no particular preference for a particular combination then there will be dynamics in the protein chain and there are multiple combinations are possible for each amino acid and then what we will see is an average of this chemical shifts.

So, for the individual protons which we talked about which are present on the amino acids, they will all be in a very narrow region here. So, such a kind of structure is called an unfolded structure. It has no specific combination, no stable helix or β-sheet and things like that. So, they will all be exchanging very rapidly and you will have an average chemical shift and all of them will fall in this area. What are these different regions? So, let us look at that.

(Refer Slide Time: 12:13)

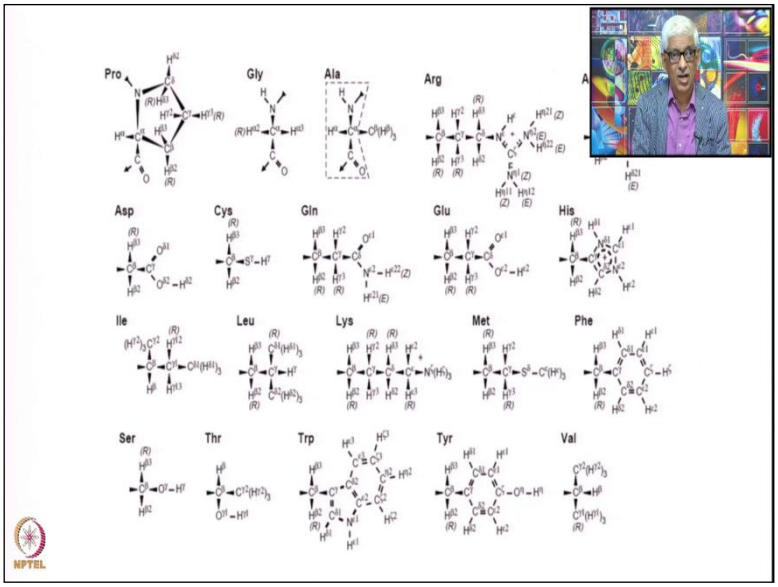


And because of these different chemical shifts, NMR provides you with a tool for determining the structures of three dimensional structures of the protein. Now here is a detailed analysis of what sort of chemical shifts are present for the different protons in your polypeptide chain. So, typically this area from here to here, these are the backbone amide protons, this is almost from 6.5 ppm up to 10.5-11 ppm.

You have the backbone of amides and here you have the aromatic protons, which are up to 8 ppm, typically about 6.5 to 8 ppm you have the aromatic protons. Then you have the side chain amide groups, where are the side chains, these are the asparagine and glutamines and also in arginine and lysine these are the 4 residues which had amide groups or the NH groups in the side chain and they will appear in this area from 6 ppm to 8 ppm.

And then along the backbone you have the α-protons. This is Cα attached to the Hα proton and then you have all other side chains here in the aliphatic groups and the methyl groups are there for the valine, isoleucine, alanine and then the threonine. So, all of these methyl groups appear in this area. So, therefore they have very characteristic chemical shift patterns for the different types of protons in your polypeptide chain.

(Refer Slide Time: 13:59)



Now here I want to show you once again the same amino acids but the nomenclatures of the individual atoms. So, these are individual residues you notice here, where is the chain going? This is the carboxyl and this is the nitrogen and these are now in the polypeptide chain, they are present in the polypeptide chain, not individual amino acids.

So, this is the nitrogen, then the Cα, carboxyl. This is the chain which is running like this and then the side chain is going like this. You have the Cβ here, this is the methylene group. So, this is Hβ1, and Hβ2, Hβ3 here. Then you have the Cγ. So, then you have this Cγ. It has 2 protons there, Hγ and then you have the Cδ, once again they have these 2 protons there, and this joins the nitrogen, therefore proline does not have a free amide proton on the backbone.

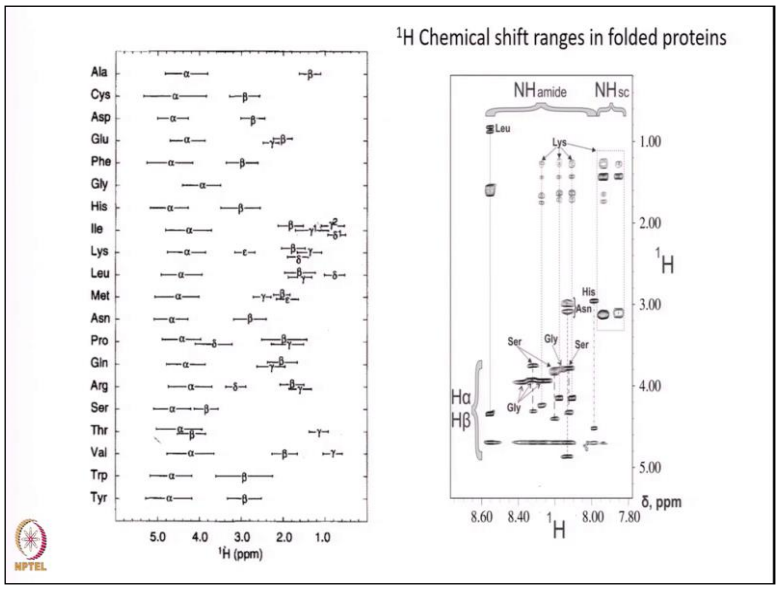
The Glycine, for example, has 2 α-protons and the chain NH-Cα-CO. Like this alanine has one methyl and one α-proton, Arginine has 2 β-protons, 2 γ-protons, 2 δ-protons, and then from this epsilon (ε) you have this one amide proton here and then you have this another carbon going over to the NH_2 group here, then the N and then NH and NH_2 there.

So, therefore the arginine is a quite elongated chain. So, it has a bit of basic nature here, because of many nitrogens present, and asparagine as we said already had $CONH_2$. You have the COH here and then you have this NH_2 there. This is the nomenclature. They are called delta, these atoms are labelled as gamma, delta and so on. So, aspartic acid again has the COOH here and the side chain Cβ, at the gamma position you have the COOH.

And cysteine, at the gamma position you have the SH and then you have the 2 β-protons there and the glutamine has again the $CONH_2$. At the gamma position, you have 2 protons. Similarly, in the glutamic acid, you have the COOH in the side chain here and that appears as the delta. Why I am going through all of this is because of these nomenclatures one will use very frequently when you describe the protein structures. We say here is the proton beta chemical shift, gamma chemical shifts, delta chemical shifts, and so forth.

Therefore one should know which proton we are talking about. Histidine has this nomenclature here you have the Nε. There is a proton here and then there is a proton there, which is attached to these are the 2 protons which are there. An exchange can happen between these 2 protons. So, similarly, we have this α β γ δ for all the amino acids. This is the typical convention which is followed in the IUPAC nomenclature of the amino acids.

(Refer Slide Time: 17:16)



So, you immediately can see here, now what we are trying to show here. Now, these are the various amino acids, Alanine, and Cysteine with the 20 different amino acids, and what are their proton chemical shift ranges? The proton chemical shift range is quite distinct although there is a certain range of overlaps here. There are many things which are overlapping in this area but alphas are appearing in this area which is already indicated there.

And these ranges indicate the variations possible depending upon the protein structure what is the kind of environment for this particular alpha group. So, depending upon that you have certain variations. Now the distinctive features are of the alanine, the beta is very distinct here. So with regard, there is only one alpha and then there is the beta and the cysteine has the beta

which is somewhere here, aspartate also has a beta here, and glutamate has the gamma and the beta.

So, you must be able to connect all of those there. The phenylalanine is alpha and at the beta, of course, there is no connection to the aromatic rings here aromatic proton positions are not shown here. Because they will also be aromatic range protons there. The glycine is only alphas, histidine is alpha-beta. So, these are called the ABX spin systems. All these are long-side chain spin systems. We have this and then you have some which are only alphas and betas.

There are many residues which have only alphas and betas. See this one is alpha beta and of course, this one is alpha-beta, this is alpha-beta, and this is alpha-beta gamma, alpha-beta, alpha-beta. There are many like that. This one is an alpha-beta, and once again this is alpha-beta. But the serine beta is very close to the alpha area this is because of the OH group which is present on the side chain.

The beta proton comes very close to the alpha. Tryptophan, this is again alpha-beta, tyrosine is alpha-beta, and threonine also has alpha beta which is very close to these. Because once again there is OH group on the threonine side chain but it has also a methyl. So, that is the gamma and that distinguishes it from the serine. Threonine and serine are distinguished because of this.

Although these areas are overlapping, the threonine has a methyl in the gamma position and the valine alpha beta gamma. Now here is an experimental spectrum and this is a so-called TOCSY spectrum, you can see here, the various amino acids which have typical peak patterns. This is essentially what is listed here explicitly shown in this. Here you have the alpha-beta which will be present.

But the betas of the serine and threonine will be there, but of course, in this sequence, there is no threonine, you only have the serine. All alphas are there for a particular section of a protein spectrum and then on this side you have the betas and the gammas and all of those and this area belongs to the lysine, and these are the side chains. This area belongs to the side chains, these appear in pairs.

Because the $CONH_2$, the 2 protons on the NH_2 are non-equivalent and they will be connected to the thing in the side chain in this manner. So they will be in pairs, and they will appear

together. So, you see the correlations, you can see these peaks in the TOCSY or you will also see them in the NOESY spectrum.

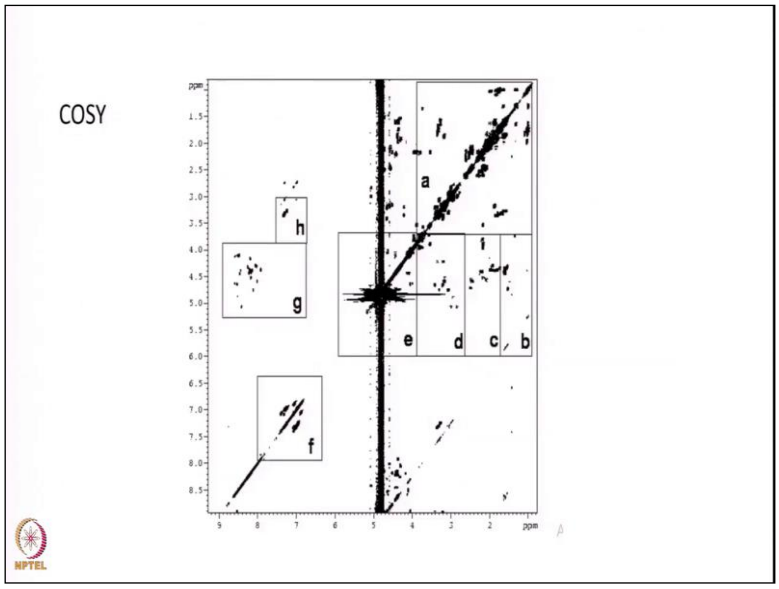
We already talked about the TOCSYs and the NOESYs. I will show you again those below and here you correlate the glycine. So, the glycine is here, and you look at the serine here, serine alpha is here and the beta is here, which is quite close. There is one glycine here another glycine there. There is no other one below or above that. So, the lines are drawn here to indicate that they all belong to a particular amino acid and from that particular amino acid you have the same pattern.

So, you look at the lysine here, this is the alpha of the lysine and goes all the way to this beta, gamma and all of those there. So, now here from histidine, histidine you are seeing to this beta and does not have any other one there. These are epsilons from the side chains of these, you will also see the betas of the asparagine and the glutamines and also sometimes you will see the epsilons of the lysine, arginine, and they will also appear in this area.

You can see here the lysine, see the beta gamma delta, they are appearing here, and the epsilon is appearing here. The epsilon is at a chemical shift of 3 ppm very distinct. And sometimes you may see these peaks, we may not see this peak, you will see them in the NOESY, but you may not see them in the TOCSY spectrum because the relay has to go all the way down.

Depending upon what is the mixing time you use in the TOCSY spectrum it may not reach up to the epsilon but it will certainly reach up to the alpha beta gamma and sometimes the delta. So, you will see distinctive intensities for these different peaks as you are looking at them. Of course, what is present here is the water.

(Refer Slide Time: 23:00)

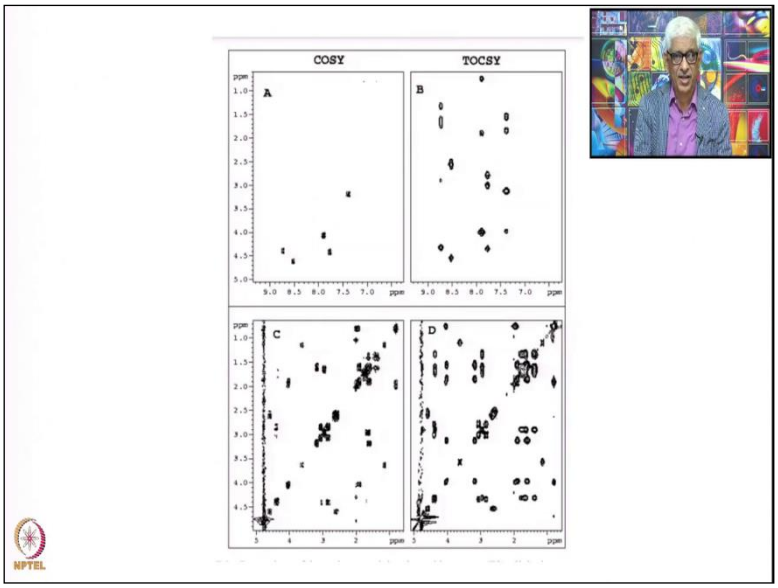


Now, this is the typical COSY spectrum of a peptide and the various regions are indicated there. So, what are these here, these ones are the aromatic protons. As we said the aromatic ring protons have correlations among themselves the coupling is between themselves. So, in the COSY spectrum, you will only see cross peaks here for the aromatic ring protons. This area belongs to the NH-CαH and NH proton is coupled to the Cα. This is a three-bond coupling and you will see those in this area.

So, therefore the COSY spectrum shows NH to the CαH. Now from the CαH, of course, you will see the Cβ here, and the beta to the gammas, you will see beta-beta etcetera. This region is a quite crowded region there. So, one can identify only from the alpha to the betas here then you have the beta gammas and then you have the gamma deltas and delta to the epsilons and the methyls etcetera. They will all come to this area.

So, this is typically categorized in certain boxes here depending upon what sort of proton pairs are involved in this generation of the cross peaks.

(Refer Slide Time: 24:11)



So, you see here the difference between the COSY and the TOCSY, and this is the same this is again the NH to the alpha. So, this appears in this area from 9 ppm to 7 ppm. You see a cross peak to the individual alpha amino acids, α-proton of the same residue in this case.

Of course, they have a fine structure which is expected because all these protons have a fine structure and that will reflect in the cross peak also. Therefore you have 1 2 3 4 5, and you can identify these 5 amino acids there and then correspond to those of course you will also in the TOCSY spectrum, you will see additional peaks, and what those additional peaks are, they will belong to the beta, the gamma, and the delta.

Therefore looking at this kind of pattern, you can identify what sort of a spin system, and what sort of amino acid it is. For example, this one has many protons which are connected to it. So, therefore this is clearly a long side chain and this one has only one fellow here. So, this is alpha and then there is only one there are 2 peaks here possibly there and these ones are beta 1 and beta 2.

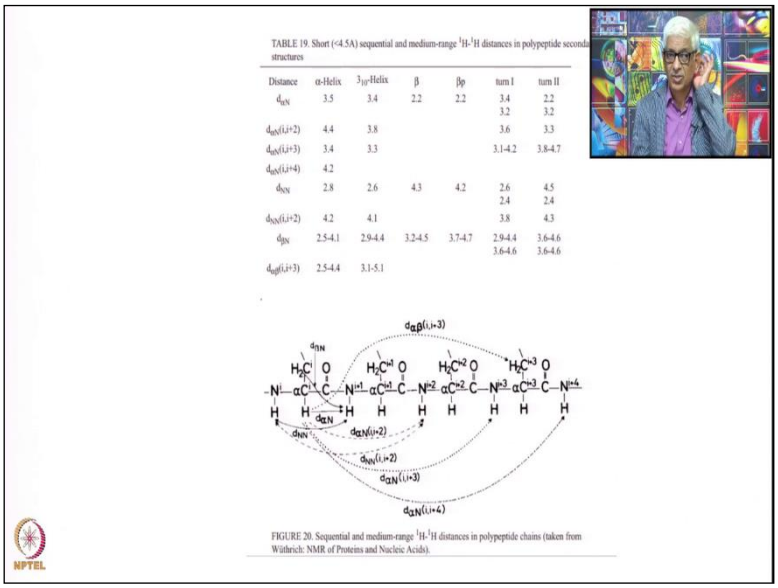
So, the relay has happened from α-proton to the beta 1 and beta 2 there. Now, this amino acid has an α-proton here, and then it goes to the beta here, and then to the gamma there. So, therefore this also is a little longer side chain. See now, it sees it goes into the methyl area once it comes into this area it is the methyl proton region. Therefore these are likely to be like valine, isoleucine, or leucine, and things like that.

Now you look at the next one here, this α-proton has 2 peaks here and these are around 3 ppm. Looking at the table that I showed you here, this is again 2 beta protons of a particular amino acid residue and these are most likely to be the aromatic ones. Aromatic ones are present in this area, this is the three peaks these are tryptophan, phenylalanine, tyrosine, and sometimes histidine.

So you might find these, it can belong to this kind of residue, and here you have the last α-proton there. Now, this peak corresponds to this here, but this one also has certain other ones residues here which are going like this. So, therefore there are threonines or beta protons or α-protons and so on and so forth. There can be some other residues overlapping, this is what you see in the TOCSY.

And this is the more detailed region or the other region of this of the same spectra. So, you have a peak coming from in the alpha proton. In the aliphatic area, you have this various correlations the beta1 and beta 2, alpha to the betas, and these will be the alpha to the beta cross peaks, then you have the beta to the gamma cross peaks, then your gamma-gamma cross peaks here. So, you will have a whole set of cross peaks which one needs to analyze by looking at the sequence that you have. Then you can fix those.

(Refer Slide Time: 27:18)



So, now that is for the identification of the individual amino acid residues. Now, you have to connect them sequentially because this spectrum only shows within the same amino acid residue. COSY does not show peaks between amino acid residues, from one residue to another

residue it does not show because there is no such proton-proton coupling which can tell you that.

However, these will appear in the NOESY spectrum because the NOESY spectrum reflects on the proton-proton distances. So, the proton-proton distances can be quite varied depending upon the structure that you might have in the protein. So, you certainly have near-neighbour interactions. Near neighbour interactions are sequential interactions, you can also have long-range interactions which will determine the secondary structure of the protein.

So, there is a brief of the distances here, the various distances which are present in the polypeptide secondary structures. Let us look at this particular we will have to go through this again in greater detail, and the polypeptide chain is indicated here NH-Cα-CO-NH-Cα-CO-NH-Cα-CO, the chain is running like this, residue I, residue i + 1, i + 2, i + 3 and so on so forth. Why you are taking 4 residues?

Because we are going to have short distances between the protons as far as 4 residues and that is in the regular secondary structures. Therefore one has to show these 4 residues and it is the various short distances indicated by lines there, and we will go through these individual ones possibly in the next class, I will stop here.